European University Institute

Max Weber Programme

# EUI Working Papers

**MWP 2008/36**

The Effect of Language at School on Identity and Political Outlooks

Oriol Aspachs-Bracons  Irma Clots-Figueras  Paolo Masella

**EUROPEAN UNIVERSITY INSTITUTE**
**MAX WEBER PROGRAMME**

# *The Effect of Language at School on Identity and Political Outlooks*

ORIOL ASPACHS-BRACONS IRMA CLOTS-FIGUERAS PAOLO MASELLA

EUI Working Paper MWP No. 2008/36

ISSN 1830-7728



Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy

http://www.eui.eu/
http://cadmus.eui.eu/

**Abstract**

The process of individual identity formation is still an enigma, as it is the capacity of public bodies to intervene in it. In 1983 the Catalan education system became bilingual, and Catalan, together with Spanish, was taught in schools. Using survey data from Catalonia and exploiting within- and between-cohort variation in exposure to the Catalan language at school, results show that individuals who have experienced greater exposure to teaching in Catalan are more likely to say that they feel more Catalan than Spanish. Interestingly, the effect appears to also be present among individuals whose parents do not have Catalan origins. In addition, the reform affects actions too, as individuals exposed to more teaching in Catalan are more likely to vote in regional elections and vote for Catalan regionalist parties. To the best of our knowledge, this is the first paper to analyze empirically how policies affect individual identity.

**Keywords**
Identity, Language, Education, Voting

# The Effect of Language at School on Identity and Political Outlooks*

Oriol Aspachs-Bracons†, Irma Clots-Figueras‡, Paolo Masella§

October 8, 2008

## 1 Introduction

> "Of all political questions, that [of education] is perhaps the most important. There cannot be a firmly established political state unless there is a teaching body with definitely recognized principles. If the child is not taught from infancy that he ought to be a republican or a monarchist, a Catholic or a free-thinker, the state will not constitute a nation; it will rest on uncertain and shifting foundations; and it will be constantly exposed to disorder and change". [Napoleon][1]

Education has often been invoked as an important instrument of nation building and State power consolidation. In his lectures delivered in Berlin

*The authors thank Oriana Bandiera, Robin Burgess, Francesco Caselli, Maitreesh Ghatak, Luigi Guiso, Eliana La Ferrara, Ramón Marimón and seminar participants at the London School of Economics, Universidad Carlos III de Madrid, Universitat Pompeu Fabra, EUI, CEMFI, Universita 'di Pisa, IMT Lucca, University of Bristol and University of Mannheim. The authors also thank participants at the EEA meetings 2007, the SAE 2007, the RES annual conference 2008, the ESPE conference 2008 and NASM 2008. This paper combines material presented in two working papers (2007a and 2007 b). Irma Clots-Figueras gratefully acknowledges financial support from the MEC grants SEJ2004-07861 and SEJ2007-67436.

†LSE and FMG

‡Universidad Carlos III de Madrid

§European University Institute

[1]Quoted in Reisner (1922), p35.

(Addresses to the German nation, 1807), Fichte explicitly emphasized the importance of state- controlled mass education in teaching Germans to be good Germans and to create the unified national sentiment needed to restore Prussian power. Between 1817 and 1825, new taxes were imposed to revitalize the Prussian school system and a separate Department of Education was created in the Ministry of Religion, Education and Public Health: every Prussian land was required to have primary schools[2].

Ethnic diversity has been shown to be correlated with lower economic growth and low-quality governments. At the same time, several social scientists (Anderson 1983, Bates 1983, Horowitz 1985) have argued that the boundaries of ethnic groups are not exogenous and change over time. They are likely to be affected by social and economic conditions and to depend on a number of economic and policy choices. Nation building policies (for example the implementation of a unique national language in schools and across the entire territory of a State) as well as multicultural policies that explicitly recognize cultural differences (regional autonomy, the devolution of powers, or the use of multiple languages in schools and in other contexts) have been often proposed as sources of conflict management in ethnically divided societies.

To date, only limited research has been carried out to examine the extent to which individual identity can be shifted by cultural policies and regulation. We know surprisingly little about whether government policies in general, and education systems in particular, can enhance the national cohesion and revitalize the national sentiment of a country.

Our paper takes a step in this direction by analyzing the effect of a particular education policy that took place in Spain on the process of identity formation. Up to 1983 Spanish was the official language of the Catalan education system. That year the education system became bilingual, and Catalan, together with Spanish, was taught in schools. Using survey data from Catalonia and exploiting within- and between-cohort variation in the exposure to the Catalan language at school, results show that individuals who experienced greater exposure to teaching in Catalan are more likely to say that they feel more Catalan than Spanish. Interestingly, the effect also appears

[2]Ramirez and Boli (1987) study extensively the political origins of mass schooling.

to be present among individuals educated in Catalonia after the reform but whose parents do not have Catalan origins: education, through language, can balance out the role of family in the cultural transmission process.

This reform can be interpreted as an example of a multicultural policy within Spain, where individuals living in Catalonia are the relevant minority and, at the same time, as an example of a nation-building policy within Catalonia, where individuals living in Catalonia who were born elsewhere or whose parents do not have Catalan origins are the relevant minority. As a result, nation-building policies and multicultural policies deliver very different outcomes in terms of the effect on individual sentiments: multicultural policies tend to favour the development of regional identities while nation-building policies tend to promote a common national feeling.

As a second step, we study the effect of the linguistic reform on political behaviour. Since individual identity is likely to affect thepolitical choices of individuals, the language education reform could foster attachment to local institutions and, in ethnically divided societies, it could increase the vote share of ethnic parties. We find that individuals who have experienced greater exposure to teaching in Catalan are more likely to declare that they voted during the 1999 regional elections and that they chose a party with a Catalanist (i.e Catalan regionalist) platform. The introduction of a bilingual education system increased the salience of the ethnic issue in the Catalan society and helped consolidate a political system organized along ethnic lines.

Our empirical argument relies on within- and between-cohort variation in exposure to Catalan language at school. The 1983 education reform affected younger cohorts more than the older generations. Within a given cohort, the exposure to Catalan varies according to the number of years they remained in education (some students left school earlier than others).

A number of factors leads us to believe that the link between the language reform and individual identity is causal. First, a number of robustness checks allow us to conclude that the results are not driven by education-specific cohort trends on Catalan identity. We also check the robustness of the results to various events that took place relatively close to the reform and that might have influenced catalan identity, such as the end of the Franco's dictatorship and the diffusion of Catalan-language media. Second, we can exclude the

idea that the reform changed the composition of the Catalan population, as comparisons between migration patterns in Catalonia and other Spanish regions show that Catalonia do not present any anomaly. Finally, we check that the results are not driven by the fact that individual years of schooling are an individual and parental choice. An identification strategy relying on the number of years for which students were taught in Catalan within compulsory education, suggests that this is not the case.

Our findings can, potentially, be of general interest and not only related to the Catalan society. The number of languages spoken is estimated to be between 6000 and 7000 and there are more than 20 States with more than one official language. Several countries, in addition to the official language, recognize other national languages, often compulsory in education, and in many other countries, a variety of languages are widely used without having the legal authority of an official language[3]. The choice of a language over an other has been often an important and divisive issue. For example, in 1956, in Sri Lanka, Sinhalese was made the official language provoking the strong reaction by the Tamil minority and until 1994 in Turkey the use of the Kurdish language in public was prohibited by law.

Multilingual societies have often adopted more than one language of instruction. Canada and India have an extensive experience in multilingual education; for four decades in India the languages of instructions were at least 3, the official language of each Indian state together with English and Hindi. In Latin American indigenous groups receive instruction in their own language and in the official language of the country, while in most Sub-Saharan African countries children are educated in their local languages during the very first grades and in the colonial language (French, English or Portuguese) during the further stages of their education.

This paper is related to a growing economic literature on the identity-formation process. In an influential series of papers, Akerlof and Kranton (2000, 2002, 2005) and Akerlof (2007) have incorporated identity, a person's sense of self, into an economic model of behavior. They proposed a utility function that depends on the individual's assigned or chosen social category, on the match between (exogenous) "prescriptions" for that category and the

[3]See UNESCO (2003) and the UN report (2004) for a discussion of the topic.

individual's given characteristics and behavior, and on his and others' actions. They have then presented several applications of their theory in the fields of the Economics of Education, the Economics of Organizations and Macroeconomics. Several theoretical studies followed the work by Akerlof and Kranton, including that by Benabou and Tirole (2007), that endogenizes identity payoffs and categorical prescription. Bisin et al. (2006). Charness, Rigotti and Rustichini (2006), Mc Leish and Oxoby (2007) and Eckel and Grossman (2005) provide laboratory experiments that show how group identity affects individual behavior. However, the number of empirical studies which study both the determinants of individual identity and how a person's sense of self affects individual behavior remains limited. In particular, there are not studies that analyze how policies affect identity. Our paper contributes to this literature by analyzing the effect of an educational policy on identity.

The paper is connected to a second field of literature, on the mechanisms of cultural transmission and on the effects of culture on economic outcomes[4]. If we consider identity as part of our cultural background, we show that education affects individual cultural attributes and also that the effect of school is comparable in magnitude with the effect of family during the cultural transmission process.

We also find connections with the literature on endogenous preferences. This literature[5] studies how political and economic institutions shape preferences through their effects on social norms, cultural transmission or through other channels. If identity is part of the utility function, our paper isolates a particular institutional arrangement (bilingual education) that is able to influence individual preferences.

This work is linked to a panoply of papers that study the relationship between ethnic diversity and economic and political outcomes. Easterly and Levine (1997) claimed that high levels of ethnic fragmentation were at the root of Africa's growth tragedy. La Porta, Lopez de Silanes, Shleifer and Vishny (1999) showed that ethnic heterogeneity is negatively correlated with

[4]See for istance Bisin and Verdier (2000), Guiso, Sapienza and Zingales (2007), Fernandez, Fogli and Olivetti (2005)

[5]See Bowles (1998) for a review of the topic.

the quality of government and, more recently, Montalvo and Reynal (2005) and (2006) explored the channels through which ethnic diversity influences economic development. They found that ethnic fractionalization lowers the rate of investment, while ethnic polarization increases the probability of civil wars. In contrast to those works, our paper tries to understand the roots of the individual ethnic identification[6].

Finally, our study is related to the literature on the determinants of turnout and political behavior (see Merlo 2006 for a review of the topic) and, in particular, to that which focuses on the relationship between education and individual political choices. Using compulsory education laws as instruments, Milligan et al. (2004) found a robust positive relationship between education and turnout for the United States. However, this was not the case for the United Kingdom. In this paper, we focus on the relationship between schooling and the political behavior of citizens in ethnically divided societies.

The structure of the paper is as follows: Section 2 provides a brief description of the reform and more generally of the institutional setting in Catalonia before and after 1983**.** Then the data and the empirical strategy employed are discussed. Section 3 presents the basic empirical evidence and Section 4 provides several robustness checks. Section 5 provides evidence on the heterogeneous effects of the reform. Section 6 provides evidence on the link between the reform and political choices. Finally, Section 7 concludes.

# 2 Background and empirical strategy

## 2.1 Background

Catalonia is a region in the north-east of Spain. Catalan is a language that evolved from vulgar Latin in the 9th century on both sides of the eastern part of the Pyrenees. Its territorial expansion went hand in hand with the expansion of the Catalano-Aragonese Crown, which was established in other Spanish regions like Valencia and the Balearic Islands, the south of France

[6]Miguel and Posner (2005) and Masella (2007) study the relationship between ethnic diversity and ethnic and national identity.

and the town of Alghero (in Sardinia, Italy). These are the regions where Catalan is present nowadays with more than 9 million speakers, making it one of the most important regional languages of the European Union. Catalonia is the region that hosts most of them, with more than 6 million speakers.

Catalan has not always been the main language in schools in Catalonia. From the middle of the twentieth century, when education became compulsory, it stopped being taught and all education was received in Spanish. In fact, Catalan was only seldom used in schools until 1978. During Franco dictatorship (1940-1975) it was banned and the Spanish language was the only one used as an official language.

The democratization process that took place in Spain during the late seventies led to the approval of a Constitution that, within a decentralized government structure, recognized the existence of a set of autonomous communities and allowed them to have their own regional parliaments. One of the most important laws approved by the new Catalan Government was the "Catalan law of language normalization (1983)" whose main goal was to promote the use of the Catalan language.

In one of its most relevant parts, the new law allowed the educational system (previously all in Spanish as an inheritance from Franco period) to become bilingual. Altough schools had been able to teach Catalan as a subject since 1978, until 1983 it was not recognized as the main language to be used in education. In addition, according to this law all students, irrespective of their origin, should be able to use both Catalan and Spanish at the end of their education, in fact- it made clear that- the Certificate of Basic Educational Attainment could not be achieved without proving proficiency in the official languages of Catalonia (Catalan and Spanish). An important feature of the law is that it stated that the language used in the education system could not separate pupils according to language differences.nd Catalan would be progressively used as students learn it.

Since the Catalan Education system had been Spanish based for such a long time period, the transition to a bilingual system was designed to be smooth. As stated in the "Order of Application of the Catalan Law of Normalization (1983)", for the first 4 years of primary education the presence of Catalan had to be smoothly increased, and it could not impede the normal

learning process of students already enrolled. For the subsequent 4 years of primary education the presence of the Catalan Language in the education system was initially introduced in social and natural science courses, and was increased in the following years. For secondary education, the Order again emphasizes that the increased presence of the Catalan language in each school had to be done in accordance with the students' prior knowledge of Catalan, in order to minimize its effect on the normal learning process.

The Catalan Law of Normalization, however, involved other changes, as the goal of the law was to increase the use of the Catalan language and make it an effective communication vehicle. In fact, it clearly established that citizens had the right to use Catalan, and that the Public Administration in Catalonia had to use both Catalan and Spanish, both in laws and documents and when communicating with citizens. It also established that the names of places had to be written in Catalan and encouraged its use by the media.

As it will be clear later on in the paper, in order to affect our identification strategy such changes should have a stronger impact on today feelings of younger and more educated respondents. We use data from the Basque Country, where similar reforms were implemented, and information on the current exposure to the media using the Catalan language, to check whether this is the case.

## 2.2 Data and descriptive statistics

The empirical analysis uses representative survey data on the institutional attachment of residents in Catalonia to the Spanish State that was published by "Centro de Investigaciones Sociologicas" in 2001.

To identify individuals' national attachment to Spain or Catalonia, we rely on the following question: "With which of the following sentences do you identify better? (1) I feel only Spanish, (2) I feel more Spanish than Catalan, (3) I feel as Spanish as Catalan, (4) I feel more Catalan than Spanish, (5) I feel only Catalan". That is, the closer the number is to 5, the more Catalan and less Spanish you feel, and vice versa.

The treated cohorts are those exposed, or potentially exposed to more years of teaching in Catalan, while the control cohorts are those that have

not been exposed to it. Thus, *exposure* depends on whether an individual was in primary or secondary education after the law was implemented in 1983, which depends on their year of birth and the number of years of education. In Spain, primary education lasts 8 years and starts at the age of 6, while secondary education lasts 4 years. An individual from the 1966 cohort who completed secondary education will have received one year of treatment, while an individual from the same cohort, but with only primary education, will have no treatment at all. Similarly, an individual from the 1970 cohort who completed secondary education will have received 5 years of treatment, while an individual from the same cohort but with only primary education will have received just 1 year of treatment. This means that the treated cohorts will be those born between 1966 and 1983, while the control cohorts will be those born between 1948 and 1965.

Figure 1 describes the relationship between cohorts, years of schooling and years of treatment. There are 3 major educational categories: (i) individuals that completed only primary education, who received 8 years of education; (ii) individuals that started but did not complete secondary education or that received some professional training,whom we assume received 10 years of education; and (iii) individuals that completed secondary education or a higher level of professional training, who went through 12 years of education[7].

Figures 2-6 show the proportions of individuals by cohort and educational level that answered the aforementioned question. As can be seen in Figure 3, there is an increase in the fraction of individuals who answer that they feel only Catalan for the cohorts born after 1966 with 12 years of education. For those individuals with less than 10 years of education the increase is for cohorts born after 1970. Furthermore, the difference between the two educational groups increases with the reform. Similar conclusions can be drawn from Figures 3-6: the reform stops the downward trend in individuals who say they feel more Catalan and it coincides with a decrease in the fraction of individuals who feel as Spanish as Catalan and with a mild decrease in the fraction of individuals who feel only Spanish.

Table 1 provides the full set of descriptive statistics for the main variables

[7]We do not consider universitary education since the law we study did not affect it. Thus, individuals with university education are included in the 12 years of education group.

used, by education and by the samples used in the main regressions.

## 2.3 Empirical strategy

Teaching in Catalan started at the beginning of the academic year 1983-1984. Students' exposure to the Catalan language reform varies according to the number of years they have been in education after the reform was implemented. Cohorts who started their primary education on or after that academic year were completely affected, while cohorts who started primary education prior to this date but were still in primary or secondary education *during or* after 1983-1984 were only partially affected by the reform.

The effect of the Catalan law is identified using a difference-in-differences approach that exploits between-cohort variation in Catalan instruction and within-cohort variation in years of education. The identification strategy used in this paper is similar to that used in Angrist and Lavy (1997) and Angrist et al (2006).

The first econometric specification to be tested is:

$$y_{ijm} = \alpha + \beta L_{ijm} + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \quad (1)$$

where the dependent variable $y_{ijm}$ indicates whether individual i, from cohort j and with m years of schooling answers 1,2,3,4 or 5 to the identification question.

$L_{ijm}$ is the length of exposure to teaching in Catalan, $\gamma_j$ are cohort dummies , $\delta_m$ is a dummy for the educational background and $X_{ijm}$ is a vector of individual-level control variables, including gender, the individual's origin, parent's origin and location dummies[8]. We run this specification, first as a linear regression and then as an ordered logit, due to the ordinal nature of the dependent variable[9]. Robust standard errors are clustered at the cohort-years of education level, to control for the fact that observations in a given

[8] We exclude from the sample all migrants who received some of their education outside Catalonia.

[9] Results for ordered probit are very similar and avalaible on request. Similarly, results for a multinomial logit and results treating the dependent variable as 5 different dummies are also very similar and available on request.

cohort-years of education group may be correlated[10].

Our identification strategy relies on the *implicit assumption* that there is no other variable that affects how certain cohort-years of education groups feel about the Catalan identity issue. If our results capture the effect of *omitted variables or of some other policies*, *these variables* should affect people who had some years of primary and secondary education after 1983 and it should affecte those individuals' identity more the longer they had remained in education. As previously mentioned, other parts of the law tried to enhance the use of Catalan in other public *spheres*; however, apriori, their effect should not be related with the level of the education of the respondents. In another section of the paper we discuss the fact that the years of education and the composition of our sample might not exogenous to the reform and we rule out the possibility that our estimates are simply driven by education-specific cohort trends in Catalan feelings or events that took place relatively close to year of the reform.

### 2.3.1 Did the reform increase the amount of education in Catalan?

The reform was expected to increase the amount of Catalan taught in schools and to also increase the likelihood that affected individuals use Catalan as standard language of expression. As a first step, in this section we investigate whether this is actually the case. The change in the language of instruction can reinforce Catalan identity- both because Catalan becomes the language used when learning and because Catalan becomes the language used by individuals.

In order to check the validity of our hypothesis, we take advantage of several questions asked during the survey. First, individuals were asked to classify how much of their education was in Catalan and how much in Spanish. We then create an index from 0 to 1, with 0 corresponding to education only in Spanish and 1 to education only in Catalan. In order to have a measure of the length of exposure of each respondent to Catalan teaching, we build a variable "Language exposure" that is the product of this index and the number of years of education.

[10] The coefficient of interest is still significant if we cluster at cohort level.

Second, in the survey, there are also several questions that provide information about the use of the Catalan language in everyday life: 1) with friends, 2) at home, 3) shopping, 4) being asked for directions in the street, 5) answering the phone, 6) interacting with civil servants 7) or with colleagues at work. We then create an index of the "social use" of Catalan, from 0 to 7, where 7 indicates that the interviewee uses the Catalan language in all of these circumstances.

In Table 2 we run OLS regressions in which we estimate the effect of the number of years each individual has been exposed to the reform on the language exposure variable and the language use variable. Results show the positive impact that the reform had on the number of years of education taught in Catalan, confirming that the reform was actually effective. In addition, the reform also positively affected the extent to which Catalan is used in everyday life.

# 3 Baseline Results

## 3.1 Difference-in-Difference Estimates

We now turn to the empirical evidence. Table 3 shows estimates of equation (1) using the sample born in 1948-83. The first and second columns (the OLS and the ordered logit specifications, respectively) indicates that, after conditioning on year of birth and years of education dummies, an increase in exposure to the reform significantly increases the probability that an individual feels more Catalan. When we include province of residence dummies and control for gender, family and individual origin (columns 3 and 4, the OLS and the ordered logit specifications, respectively) the coefficient is still positive and significant and also larger in size.[11].The intensity of Catalan identity is, as predictable, correlated with the origin of the interviewee and with that of his/her parents. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in

[11] Results do not change if we include controls for the profession and the socio economic status of the respondent.

Catalonia but with only one parent born in Catalonia 4) individuals who were born in Catalonia and whose parents were both born in Catalonia. We find that Catalan identity is strongest among respondents who were born in Catalonia whose parents were both born in Catalonia and it is weakest among respondents who were not born in Catalonia.

In order to interpret the ordered logit coefficient reported in column 4 of table 3, Figure 7 plots how the predicted probability of each answer changes with the intensity of the reform, i.e., the number of years of Catalan instruction. On the one hand, while the probability that an individual with 0 years of treatment says "I am only Catalan" is less than 0.1, the probability that a respondent with full treatment says "I am only Catalan" is almost 0.5. On the other hand, the probability that an individual says " I am as Spanish as Catalan" appears to strongly decrease with years of treatment (from more than 0.5 with 0 treatment to just over 0.1 with 12 years of treatment). The probability of choosing "I am only Spanish" and "I am more Spanish than Catalan" decrease slightly, while the probability of choosing "I am more Catalan than Spanish" shows a small increase with the length of exposure to the reform. Thus, Figure 8 suggests that, as a consequence of the reform, individuals are less likely to answer that they feel as Spanish as Catalan and more likely to say that they feel only Catalan.

To gauge the size of the impact of the introducation of bilingualism in Catalan schools, we then use the estimates in column 4 of table 3 to calculate the predicted identity (i.e., the predicted probability of choosing each of the five answers) of four different class of individuals: 1) respondents with 0 years of treatment who were born in Catalonia and whose parents were both born in Catalonia 2) respondents with 12 years of treatment (full treatment) who were born in Catalonia and whose parents were both born in Catalonia 3) respondents with 0 years of treatment who were born in Catalonia and whose parents were both born outside Catalonia 4) respondents with 12 years of treatment (full treatment) who were born in Catalonia and whose parents were both born outside Catalonia. Table 4 shows that respondents with 0 years of treatment who were born in Catalonia and whose parents were both born in Catalonia are likely to report very similar identity feelings to respondents with 12 years of treatment (full treatment) who were born in

Catalonia but whose parents were both born outside Catalonia.

This allows us to conclude that the educational reform has been able to balance out the intergenerational transmission of identity. In other words, school (through the language of instruction) and family turn out to produce a very similar effect on individual identity.

# 4 Robustness

## 4.1 Controlling for differential trends

The estimates in table 3 could be the artifact of education-specific cohort trends in Catalan feelings. Different cohorts might have been raised by parents with different values and preferences (the younger parents being more pro-Catalan), and, at the same time, parents with stronger Catalan preferences could have invested more in their children's education.

In order to investigate whether this is the case, and similarly to Angrist (2006), we propose the following two exercises. As a first step, we consider only cohorts who were not affected by the reform (cohorts born in 1930-65) and, among them, we assign a pseudo-treatment to the younger cohorts (cohorts born 1948-65)[12]. This pseudo-treatment consists of the length of exposure to the reform of individuals with the same level of education but born 18 years later. We then run the specification proposed by equation (1) using both an ordinary least squares procedure and an ordered logit one. If the estimates in Table 3 merely reflected the existence of differential trends we would expect the coefficient of the psuedo treatment variable similarly sized positive and significantly different from zero. Results from the experiment in columns 1 and 2 of Table 5 show there is no evidence of education-specific cohort trends in Catalan feelings; the coefficient of the pseudo treatment variable is very small and not significant.

As a second exercise, we focus on a triple differences identification strategy using earlier cohorts of Catalans (who never experienced a change in language of instruction) to adjust for differential trends. We consider cohorts born

[12] In other words, we assume the reform was implemented in 1966 rather then in 1983.

1921-80 and we use the following specification,

$$y_{ijm} = \alpha + \beta L_{ijm} + \theta l_{ijm} + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \quad (2)$$

where $L_{ijm}$ is a term that captures real exposure to teaching in Catalan and $l_{ijm}$ captures spurious effects.

$L_{ijm}$ are the years of real treatment for cohorts 1966-80; while $l_{ijm}$ equals years of pseudo treatment for cohorts 1936-50 (where now the pseudo-treatment consists of the intensity of the reform received by individuals with the same number of years of education but born 30 years later) and years of real treatment for cohorts 1966-1980. $\beta$ are the triple differences estimates, i.e. the treatment effects from the real experiment minus the pseudo treatment effects estimated using cohorts 1921-1950. Columns 3 and 4 report the results for this specification. Triple difference estimates are positive and significant, while the coefficient that would capture spurious trends is negative and not significant. We can conclude that, among cohorts affected by the reform, there are no education-specific cohort trends.[13]

## 4.2 Controlling for contemporaneous events

In this section we address confounding effects due to contemporaneous events that may have affected the feelings today of young and more educated Catalan residents. In particular the estimates in table 3 might be capturing the effect of the end of the dictatorship (Franco died in the 1975) and the beginning of the democracy;THIS IS THE CASE BECAUSE one, indeed, may think that cohorts not affected by the dictatorship feel more free today to express their feelings and declare their Catalan identity.

As a first test, we draw on Aspachs et al (2007) and consider another Spanish region: the Basque Country, where similar reforms and linguistic changes took place. In 1983 the education system became bilingual in the Basque region as well. However, whereas in Catalonia the reform was compulsory and homogenously spread across the Catalan population, in the Basque

[13]Because of lack of data about very old cohorts we could not use a larger number of cohorts, such as, for example, from 1909 to 1983

Country parents could choose the language used to educate their children[14]. The Basque and the Catalan languages are also very different: Basque is a non-Indo European language and it presents a completely different structure for native Spanish speakers. Moreover, the salience of the language issue is quite different in the two regions. While language is the main vehicle for identification within the Catalan group, this may not be the case within the Basque group. As a result, we are likely to find that the reform in the Basque Country had no effect on the identity of residents (or lower than in Catalonia). However, since the death of Franco is likely to have affected residents of Catalonia and the Basque Country in the same way and education-specific cohort trends are likely to follow similar patterns in both regions, if our results were driven by such factors, the coefficient of the variable of interest should be positive and significant also using the Basque sample. We use representative survey data on the institutional attachment of residents in the Basque Country to the Spanish State. This survey was also published by CIS in 2001 and is very similar to the survey conducted in Catalonia. The only difference is that the Basque survey was conducted one month before the Catalan one. When we apply the same empirical strategy to the Basque sample we find that the coefficient of the variable of interest (years of treatment) is not significantly different from zero. Results are shown in columns 1 and 2 of Table 6. Thus, the comparison between Catalonia and Basque Country helps us to exclude the possibility that the findings discussed in the previous sections are driven either by the existence of education-specific cohort trends or by the end of the Franco's dictatorship (or by any other event that took place in both regions and was contemporaneous with the reform)[15].

Since the middle of the eighties, a TV channel (TV3) has broadcast in the Catalan language and the diffusion of the daily press in Catalan has increased

[14] See Aspachs et al. (2007) for a more accurate discussion of the topic.

[15] Regressions in which we look at the effect of number of years of education after Franco's death and before the reform that each individual had show that this effect is not significant and small. The same is true when we introduce the original intensity variable in the regressions together with a variable that indicates the number of years of education after Franco's death for individuals that were not affected by the reform. The same conclusions can be drawn if, instead of considering Franco's death we consider the formation of the Catalan government at the end of 1977 as the relevant event. This is all available from the authors on request.

considerably; younger cohorts have been, in relative terms, more exposed to Catalan-language media. As a result, the between-cohorts variation we capture could be driven by media exposure rather than the length of exposure to Catalan instruction. Unfortunately we cannot observe the life-time exposure to Catalan-language media of respondents (which, however, is also likely to be affected by the 1983 reform) but we do have information on their current behaviour. We then construct 3 dummy variables that describe respectively whether or not the respondents today are inclined to (i) frequently watch Catalan language TV (ii) frequently read Catalan language press (iii) frequently listen to Catalan language radio stations. Table 6 shows results when we control for current exposure to Catalan language media: the coefficient of interest is still positive and significant (although smaller in size). As a second exercise, we check whether the effect of the 1983 reform varies with the current exposure to Catalan language media[16]. Table 7 shows that the effect of the reform is homogenous and independent of whether the respondent today watches Catalan language TV or reads Catalan language press or listens to Catalan language radio. Thus, the introduction of biligualism in Catalan schools seems to have an effect on Catalan feelings over and above the effect of exposure to Catalan media.

## 4.3 Controlling for the endogeneity of the level of education

The identification strategy previously described relies both on between-cohorts variation (older cohorts have not been affected by the reform, while the younger ones have been) and on within cohorts variation (respondents from the same cohort with a different level of education have been exposed to a different intensity of treatment). However, the number of years of schooling is typically an individual choice (or a parental one) and may have been affected by the introduction of the linguistic reform. One may think that respondents with very intense Spanish (and anti-Catalan) feelings, for instance migrants from another region of Spain, may feel less comfortable about receiving ed-

[16] We introduce interactions between our measure of exposure to the language reform and each of the 3 dummies previously described.

ucation in Catalan and then might have decided to drop out from school as a consequence of the reform.

As an initial check to confirm whether or not this is the case, we add as a control into the regression a dummy variable that is equal to one if the individual spoke Catalan at home with his or her parents. An individual who is used to speak Catalan is more likely to have a stronger Catalan identity and, at the same time, is less likely to have experienced difficulties learning Catalan at school and then dropped out from school after the reform. The results of this specification are shown in columns 1 and 2 of Table 9 and are very similar to those obtained previously. This is reassuring, as the language spoken at home could be the main source of endogeneity.[17]

As a second step, in order to avoid considering any within-cohort variation, we instrument $L_{i,m}$ with a measure of exposure to the reform that is not the result of an individual choice: the number of years of compulsory education for each individual[18]. Each respondent belonging to the same cohort will be subject to the same amount of exposure to compulsory education. This should provide an exogenous measure of years of education. Results are reported in Tables 8 and 9. As is predictable, the first stage of the 2SLS analysis (Table 8) shows a strong correlation between the variable of interest and its instrument. Column 3 reports the second stage; the coefficient is significant and slightly smaller than the one we obtained before. Finally, in columns 4 and 5 we report the results when we use the reduced form (this allows us to use an ordered logit model too). As before, results are consistent with our previous findings[19].

Then, as an additional test, we control for the cumulative distribution function for education using data from the Spanish Census 2001, i.e., for each individual, we compute the fraction of people with an equal, or lower, level of education in her cohort. We also control for the square of this variable.

[17] Results are robust to controlling for interactions between our proxy for the language spoken at home and cohort dummies.

[18] Schooling was previously compulsory only until the age of 12. This was then changed to 14 in 1970 with the General Educational Law. In 1990, with the approval of the Law for the General Order of the Educational System (LOGSE), the first two years of secondary education became compulsory as well, so students could leave education at the age of 16.

[19] These results include age and age squared. They are also robust to the inclusion of cubic and quartic trends. Results available from the authors on request.

This accounts for the possibility that the proportion of students who reach higher levels of education changes over time; any variation in educational attainment for treated cohorts should be captured by the introduction of the cumulative distribution function. Coefficients, shown in columns 6 and 7, remain positive and significant. Results in this section suggest that years of education are unlikely to have changed endogenously as a consequence of the reform. This is also supported by the fact that schools provided facilities and extra time for students for whom Catalan was not their mother tongue.

## 4.4 Controlling for the heterogeneity of the set of cohorts used

In our first specification, we were comparing individuals born in 1948 with those born in 1983. While this allows us to use the maximum data available, it may also introduce some noise. In order to mitigate this effect, we perform the same regression using smaller samples (the smallest includes only respondents born between 1961 and 1970). As the cohorts we are comparing are more similar, the results should be more reliable. Table 10 shows that, when we restrict the sample and compare cohorts that have similar age, results improve noticeably. This may be because we are comparing cohorts that are more similar among and that have been affected by similar shocks. This allows us to increase the precision of our estimates and to better capture the effects of the reform. Figure 8 plots the OLS coefficients, together with their 95% confidence interval of an exercise in which we progressivelly reduce the sample. As the sample is decreased, the coefficient corresponding to the intensity of the reform remains very stable and significant.

## 4.5 Controlling for the endogeneity of migration flows

The reform might have changed migration patterns into and out of Catalonia. Schooling in Catalan could have implied an additional cost of migration to Catalonia for Spanish speakers. This additional migration cost could have been higher for potential migrants with very intense Spanish sentiments who, as a result, could have decided not to migrate towards Catalonia. Similarly,

people with very intense Spanish sentiments could have decided to leave Catalonia after the reform. If this is the case, our results could be capturing a change in the composition of Catalan society rather than the effect of the reform.

As a first step, we check if our results hold only for individuals born in Catalonia. Since the youngest cohort was born in 1983, we restrict the sample to individuals whose parents were already in Catalonia when the reform was implemented. For these individuals, their parents' decision to migrate towards Catalonia is less likely to be affected by the linguistic reform. Columns 1 and 2 of Table 11 show that the results are robust to this check. However, it may still be the case that parents somehow anticipated the 1983 reform and then took the migration decision before 1983. To avoid this anticipation problem we restrict even further the sample and we consider only cohorts born in Catalonia between 1951 and 1978. Columns 3 and 4 of Table 11 show that results are robust to this robustness check.

Table 11 should be sufficient to convince that the results we find are not related to a change in the pattern of migration inflows in Catalonia. However, it can still be the case that patterns of migration outflows are affected by the reform and, at the same time, lie at the root of our results. Unfortunately the survey we have does not allow for any checks that could help us in this direction. Using "Residential Variation Data " (this is the only source of annual migration flows that goes back earlier than 1987), we provide some evidence that migration outflows were not affected by the introduction of bilingualism[20]. We consider the following specification,

$$Y_{i,t} = \alpha Cat_{i,t} + \beta After_{i,t} + \gamma Cat_{i,t} * After_{i,t} + \varepsilon_{it}$$

where $Y_{i,t}$ are the outflows (in per capita terms) of the region i in year t (from 1978 to 1987), $Cat_{i,t}$ is a Dummy variable equal to 1 if the observation refers to Catalan migration outflows and $After_{i,t}$ is a Dummy variable equal to 1 if the observation refers to years from 1983 onwards (so after the reform was implemented). In the first specification we consider all the Spanish

[20]Bover and Velilla (2001) discuss extensively migration trends in Spain in the 20th century. And we thank Olympia Bover for kindly sharing their data with us.

regions, in the second one only the 5 richest regions[21]; the coefficient of the interaction term (Colums 5 and 6 of Table 11) is always very close to zero and not significant. This suggests that per capita outflows in Catalonia were not affected by the linguistic law.[22].

## 4.6 Using other survey data

Altough the survey run in 2001 is clearly the best survey available in terms of number of observations and variety of questions (as described in previous sections of the paper, the survey contains also important informations on the current exposure to Catalan media and on the use of the Catalan language in the everyday life), the "Centro de Investigaciones Sociologicas" (CIS) provides also other surveys (one in 1998 and two in 2006, the first after the referendum on amending the 1979 Catalan Statute of Autonomy[23] and the second after the 2006 Catalan regional elections) that use a different set of individuals but include questions on identity, the origin of the respondents and of the parents of the respondents. The "Panel de Desigualtats socials a Catalunya (PaD)", run by a different agency, the "Fundacio' Jaume Bofill", also provides data on the identity feelings of respondents in three of the four waves available (2001-2002, 2004, 2005).

As a further robustness check we then consider the "Panel de Desigualtats socials a Catalunya (PaD)" and the other three surveys provided by the CIS. We adopt the same identification strategy used in the 2001 baseline survey[24]. Results in Table 12 show that the coefficient of our measure of exposure is positive and significant in each of the four different surveys analyzed[25]; the

[21] El Pais Vasco, La Rioja, Madrid, Navarra, Valencia.

[22] Migration patterns in and out of Catalonia changed the composition of the Catalan population. The fraction of respondents whose parents were both born in Catalonia could thus be larger among control cohorts than among treated cohorts, which could have biased downwards our results. In order to control for this, we add as controls the fraction of respondents of each origin by cohort-education group. Results are robust to this specification and available from the authors on request.

[23] The goal of the referendum was to further expand the authority of the Catalan government. It was approved by 73.24% of the census, and became effective as of August 9, 2006.

[24] In the case of the PaD we pool all the available waves together and we include survey-year dummies.

[25] *We also considered each of the three waves of the Pad separately. The coefficient of*

sizes of the coefficient are also quite similar to the one obtained using the baseline survey (roughly the same when we use the 1998 CIS survey).

# 5 Heterogeneous Effects

## 5.1 Effect for each level of treatment

The effect of treatment on identity can be very different for each level of treatment. In fact, the number of years of treatment can vary from 1 to 12. Those affected by one year will have been affected by the reform at the end of their primary or secondary education. Those affected by 8 years will have been affected, either during primary education, if they only completed primary, or during secondary education and the second half of primary education, provided they completed secondary education.

In this section we investigate how the effect increases with years of treatment. As a result, and similarly to Angrist et al. (2006), we introduce a dummy for each possible level of treatment and run the following specification:

$$y_{ijm} = \alpha + \sum_{k=1}^{12} \beta_k I(L_{jm} = k) + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \qquad (3)$$

where $I()$ is the indicator function. This specification is run both as an OLS and as an ordered logit model.

Columns 1 and 3 in Table 13 report ordinary least square estimates of the previous equation. In column 1, only dummies for years of education and cohort are included. In column 3, individual controls and province dummies are added to the regressions. Results in both columns go in the same direction. Identity is affected by the reform. Treatment effects do not increase monotonically with years of treatment, even if the effect seems to be stronger after 5 years of treatment. Ordered logit estimates for the same two specifications are provided in columns 2 and 4.

Figure 9 gives a clearer picture and confirms the patterns proposed in the previous section. In this figure we plot how a change from 0 to 1 in each of the

*our measure of exposure to catalan instruction is positive in each of the three waves and it is not significant at the standard level only in the last wave.*

treatment dummies affects the probability of each answer. This corresponds to the ordered logit regression in column 4. The probability of answering "I am only Catalan" increases with years of treatment, while the probability of answering " I am as Spanish as Catalan" appears to decrease strongly with years of treatment. The effect of the reform on the probability of choosing the other 3 categories (only Spanish/ more Spanish than Catalan/ more Catalan than Spanish) does not seem to be very relevant, as the lines fluctuate around zero.

## 5.2 Effect for each level of education

If the reform was applied gradually, or if all schools did not adopt the reform at the same time, our results would be downward biased, as we would be considering as exposed to the reform individuals that were not affected. This would imply that individuals who were exposed at the end of their primary education or at the end of their secondary or primary education would not be as affected as individuals who were exposed during their whole educational period. In addition, the effect of the reform will certainly vary if individuals were exposed for the first time to teaching in Catalan when they were at secondary education, at primary education or during both periods. For this reason, we have created four dummy variables: the first is equal to one if the individual were exposed in secondary education only, the second is equal to one if the individual were exposed only during the four years of primary education and did not continue in education, the third is equal to one if the individual were exposed for more than 4 years of primary education and did not continue in education and the last one that is equal to one if the individual were exposed both in secondary and primary education. The reference category will then be individuals who have not been exposed to the reform.

Results are shown in columns 1 and 2 of Panel A in Table 14. The OLS coefficients suggest that the earlier individuals have been exposed to the reform, the larger is the effect: the coefficients for individuals affected at half or more than half of primary education and at primary and secondary education are significant larger than that for individuals affected in secondary.

In addition, the coefficient for individuals affected both in secondary and primary is larger than that for individuals affected only in primary but more than four years, and larger than that for individuals affected less than four years of primary education, suggesting that the longer the period of exposure the greater the effect of the reform and that the earlier the individual is affected by the reform, the larger is the effect as well. Marginal effects for the ordered logit results are reported in Panel B and are consistent with the interpreation of the OLS results. As in the previous section, the effect of the reform was to mainly increase the probability of answering "only Catalan" and decreasing the probability of answering "as Spanish as Catalan".

## 5.3 Different Origins

The introduction of a Catalan-based schooling system can be interpreted as an example of a nation-building policy within the Catalan region. It is therefore interesting to check whether the reform has had an impact on the feelings of minorities living in Catalonia. We define as a minority group all of the respondents of non-Catalan origin or with parents without Catalan origin (we also distinguish between individuals from families with at least one parent born in Catalonia and individuals from families with neither of the parents born in Catalonia). We construct 4 dummy variables for these categories and interact them with $L_{ijm}$[26].

OLS results in column 3 of Panel A in Table 14 show that the reform had a positive and uniform impact on the whole population; the effect is positive not only on respondents of Catalan origin and whose parents were both born in Catalonia, but also on individuals born in Catalonia and from families with no parents born in Catalonia or with only one parent with Catalan origin. Marginal effects for the ordered logit in column 4 are reported in Panel B and are consistent with the OLS interpretation, even if coefficients for individuals not born in Catalonia are not always significant. Overall, however, results suggest that the new schooling system successfully increased the level of integration of minorities within Catalonia

[26] We also use other specifications: without cohort dummies, and with the instrumentation previously explained. All provide similar results. This is available from the authors upon request.

### 5.4 Language Spoken at Home

The effect of the reform could also vary according to the language individuals spoke at home when they were young, that is, at the time they were in education. In this section we analyze this issue by interacting the variable that captures the intensity of the reform with two dummy variables that indicate whether the individual spoke Catalan at home or not. Results are shown in columns 3 and 4 of Panel A in Table 14, for the OLS and ordered logit specifications. OLS results show how all individuals are affected by the reform, regardless of whether they spoke Catalan at home or not. However, the effect is slightly larger for those individuals who spoke Catalan at home. Marginal effects for the ordered logit specification are reported in Panel B and show a small decrease in the probability of answering "Only Spanish" or "More Spanish", a larger decrease in the probability of answering "As Spanish as Catalan" and an increase in the probability of answering "More Catalan" and "Only Catalan". However, the effect is always slightly larger for individuals who spoke Catalan at home.

## 6 The effect of the reform on political behavior

So far we have seen that the language reform affects identity, as people more affected by it tend to report that they feel more Catalan than Spanish. In this section we try to analyze whether this change in preferences also affected actions. In fact, we analyze whether voting decisions changed due to the reform. If the reform changed individuals' identity, this could have induced them to vote more in regional elections and to vote for political parties that push the Catalan identity issue.

The CIS 2001 survey includes a question in which individuals are asked who they voted for in the 1999 elections for the Catalan Parliament. Indivuals then answer which political party they voted for and whether they turned out to vote or not. This feature of the data will allow us to compare the voting behaviour of individuals affected by the reform with the voting behaviour of individuals not affected and then to perform a direct test on whether the

reform increased participation in Catalan elections and increased the salience of the ethnic issue.

In Catalonia, as basic democratic rights came to be recognized, the old political parties, which had existed in secrecy during the dictatorship, were legalized and new political parties also emerged. Of the five parties in the Catalan Parliament at the time of the survey, one was created during the transition: Convergència i Unió (CIU), the party that was in power from 1980 to 2003. The other four had existed, with their own variations, prior to the dictatorship: Partit dels Socialistes de Catalunya (PSC), Esquerra Republicana de Catalunya (ERC), Partido Popular (PP), and Iniciativa per Catalunya Verds (ICV). Three of these, CIU, ERC and ICV, are Catalan-only parties in the sense that they only present candidates in Catalonia, while the other two exist throughout the Spanish territory. In our wider classification of Catalan and non-Catalan parties, we consider as Catalan all parties that have in their programmes the approval of a law that would give Catalonia the right to self-determination. These parties are CIU, ERC and ICV. Then we use a narrower classification of the Catalan parties, in which we include only those parties that are in favour of an independent Catalonia: CIU and ERC.

Using the same identification strategy as in the previous part of the paper, we analyze the effect of the linguistic reform on political behaviour, namely turnout and party preferences using specification (1). We now use several dependent variables: $y_{ijm}$, indicates whether or not individual i, from cohort j and with m years of schooling:

(1) declares that he/she voted in the 1999 regional election. This specification checks if the linguistic reform stimulated turnout to the regional elections.

(2) declares that he/she voted for a Catalanist party in the 1999 regional election (we restrict the sample only to people who declared that they voted in that election). As a first step we consider IC, ERC and CIU as Catalanist parties, then look only at ERC and CIU. As a last step, we consider only CIU (the party in power when the reform was implemented). This specification checks whether the linguistic reform increased the share of votes for Catalanist parties in the regional elections.

(3) declares that he/she voted for a Catalanist party in the 1999 regional election (we do not restrict the sample). As in point (2), firstly we consider IC, ERC and CIU as Catalanist parties, then we look only at ERC and CIU and as a last step, we consider only CIU. This specification checks if the linguistic reform increased the total number of votes of Catalanist parties in regional elections.

As before, the identification strategy relies on the fact that there is no other variable that affects the political behavior of the cohort-years of education groups. We run this specification, first as a linear regression and then as a logit, due to the binomial nature of the dependent variable[27]. Robust standard errors are clustered at the cohort-years of education level, to control for the fact that observations in a given cohort-years of education group may be correlated.

Table 15 reports estimates for Equation (1) using as dependent variable a dummy equal to one if the respondent voted in the 1999 regional election. In Columns 1 and 2, we show estimates when we only control for years of education and year of birth dummies, for the OLS and the logit specifications, respectively. In Columns 3 and 4, we add province of residence dummies and control for gender, family and individual origin. Coefficients are always positive and significant. Results suggest that those who were more exposed to Catalan instruction tend to be significantly more likely to vote; in fact, an extra year of education in Catalan increases the likelihood that the respondent declared that he/she voted by 4.5%.

Tables 16 and 17 report estimates for equation (1) using as dependent variable a dummy equal to one if the respondent voted for a Catalanist party in the 1999 regional election, when we restrict the sample only to people who declared that they voted in that election (Table 13) and when we do not restrict the sample (Table 14). In Columns 1 and 2 of both tables we show estimates for the OLS and the logit specifications, when we only control for years of education and year of birth fixed effects and when we classify IC, ERC and CIU as Catalanist parties. In Columns 3 and 4 of both tables we consider only CIU and ERC as Catalanist parties. In Columns 5 and 6 of both tables we focus only on CIU. Results indicate that an increase in exposure to

[27] The results for probit are very similar and available on request.

the reform increases the probability that an individual votes for a Catalanist party; moreover most of the effects we find in Tables 16 and 17 seem to come from the increase in the probability of choosing CIU. In Columns 7-12 of both tables we add province of residence dummies and control for gender, family and individual origin. The coefficients are still positive and significant. As shown in Column 8 of Table 17 (restricted sample), an increase from 4 years of treatment (the average amount of treatment) to 8 years of treatment (the average amount of treatment plus one standard deviation) would increase the likelihood that the respondent declared that he/she voted for a party with Catalan affiliation by almost 20%. Column 8 of Table 17 (unrestricted sample) shows that an increase from 4 to 8 years of treatment would increase the likelihood that the respondent declared that he/she voted for a party with Catalan affiliation by 27%.[28]

# 7 Conclusions

Much has been said about the possibility that policies can affect individual identity, sentiments and preferences. However, to date, not many studies have been conducted on this matter. We have considered the 1983 educational reform, by which the Catalan education system became bilingual and Catalan, together with Spanish, was taught in schools and we have found a positive effect of this policy on Catalan identity. The size of the effect is large and extends to individuals whose parents have no Catalan origins. Education, through language, can balance out the role of family in the cultural transmission process. The Catalan case allows us also to compare multicultural policies and nation building policies. We interpret the reform as an example of multicultural policies within Spain and of nation building policies within Catalonia. We conclude that while multicultural policies stimulate re-

[28] Interestingly, we use a two stage least square specification as well. We use the level of Catalan identity as explanatory variable for political choices and we instrument identity with the lenght of exposure to the Catalan instruction. The coefficient of identity is positive and significant, suggesting that identity might truly be the channhel through which the reform affected political behavior. This specification, however, relies on the assumption that the change in the language of instruction affected political choices only through identity.

gional identities, nation building policies encourage the growth of a shared national sentiment. The survey used for Catalonia further allows us to show that the language reform increased turnout and voting for Catalan parties. The shift in identity may be one of the channels through which this takes place.

Nowadays most countries in the world could be classified as multilingual, however, not all of them have multilingual educational systems. The UNESCO report "Education in a Multiligual World" explains that the choice of language of education constitutes an important challenge in the development of educational policies. In addition, mother tongue instruction is considered important for the quality of the education provided. Language can be regarded, not only as a communication tool, but also an attribute of empowerment and cultural identity. Research of this sort together with the study of the effects of such reforms on political and economic outcomes can be highly relevant in ethnically divided societies where policies (linguistic and not) can be seen as mechanisms of integration and conflict reduction. This paper constitutes a first step in this direction, but more research and data from diferent countries is needed.

# References


Akerlof, G.A. (2007) "The Missing Motivation in Macroeconomics" *American Economic Review* 97(1), 5-36

Akerlof, G.A. and R.E. Kranton (2000), "Economics and identity" *Quarterly Journal of Economics* 115, 715-753

Akerlof, G.A. and R.E. Kranton (2002), "Identity and Schooling: Some Lessons for the Economics of Education" *Journal of Economic Literature,* 40(4), 1167-1201

Akerlof, G.A. and R.E. Kranton (2005), "Identity and Economics of Organization" *Journal of Economic Perspectives,* 19, 9-32

Anderson, B (1983). Imagined Communities: Reflections on the Origins and Spread of Nationalism, London, Verso.

Angrist, J., A. Chin and R. Godoy (2006) "Is Spanish-only schooling responsible for the Puerto Rican language gap?" *Journal of Development Economics* forthcoming

Angrist, J. and V. Lavy (1997) "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco" *Journal of Labor Economics,* 15, S48-S76 (January, Part 2)

Aspachs-Bracons, Oriol, Irma Clots-Figueras and Paolo Masella (2007a)."Identity and Language Policies" Universidad Carlos III Working Papers 07-77(46).

Aspachs-Bracons, Oriol, Irma Clots-Figueras and Paolo Masella (2007b)."Education and Political Behavior. Evidence from the Catalan Linguistic Reform" Universidad Carlos III Working Papers 07-78(47).

Aspachs, O., I. Clots J, Costa and P. Masella (2008) "Compulsory Language Educational Policies and Identity Formation" (with ) Journal of European Economic Association, 6 (2-3), April-May

Bates, R (1983) "Modernization, Ethnic competition and the rationality of Politics in Contemporary Africa" in *State versus Ethnic Claims: African Policy Dilemmas*

Bénabou, R. and Tirole, J. 2007. "Identity, Dignity and Taboos: Beliefs as Assets," *CEPR Discussion Papers 6123, C.E.P.R. Discussion Papers*

Bisin A and T. Verdier (2000) "The economics of cultural transmission and the dynamics of preferences" *Journal of Economic Theory* 97, 298-319

Bisin A, E. Patacchini , T. Verdier , Y. Zenou (2006) "'Bend it like Beckam': identity, socialization and assimilation" *CEPR Discussion paper* No 5662

Bover, O. and P Velilla (2001) "Migrations in Spain: Historical Background and Current Trends" *Banco de España Working Papers* No 9909

Charness, G Rigotti, L and Rustichini, A (2006)" Individual Behavior and Group Membership" forthcoming AER

Bowles, S (1978) "Capitalist Development and Educational Structure" *World development*, 6, 783-796.

Easterly, W and R. Levine (1997) "Africa's Growth Tragedy: Policies and Ethnic Divisions" *Quarterly Journal of Economics*, Vol. 112, Issue 4

Eckel, C.E. & Grossman, P.J. (2005). Managing diversity by creating team identity. Journal of Economic Behavior and Organization, 58(3), 371-392.

Fernandez,R , A Fogli and C. Olivettu "Marrying your mom: Preference transmission and Womens'Labor and Education choice" *NBER working paper*

Horowitz, D. (1985). Ethnic Groups in Conflict. *Berkeley: University of California Press.*

La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R. Vishny (1999)"The Quality of Government" *Journal of Law, Economics and Organization,* 222-279

Masella, P (2007)"National identity and ethnic diversity. Theory and cross-country evidence" *working paper*

Merlo, A (2007) "Whither Political Economy? Theories, Facts and Issues", in R.Blundell, W. Newey and T. Persson (eds.), Advances in Economics and Economet-rics, Theory and Applications: Ninth World Congress of the Econometric Society, Cambridge: Cambridge University Press. -

McLeish, K . and Oxoby, R . (2007). "Identity, Cooperation, and Punishment" IZA Discussion Paper No. 2572

Miguel T and Posner D. (2004) "Sources of ethnic identification in Africa" *working paper*

Milligan, K. , Moretti, E. and Oreopulos P. (2004) "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom" Journal of Political Economy 88, 1667-1695

Montalvo J. and Reynal M. (2005) "Ethnic diversity and economic development" *Journal of Development economics*, 76, 293-323

Montalvo J. and Reynal M. (2006) "Ethnic polarization, potential conflict and civil wars" *American Economic Review* 95 (3), 796-816

Ramirez F. and Boli J. (1987) "The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization" *Sociology of Education*, Vol. 60, No. 1. (Jan., 1987), pp. 2-17.

Reisner, E (1922) "Nationalism and Education since 1789" *New York, Macmillan*

UN report (2004) "Building multicultural democracy"

UNESCO (2003) "Education in a Multiligual World"

# 8 Appendix

## 8.1 Definition of the variables

-*Identity:* ordered variable which assumes the following values: (1) if the respondent answered "I feel only Spanish" (2) if the respondent answered "I feel more Spanish than Catalan" (3) if the respondent answered "I feel as Spanish as Catalan" (4) if the respondent answered "I feel more Catalan than Spanish" and (5) if the respondent answered "I feel only Catalan". Source: CIS

-*Years of education:* (i) if the respondent answered that he completed only primary education, we assume that he received 8 years of education; (ii) if the respondent answered that he started but did not complete secondary education or that he received some professional training, we assume that he received 10 years of education; and (iii) if the respondent answered that he completed secondary education or a higher level of professional training, we assume that he received 12 years of education. Source: CIS

-*Catalan origin, Catalan family:* dummy equal to 1 if the respondent answered that he was born in Catalonia and both his parents were born in Catalonia. Source: CIS

-*Catalan origin, mixed family:* dummy equal to 1 if the respondent answered that he was born in Catalonia and only one of his parents were born in Catalonia. Source: CIS

-*Catalan origin, non Catalan family:* dummy equal to 1 if the respondent answered that he was born in Catalonia and neither of his parents were born in Catalonia. Source: CIS

-*non Catalan origin:* dummy equal to 1 if the respondent answered that he was not born in Catalonia. Source: CIS

-*Language exposure:* We use an ordered variable which assumes the following values: (1) if the respondent answered that his education was only in Spanish (2) if the respondent answered that his education was more in Spanish than in Catalan (3) if the respondent answered that his education was half in Spanish half in Catalan (4) if the respondent answered that his education was more in Catalan than in Spanish (5) if the respondent answered that his education was only in Catalan. We then create an index that goes from 0 to 1, with 0 corresponding to education only in Spanish and 1

to education only in Catalan. "Language exposure" is the product between this index and years of education. Source: CIS

-*Catalan social use:* respondents were asked to provide information about the use of Catalan language 1) when they meet friends, 2) when they are at home, 3) when they go shopping, 4) if they are asked for directions in the street, 5) when they answer the phone, 6) when they interact with civic servants 7) or when they are with colleagues at work. We then create an index of the "social use" of Catalan, that goes from 0 to 7, where 7 indicates that the interviewee uses the Catalan language in all of these circumstances.

-*Language at home:* dummy equal to 1 if the individual spoke Catalan at home with his parents

-*Female:* dummy equal to 1 if the respondent is female

-*Voting:* dummy equal to (1), if the respondent declares to have voted in the 1999 regional election. Source: CIS

-*Catalan Voting (IC+ERC+CIU):* dummy equal to (1) if the respondent declares to have voted for a Catalanist party in the 1999 regional election. We consider IC, ERC and CIU as Catalanist parties Source: CIS

-*Catalan Voting (ERC+CIU):* dummy equal to (1) if the respondent declares to have voted for a Catalanist party in the 1999 regional election. We consider ERC and CIU as Catalanist parties Source: CIS

-*Catalan Voting (CIU):* dummy equal to (1) if the respondent declares to have voted for a Catalanist party in the 1999 regional election. We consider CIU as the only Catalanist parties Source: CIS

-*Catalan TV: :* dummy equal to 1 if the individual currently watches only Catalan TV. Source: CIS

-*Catalan radio: :* dummy equal to 1 if the individual currently lissens only to Catalan radio. Source: CIS

-*Catalan newspapers:* dummy equal to 1 if the individual currently reads only to Catalan newspapers. Source: CIS

-*outflows:* annual per capita outflows. Source: Residential Variation Data

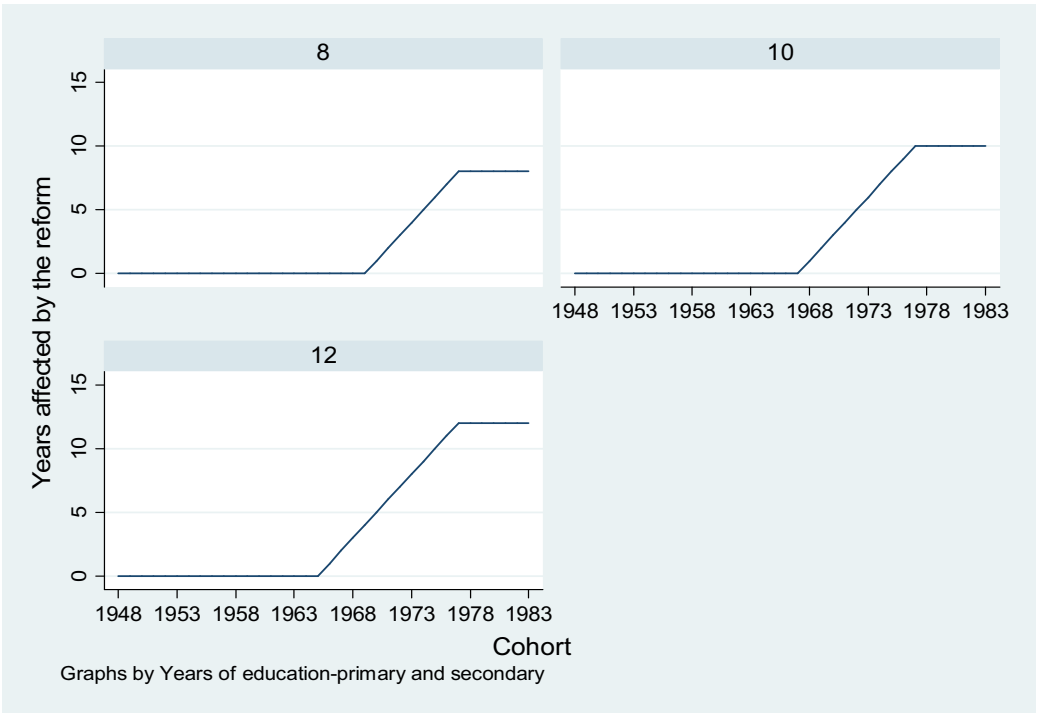
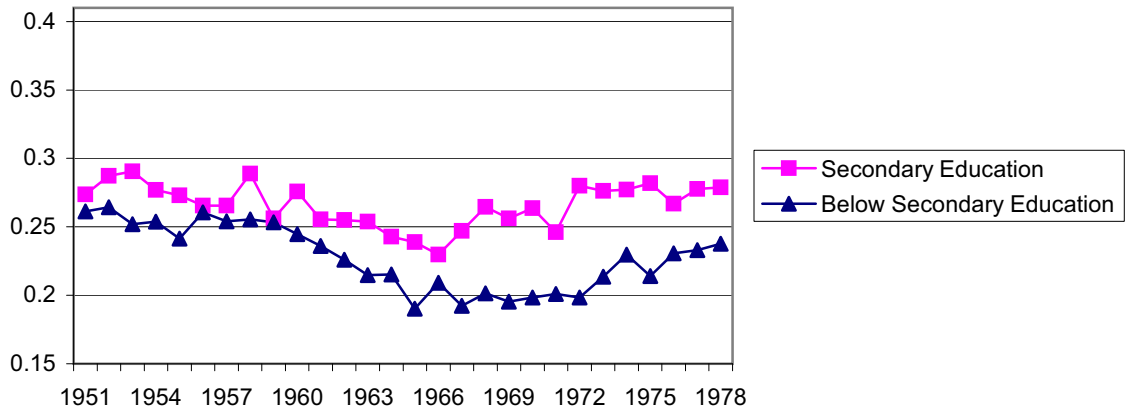


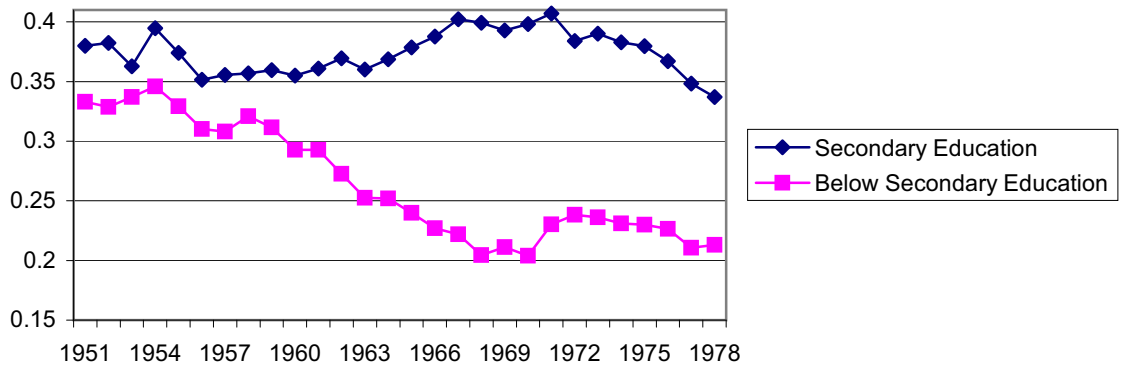

**Figure 1: Treatment by Cohort and Years of Education**

**Figure 2: Fraction of Catalans that feel only Catalan by level of education**



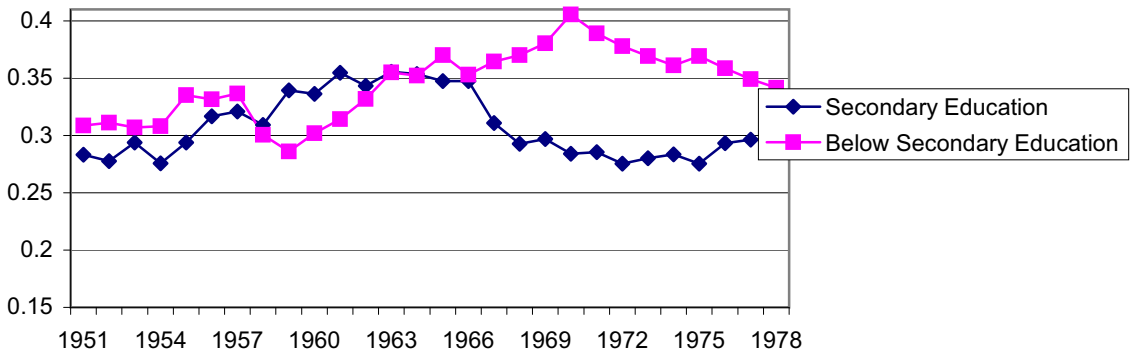

Note: The graph plots the fraction of respondents with and without secondary education who choose *I feel only Catalan* against year of birth. All series are (5,1,5) moving averages.

**Figure 3: Fraction of Catalans that feel more Catalan than Spanish by level of education**



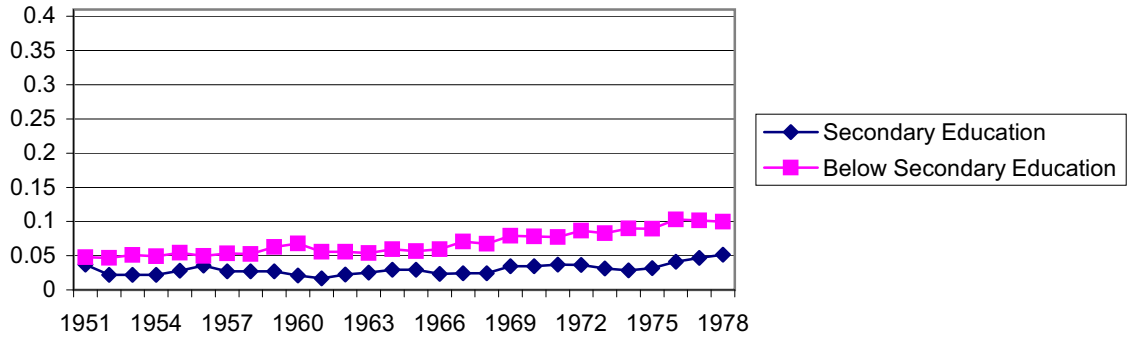

Note: The graph plots the fraction of respondents with and without secondary education who choose *I feel more Catalan than Spanish* against year of birth. All series are (5,1,5) moving averages.

**Figure 4: Fraction of catalans that feel as Catalan as Spanish by level of education**



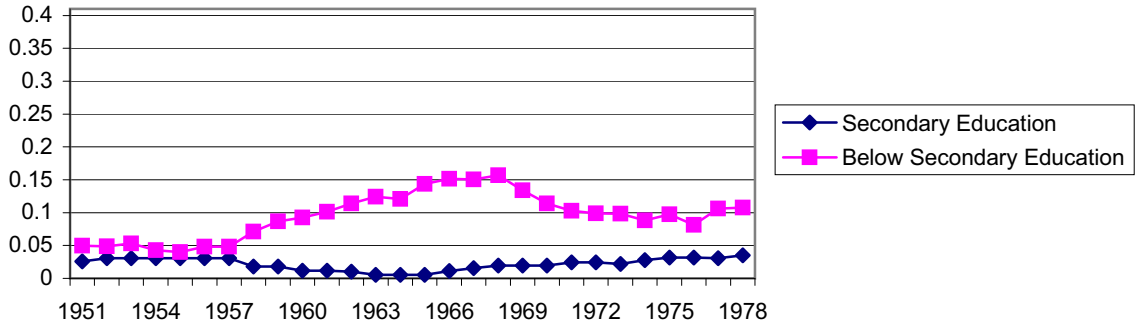

Note: The graph plots the fraction of respondents with and without secondary education who choose *I feel more Spanish than Catalan* against year of birth. All series are (5,1,5) moving averages.

**Figure 5: Fraction of Catalans that feel more Spanish than catalan by level of education**




Note: The graph plots the fraction of respondents with and without secondary education who choose *I feel as Spanish as Catalan* against year of birth. All series are (5,1,5) moving averages.

**Figure 6: Fraction of Catalans that feel only Spanish by level of education**




Note: The graph plots the fraction of respondents with and without secondary education who choose *I feel only Spanish* against year of birth. All series are (5,1,5) moving averages.

**TABLE 1**
**Descriptive stats (1467 obs )**

| Variable | mean | Std. Dev. |
|---|---|---|
| Only Spanish | 0.0579 | 0.2336 |
| More Spanish | 0.0549 | 0.2278 |
| As Spanish as Catalan | 0.3950 | 0.4890 |
| More Catalan | 0.3006 | 0.4587 |
| Only Catalan | 0.1916 | 0.3937 |
| Years Affected by Reform | 4.0120 | 4.6399 |
| No Catalan Origin | 0.1113 | 0.3146 |
| Mixed Prents | 0.1719 | 0.3774 |
| No Catalan parents | 0.2821 | 0.4502 |
| Years of education (primary and secondary) | 10.3231 | 1.8893 |
| Female | 0.4970 | 0.5002 |
| Age | 34.8099 | 9.7834 |
| Language at home | 0.4411 | 0.4967 |
| % of Catalan taught at School (1452 obs) | 0.2907 | 0.3392 |
| Age at Arrival | 0.2542 | 1.0075 |
| Years Affected by Reform (within compulsory) | 2.9259 | 3.6045 |
| | | |
| *Politics Sample (1248 electors, 941 voted)* | | |
| Turnout | 0.7721 | 0.4196 |
| Voted Catalan (CIU and ERC) voters | 0.5885 | 0.4924 |
| Voted Catalan (CIU, IC and ERC) voters | 0.6629 | 0.4730 |
| Voted Catalan (CIU and ERC) electors | 0.4451 | 0.4972 |
| Voted Catalan (CIU, IC and ERC) electors | 0.5013 | 0.5002 |

## Table 2: Did the reform increase the amount of education in Catalan?

| | Language exposure | Use of Catalan |
|---|---|---|
| | [1] | [2] |
| Intensity | .695*** | .179** |
| | (.08) | (.08) |
| INDIVIDUAL CONTROLS | YES | YES |
| YEARS OF EDUCATION | YES | YES |
| YEAR OF BIRTH | YES | YES |
| PROVINCES | YES | YES |
| No. obs. | 1464 | 1321 |
| Rsq | 0.524 | 0.519 |
| Pseudo Rsq | | |

Note: Individuals were asked to classify from 1 to 5 how much of their education was in Catalan and how much in Spanish; we then create an index from 0 to 1, with 0 corresponding to education only in Spanish and 1 to education only in Catalan. "Language exposure" is the product of this index and the years of education. We consider questions about the use of the Catalan language in everyday life: 1) when the interviewee meets friends, 2) when he is at home, 3) when he goes shopping, 4) if he is asked for directions in the street, 5) when he answers the phone, 6) when he interacts with civil servants 7) or when he is with colleagues at work. Use of Catalan is an index that goes from 0 to 7, where 7 indicates that the interviewee uses the Catalan language in all of these circumstances. All the specifications include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondent. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%

## Table 3: Baseline Results

*Dependent variable: Identity*

| | *OLS* [1] | *Ord. Logit* [2] | *OLS* [3] | *Ord. Logit* [4] |
|---|---|---|---|---|
| intensity | .067** | .147*** | .082*** | .183*** |
| | (.029) | (.055) | (.026) | (.058) |
| non Catalan origin | | | -1.335*** | -3.038*** |
| | | | (.1) | (.247) |
| mixed family | | | -.493*** | -1.074*** |
| | | | (.073) | (.16) |
| non Catalan family | | | -.958*** | -2.187*** |
| | | | (.072) | (.163) |
| YEARS OF EDUCATION | YES | YES | YES | YES |
| YEAR OF BIRTH | YES | YES | YES | YES |
| PROVINCES | NO | NO | YES | YES |
| No. obs. | 1467 | 1467 | 1467 | 1467 |
| Rsq | .088 | | .324 | |
| Pseudo-Rsq | | .032 | | .149 |

Note: The specifications reported in Columns 1 and 2 include years of education and year of birth fixed effects. The specifications reported in Columns 3 and 4 include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia (omitted category). Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%.

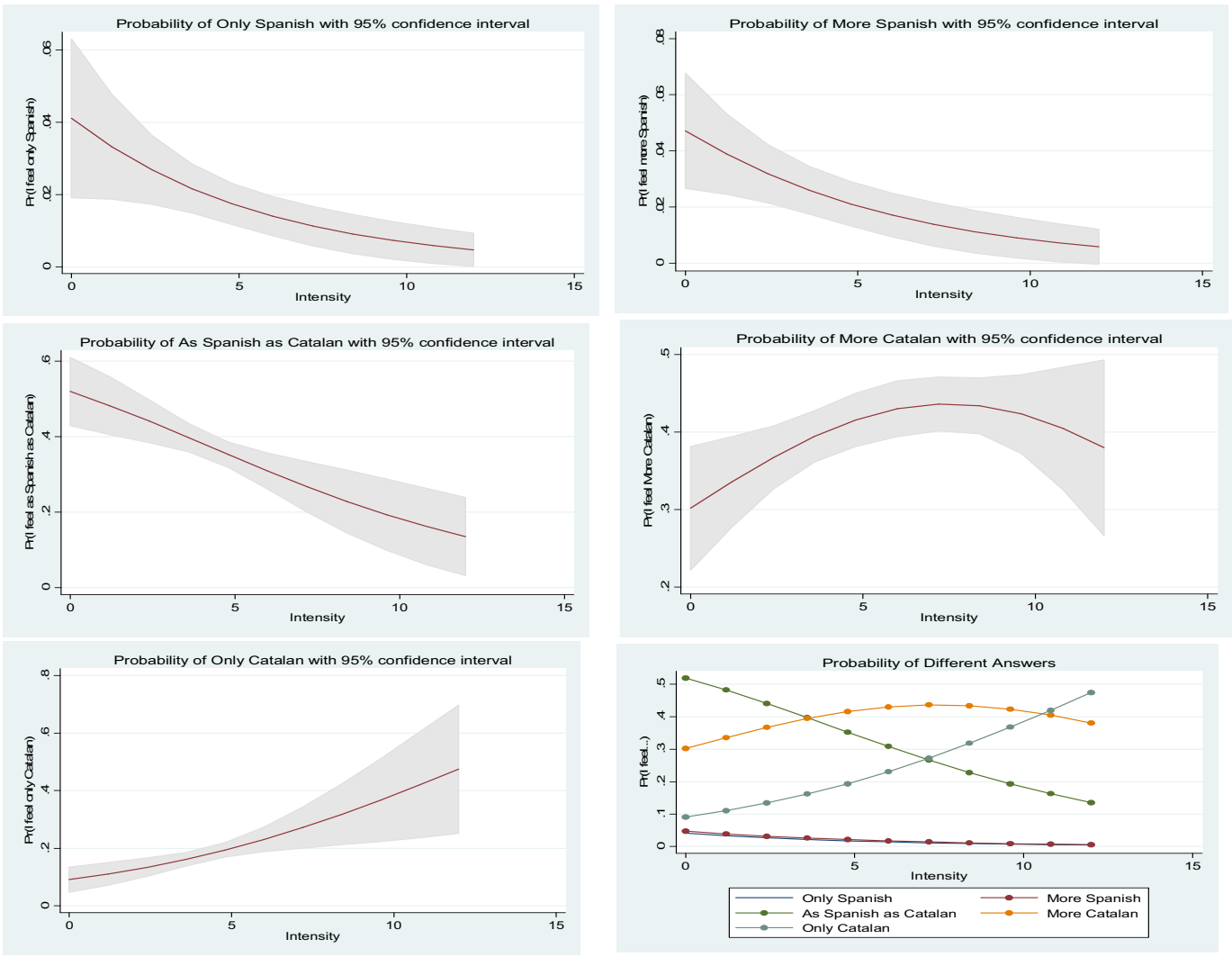



**Figure 7: Change in probabilities (specification 1)**

Note: we plot how the probability of each answer changes with the intensity of the reform, i.e. the number of years of Catalan instruction. This corresponds to the ordered logit regression in Column 4 of Table 3.

## Table 4: Baseline Results: Interpretation

*Predict the probability of each answer by Origin and intensity of treatment*

| | Cat Family *intensity=0* [1] | Cat family *intensity=12* [2] | No cat family *intensity=0* [3] | No cat family *intensity=12* [4] |
|---|---|---|---|---|
| Prob "Only Spanish" | 0.0162 | 0.0018 | 0.1276 | 0.0159 |
| Prob "More Spanish than Catalan" | 0.0196 | 0.0023 | 0.1207 | 0.0192 |
| Prob "As Spanish as Catalan" | 0.3367 | 0.0573 | 0.5926 | 0.3329 |
| Prob "More Catalan than Spanish" | 0.4216 | 0.2366 | 0.1307 | 0.4229 |
| Prob "Only Catalan" | 0.206 | 0.702 | 0.0283 | 0.2091 |

## Table 5: Controlling for cohort-education trends

*Dependent variable: Identity*

| | *OLS* [1] | *Ord. Logit* [2] | *OLS* [3] | *Ord. Logit* [4] |
|---|---|---|---|---|
| intensity | | | .115*** | .232** |
| | | | (.037) | (.93) |
| pseudo-intensity | -.015 | -.07 | -.022 | -.024 |
| | (.032) | (.075) | (.036) | (.094) |
| INDIVIDUAL CONTROLS | YES | YES | YES | YES |
| YEARS OF EDUCATION | YES | YES | YES | YES |
| YEAR OF BIRTH | YES | YES | YES | YES |
| PROVINCES | YES | YES | YES | YES |
| | 1930-1965 | 1930-1965 | 1921-1980 | 1921-1980 |
| No. obs. | 1008 | 1008 | 1860 | 1860 |
| Rsq | .291 | | .295 | |
| Pseudo Rsq | | .131 | | .135 |

Note: In the specifications reported in Columns 1 and 2 we consider only cohorts who are not affected by the reform (1930-1965) and we assign a pseudo-treatment to the younger cohorts (1948 - 1965). In the specifications reported in Columns 3 and 4 we consider cohorts born between 1921 and 1980 and the variable pseudo-intensity equals years of pseudo treatment for cohorts 1936-1950 and years of real treatment for cohorts 1966-1980. All the specifications include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%

**Table 6: Controlling for contemporaneous events (1)**

*Dependent variable: Identity*

| | *OLS* [1] | *Ord. Logit* [2] | *OLS* [3] | *Ord. Logit* [4] | *OLS* [5] | *Ord. Logit* [6] | *OLS* [7] | *Ord. Logit* [8] | *OLS* [9] | *Ord. Logit* [10] |
|---|---|---|---|---|---|---|---|---|---|---|
| intensity | .004 | .021 | .056** | .124** | .064*** | 0.143*** | .066*** | 0.154*** | .047* | .106* |
| | (.025) | (.054) | (.025) | (.059) | (.024) | (.057) | (.024) | (.057) | (.024) | (.058) |
| Catalan TV | | | .634*** | 1.54*** | | | | | .692*** | 1.089*** |
| | | | (.056) | (.157) | | | | | (.174) | (.161) |
| Catalan radio | | | | | .554*** | 1.335*** | | | .253*** | .667*** |
| | | | | | (.045) | (.118) | | | (.054) | (.152) |
| Catalan press | | | | | | | .547*** | 1.278*** | 0.271 | .692*** |
| | | | | | | | (.054) | (.014) | (.065) | (.174) |
| INDIVIDUAL CONTROLS | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| YEARS OF EDUCATION | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| YEAR OF BIRTH | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| PROVINCES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| | Basque | Basque | Catalonia | Catalonia | Catalonia | Catalonia | Catalonia | Catalonia | Catalonia | Catalonia |
| No. obs. | 1260 | 1260 | 1467 | 1467 | 1467 | 1467 | 1467 | 1467 | 1467 | 1467 |
| Rsq | .39 | | .295 | | .371 | | .365 | | .413 | |
| Pseudo Rsq | | .106 | | .189 | | .178 | | .173 | | .205 |

Note: In Columns 1 and 2, we use the standard identification strategy, but a different dataset: we use representative survey data on the institutional attachment of residents in the Basque Country to the Spanish State.In Columns 3 and 4 we use the Catalan sample and include as control a Dummy variable equal to 1 if the respondents today are inclined to frequently watch Catalan language TV. In Columns 5 and 6 we include as control a Dummy variable equal to 1 if the respondents today are inclined to frequently read Catalan language press . In Columns 7 and 8 we include as control a Dummy variable equal to 1 if the respondents today are inclined to frequently listen to Catalan language radio stations. In Columns 9 and 10 we control for current exposure to all the possible media sources. All the specifications include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%

## Table 7: Controlling for contemporaneous events (2)

*Dependent variable: Identity*

| | *OLS* [1] | *Ord. Logit* [2] | *OLS* [3] | *Ord. Logit* [4] | *OLS* [5] | *Ord. Logit* [6] |
|---|---|---|---|---|---|---|
| intensity*Catalan TV | .067** | .157** | | | | |
| | (.027) | (.063) | | | | |
| intensity*no Catalan TV | .05* | .111* | | | | |
| | (.025) | (.06) | | | | |
| intensity*Catalan radio | | | .066*** | .154*** | | |
| | | | (.025) | (.059) | | |
| intensity*no Catalan radio | | | .063** | .137** | | |
| | | | (.025) | (.059) | | |
| intensity*Catalan press | | | | | .065** | .157*** |
| | | | | | (.026) | (.061) |
| intensity*no Catalan press | | | | | .067*** | .152*** |
| | | | | | (.025) | (.058) |
| No. obs. | 1467 | 1467 | 1467 | 1467 | 1467 | 1467 |
| Rsq | .391 | | .371 | | .365 | |
| Pseudo Rsq | | .19 | | .178 | | .173 |

Note: In Columns 1 and 2 we use the Catalan sample and include as controls a Dummy variable equal to 1 if the respondents today are inclined to frequently watch Catalan language TV and an interaction between this Dummy and the variable *intensity*. In Columns 3 and 4 we include as controls a Dummy variable equal to 1 if the respondents today are inclined to frequently read Catalan language press and an interaction between this Dummy and the variable *intensity*. In Columns 5 and 6 we include as controls a Dummy variable equal to 1 if the respondents today are inclined to frequently listen to Catalan language radio stations and an interaction between this Dummy and the variable *intensity*. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%

## Table 8: First stage

*Dependent variable: Intensity*

| | *OLS* [1] |
|---|---|
| Intensity compulsory | 1.058*** |
| | (.028) |
| non Catalan origin | .084 |
| | (.116) |
| Catalan origin mixed family | -.108 |
| | (.089) |
| Catalan origin non catalan family | -.122 |
| | (.086) |
| YEARS OF EDUCATION | YES |
| YEAR OF BIRTH | YES |
| PROVINCES | YES |
| AGE and AGE SQUARED | YES |
| No. obs. | 1467 |
| Rsq | .951 |

Note: The dependent variable is the number of years of education under the reform. The specification includes the number of years of compulsory education under the reform, years of education fixed effects, province fixed effects, the age and the age squared of the respondent and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%

## Table 9: Controlling for endogenity of years of education

*Dependent variable: Identity*

| | *OLS* [1] | *Ord. Logit* [2] | *2SLS* [3] | *OLS* [4] | *Ord. Logit* [5] | *OLS* [6] | *Ord. Logit* [7] |
|---|---|---|---|---|---|---|---|
| intensity | | **.158*** ** | **.054** ** | | | **.099*** ** | **.249*** ** |
| | **(.026)** | **(.059)** | **(.023)** | | | **(.034)** | **(.074)** |
| intensity compulsory | | | | **.056** ** | **.105** ** | | |
| | | | | **-0.025** | **(.049)** | | |
| language at home | .688*** | 1.669*** | | | | | |
| | (.096) | (.235) | | | | | |
| AGE AT ARRIVAL | NO | NO | NO | NO | NO | NO | NO |
| CONTROLING FOR CDF | NO | NO | NO | NO | NO | YES | YES |
| INDIVIDUAL CONTROLS | YES | YES | YES | YES | YES | YES | YES |
| YEARS OF EDUCATION | YES | YES | OLS | YES | YES | YES | YES |
| YEAR OF BIRTH | YES | YES | NO | NO | NO | YES | YES |
| | YES | YES | YES | YES | YES | YES | YES |
| AGE and AGE SQUARED | NO | NO | YES | YES | YES | NO | NO |
| No. obs. | 1467 | 1467 | 1467 | 1467 | 1467 | 1467 | 1467 |
| Rsq | .361 | | .299 | .296 | | .325 | |
| Pseudo Rsq | | .171 | | | .135 | | .149 |

Note: The specifications reported in Columns 1 and 2 include cohort fixed effects. The specifications reported in Columns 3 and 4 include the age and the age squared of the respondents. The specification reported in Columns 5 includes the age and the age squared of the respondents and the variable intensity is instrumented by the variable intensity compulsory (the first stage is in Table 3.10). The specifications reported in Columns 6 and 7 include the age and the age squared of the respondents. All the specifications include years of education fixed effects, province fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level (Columns 1-2) and cohort level (Columns 3-7). * Significant at 10%, ** significant at 5%, *** significant at 1%

## Table 10: Other robustness check: Restricting the sample

| | *Dependent variable: Identity* | | |
|---|---|---|---|
| **sample** | **ols** | **ologit** | **obs** |
| 1948-1983 | .082*** | .183*** | 1467 |
| | (.026) | (.058) | |
| 1949-1982 | .092*** | .212*** | 1397 |
| | (.027) | (.061) | |
| 1950-1981 | .097*** | .225*** | 1326 |
| | (.027) | (.063) | |
| 1951-1980 | .108*** | .249*** | 1245 |
| | (.027) | (.064) | |
| 1952-1979 | .095*** | .226*** | 1165 |
| | (.029) | (.068) | |
| 1953-1978 | .075** | .183*** | 1091 |
| | (.029) | (.068) | |
| 1954-1977 | .094*** | .227*** | 1022 |
| | (.03) | (.069) | |
| 1955-1976 | .103*** | .259*** | 938 |
| | (.032) | (.074) | |
| 1956-1975 | .115*** | .289*** | 853 |
| | (.031) | (.072) | |
| 1957-1974 | .109*** | .277*** | 778 |
| | (.033) | (.076) | |
| 1958-1973 | .119*** | .329*** | 704 |
| | (.035) | (.08) | |
| 1959-1972 | .104*** | .292*** | 615 |
| | (.038) | (.086) | |
| 1960-1971 | .081* | .242** | 529 |
| | (.042) | (.095) | |
| 1961-1970 | .121*** | .322*** | 442 |
| | (.039) | (.098) | |

Note: All the specifications include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondent. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%.

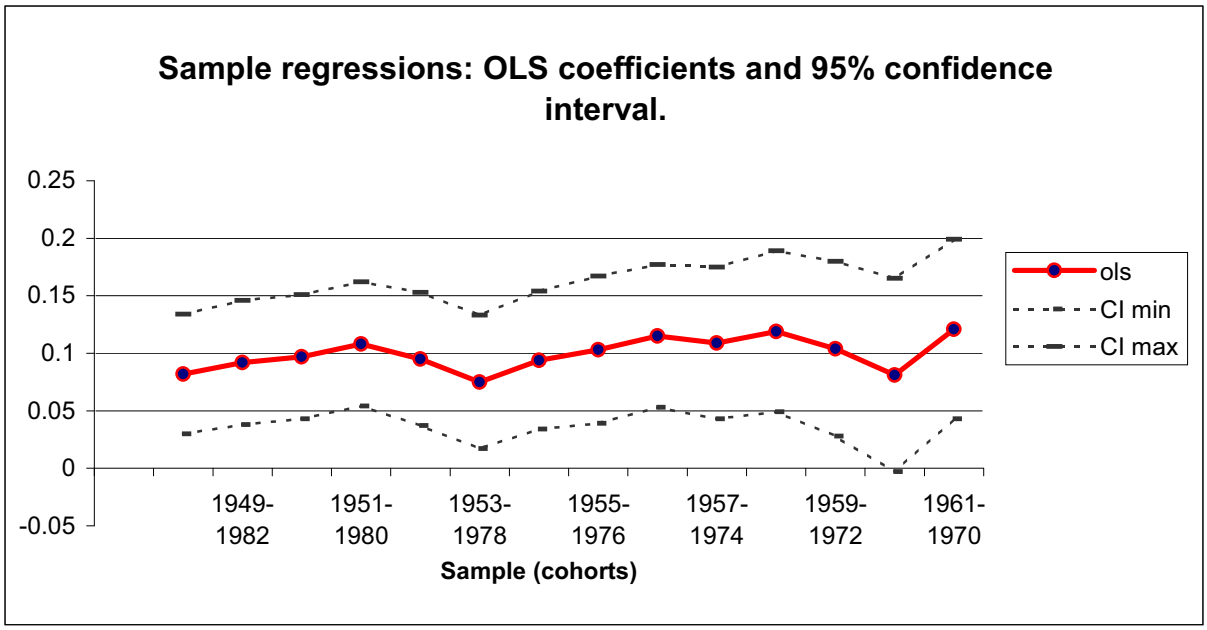



**Figure 8: Sample restrictions.**

## Table 11: Other robustness checks: Migration.

| *Dependent variable:* | *Identity* | | | | *Outflows* | |
|---|---|---|---|---|---|---|
| | *OLS* [1] | *Ord. Logit* [2] | *OLS* [3] | *Ord. Logit* [4] | *OLS* [5] | *OLS* [6] |
| intensity | .069** | 0.164*** | .07** | 0.18*** | | |
| | (.027) | (.063) | (.033) | (.079) | | |
| Cat | | | | | -.029 | -0.092 |
| | | | | | (.070) | (.069) |
| After1983 | | | | | 0.034 | .018 |
| | | | | | (.024) | (.039) |
| CatXAfter1983 | | | | | -0.032 | -.016 |
| | | | | | (.099) | (.097) |
| SAMPLE | 1948-198 catorigin | 1948-198 catorigin | 1953-197 catorigin | 1953-197 catorigin | all Spain | 5 richest |
| INDIVIDUAL CONTROLS | YES | YES | YES | YES | NO | NO |
| YEARS OF EDUCATION | YES | YES | YES | YES | NO | NO |
| YEAR OF BIRTH | YES | YES | YES | YES | NO | NO |
| PROVINCES | YES | YES | YES | YES | NO | NO |
| No. obs. | 1333 | 1333 | 994 | 994 | 170 | 60 |
| Rsq | 0.288 | | 0.287 | | | |
| Pseudo Rsq | | 0.135 | | 0.138 | .017 | 0.073 |

Note: In Columns 1 and 4 *identity* is the dependent variable. Specifications 1-4 include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. In the specifications reported in Columns 1 and 2, the sample considered consists only of respondents born in Catalonia and in the specifications reported in Columns 3 and 4 is further restricted to individuals born between 1953 and 1978. Robust standard errors are reported between parenthesis and in Specifications 1-4 are clustered at the cohort-years of education level. In Columns 5 and 6 *outflows* is the dependent variable. In Column 5 the sample consists of all the Spanish regions, in Colum 6 only of the 5 richest. * Significant at 10%, ** significant at 5%, *** significant at 1%.

**Table 12: Baseline Results with other surveys**

*Dependent variable: Identity*

| | | Intensity | Observations |
|---|---|---|---|
| CIS 2001 (baseline) | OLS | .082*** | 1467 |
| | | (.026) | |
| | Ord Logit | .183*** | 1467 |
| | | (.058) | |
| CIS 1998 (language) | OLS | .084* | 535 |
| | | (.044) | |
| | Ord Logit | .171* | 535 |
| | | (.089) | |
| CIS 2006 (post referendum) | OLS | .061* | 1062 |
| | | (.034) | |
| | Ord Logit | .200** | 1062 |
| | | (.078) | |
| CIS 2006 (post election) | OLS | .113*** | 1017 |
| | | (.034) | |
| | Ord Logit | .232*** | 1017 |
| | | (.074) | |
| Bofill | OLS | .065*** | 7521 |
| | | (.009) | |
| | Ord Logit | .154*** | 7521 |
| | | (.023) | |

## Table 13: Effect for each level of treatment

*Dependent variable: Identity*

| | *OLS* [1] | *Ord. Logit* [2] | *OLS* [3] | *Ord. Logit* [4] |
|---|---|---|---|---|
| 1 YEAR OF TREATMENT | -.185 | -.343 | -.117 | -.445 |
| | (.158) | (.326) | (.127) | (.313) |
| 2 YEARS OF TREATMENT | .039 | .118 | .318* | .751* |
| | (.185) | (.354) | (.178) | (.422) |
| 3 YEARS OF TREATMENT | .665*** | 1.29*** | .548*** | 1.147*** |
| | (.162) | (.329) | (.124) | (.303) |
| 4 YEARS OF TREATMENT | -0.91 | .062 | .285 | .721* |
| | (.21) | (.373) | (.177) | (.398) |
| 5 YEARS OF TREATMENT | 1.033*** | 1.981*** | .785*** | 1.6*** |
| | (.211) | (.434) | (.163) | (.407) |
| 6 YEARS OF TREATMENT | .383 | .904** | .594*** | 1.479*** |
| | (.245) | (.443) | (.227) | (.508) |
| 7 YEARS OF TREATMENT | .992*** | 2.013*** | .934*** | 1.967*** |
| | (.226) | (.447) | (.194) | (.471) |
| 8 YEARS OF TREATMENT | .562** | 1.264** | .837*** | 1.977*** |
| | (.28) | (.505) | (.264) | (.588) |
| 9 YEARS OF TREATMENT | .764*** | 1.561*** | .577** | 1.185** |
| | (.274) | (.538) | (.233) | (.557) |
| 10 YEARS OF TREATMENT | .464 | 1.189** | .823*** | 2.011*** |
| | (.329) | (.6) | (.303) | (.679) |
| 11 YEARS OF TREATMENT | .77** | 1.57** | .772*** | 1.575** |
| | (.317) | (.624) | (.262) | (.63) |
| 12 YEARS OF TREATMENT | .695** | 1.585** | 1.096*** | 2.485*** |
| | (.35) | (.637) | (.321) | (.719) |
| INDIVIDUAL CONTROLS | NO | NO | YES | YES |
| YEARS OF EDUCATION | YES | YES | YES | YES |
| YEAR OF BIRTH | YES | YES | YES | YES |
| PROVINCES | NO | NO | YES | YES |
| No. obs. | 1467 | 1467 | 1467 | 1467 |
| Rsq | .1 | | .33 | |
| Pseudo Rsq | | .036 | | .152 |

Note: The specifications reported in Columns 1 and 2 include years of education and year of birth fixed effects. The specifications reported in Columns 3 and 4 include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondent. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%.

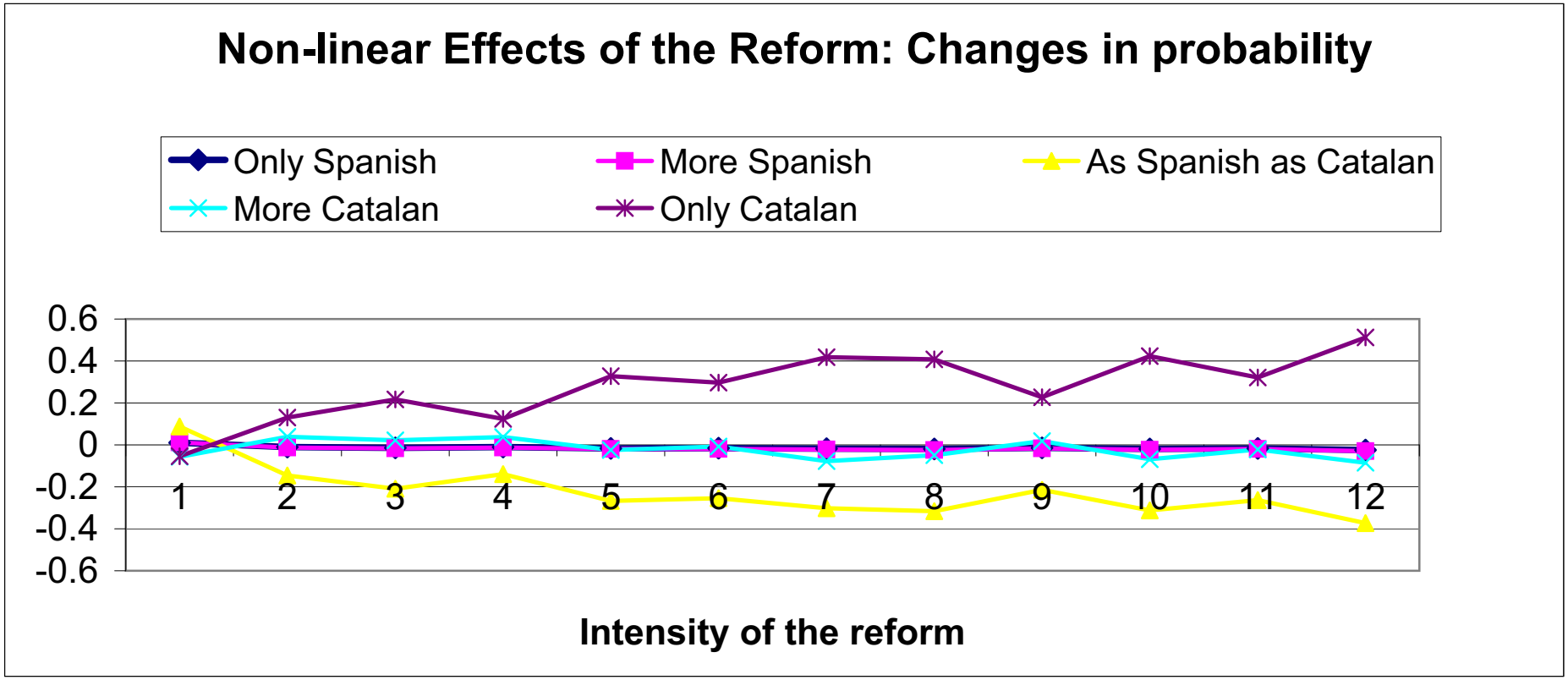



**Figure 9**

Note: we plot how the effect of the reform changes with the intensity of the reform, i.e. the number of years of Catalan instruction. This corresponds to the ordered logit regression in Column 4 of Table 12.

**Table 14: Heterogeneous Effects**

**PANEL A:** *Dependent variable: Identity*

| | *OLS* [1] | *Ord. Logit* [2] | *OLS* [3] | *Ord. Logit* [4] | *OLS* [5] | *Ord. Logit* [6] |
|---|---|---|---|---|---|---|
| Affected secondary | 0.182 | 0.368 | | | | |
| | (0.163) | (0.389) | | | | |
| Affected primary (less 4 yrs) | 1.205* | 3.556 | | | | |
| | (0.706) | (2.221) | | | | |
| Affected primary (more 4 yrs) | 1.334* | 3.876* | | | | |
| | (0.697) | (2.206) | | | | |
| Affected primary and sec | 1.503** | 4.214* | | | | |
| | (0.693) | (2.217) | | | | |
| intensity*non Cat.origin | | | .062* | .115 | | |
| | | | (.035) | (.078) | | |
| intensity*Cat.family | | | 0.086*** | 0.197*** | | |
| | | | (.026) | (.06) | | |
| intensity*mixed family | | | 0.081*** | 0.172*** | | |
| | | | (.029) | (.065) | | |
| intensity*non Cat.family | | | 0.075*** | 0.163** | | |
| | | | (.028) | (.063) | | |
| non catalan origin | | | -1.272*** | -2.835*** | | |
| | | | -0.114 | (.272) | | |
| mixed family | | | -.472*** | -.976*** | | |
| | | | (.086) | (.184) | | |
| non catalan family | | | -.913*** | -2.047*** | | |
| | | | (.095) | (.210) | | |
| Intensity (Catalan at home) | | | | | 0.073 [0.027]*** | 0.182 [0.062]*** |
| Intensity (no Catalan at home) | | | | | 0.052 [0.026]* | 0.122 [0.061]** |
| Catalan at home | | | | | 0.592 [0.119]*** | 1.405 [0.285]*** |
| YEARS OF EDUCATION | | | YES | YES | | |
| YEAR OF BIRTH | | | YES | YES | | |
| PROVINCES | | | YES | YES | | |
| No. obs. | 1467 | 1467 | 1467 | 1467 | 1467 | 1467 |
| Rsq | 0.32 | | | | 0.36 | |

In Columns 1 and 2 the reference category are theindividuals who have not been exposed to the reform, in Columns 3 and 4 the individuals who were not born in Catalonia, in Columns 5 and 6 the individuals who did not speak the Catalan language at home. Individual controls include a dummy variable for gender, for individuals not born in Catalonia, individuals born in Catalonia but with non-catalan parents and individuals born in Catalonia with mixed parents. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at the 10%, ** significant at the 5%, *** significant at the 1%.

**PANEL B**

Marginal effects

| | Only Spanish | | More Spanish | | As Spanish as Catalan More Catalan | | | | Only Catalan | |
|---|---|---|---|---|---|---|---|---|---|---|
| | coeff | se | coeff | se | coeff | se | coeff | se | coeff | se |
| Affected secondary | -0.009 | 0.008 | -0.010 | 0.010 | -0.073 | 0.078 | 0.046 | 0.042 | 0.046 | 0.054 |
| Affected primary (less 4 yrs) | -0.030 | 0.005 | -0.035 | 0.007 | -0.438 | 0.074 | -0.203 | 0.242 | 0.706 | 0.319 |
| Affected primary (more 4 yers) | -0.039 | 0.010 | -0.044 | 0.011 | -0.483 | 0.087 | -0.174 | 0.193 | 0.740 | 0.295 |
| Affected primary and sec | -0.104 | 0.070 | -0.102 | 0.056 | -0.544 | 0.075 | 0.073 | 0.137 | 0.677 | 0.330 |
| intensity*non Cat.origin | -0.003 | 0.002 | -0.004 | 0.002 | -0.022 | 0.015 | 0.016 | 0.011 | 0.013 | 0.009 |
| intensity*non Cat.family | -0.005 | 0.002 | -0.005 | 0.002 | -0.031 | 0.012 | 0.023 | 0.009 | 0.018 | 0.007 |
| intensity*mixed family | -0.005 | 0.002 | -0.005 | 0.002 | -0.033 | 0.013 | 0.024 | 0.010 | 0.019 | 0.007 |
| intensity*Cat.family | -0.005 | 0.002 | -0.006 | 0.002 | -0.038 | 0.012 | 0.028 | 0.009 | 0.022 | 0.007 |
| non Catalan origin | 0.244 | 0.050 | 0.160 | 0.023 | 0.090 | 0.042 | -0.339 | 0.021 | -0.154 | 0.011 |
| mixed family | 0.037 | 0.011 | 0.039 | 0.011 | 0.156 | 0.022 | -0.146 | 0.028 | -0.086 | 0.013 |
| non Catalan family | 0.096 | 0.020 | 0.091 | 0.013 | 0.262 | 0.022 | -0.276 | 0.024 | -0.173 | 0.017 |
| Intensity(Catalan at home) | -0.005 | 0.002 | -0.005 | 0.002 | -0.036 | 0.012 | 0.027 | 0.010 | 0.018 | 0.006 |
| Intensity (no Catalan at home) | -0.003 | 0.002 | -0.003 | 0.002 | -0.024 | 0.012 | 0.018 | 0.009 | 0.012 | 0.006 |
| Catalan at home | -0.035 | 0.009 | -0.038 | 0.009 | -0.264 | 0.050 | 0.181 | 0.032 | 0.156 | 0.034 |

**Table 15: Reform and Turnout**

*Dependent variable:Turnout*

| | *OLS* [1] | *Logit* [2] | *OLS* [3] | *Logit* [4] |
|---|---|---|---|---|
| intensity | .056*** | .045*** | .057*** | .045*** |
| | (.018) | (.015) | (.018) | (.015) |
| non Catalan origin | | | -.113* | -.142* |
| | | | (.057) | (.073) |
| Cat origin mixed family | | | -.084** | -.095*** |
| | | | (.039) | (.045) |
| Cat origin non catalan family | | | -.1291*** | -.144*** |
| | | | (.038) | (.043) |
| YEARS OF EDUCATION | YES | YES | YES | YES |
| YEAR OF BIRTH | YES | YES | YES | YES |
| PROVINCES | NO | NO | YES | YES |
| No. obs. | 1248 | 1248 | 1248 | 1248 |
| Rsq | .078 | | .099 | |
| Pseudo-Rsq | | .071 | | .093 |

Note: The dependent variable is a Dummy variable which is equal to 1 if the respondent voted during the egional elections. The specifications reported in Columns 1 and 2 include years of education and year of birth fixed effects. The specifications reported in Columns 3 and 4 include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondent. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia (omitted category). Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%. Marginal effects evaluated at the average are reported for the logit regressions.

**Table 16: Reform and Ethnic voting (1)**

*Dependent variable:Probability of ethnic voting (among voters)*

| | CIU+ERC+IC *OLS* [1] | CIU+ERC+IC *Logit* [2] | CIU+ERC *OLS* [3] | CIU+ERC *Logit* [4] | CIU *OLS* [5] | CIU *Logit* [6] | CIU+ERC+IC *OLS* [7] | CIU+ERC+IC *Logit* [8] | CIU+ERC *OLS* [9] | CIU+ERC *Logit* [10] | CIU *OLS* [11] | CIU *Logit* [12] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| intensity | .041*** | .046*** | .047** | .049** | .049*** | .053*** | .040** | .051*** | .046** | .057** | .049** | .052** |
| | (.018) | (.017) | (.020) | (.087) | (.020) | (.022) | (.016) | (.017) | (.019) | (.023) | (.020) | (.022) |
| non Catalan origin | | | | | | | -.359*** | -.408*** | -.399*** | -.423*** | -.169** | -.171** |
| | | | | | | | (.072) | (.074) | (.072) | (.063) | (.075) | (.070) |
| Cat origin mixed family | | | | | | | -.135*** | -.179*** | -.183*** | -.220*** | -.044 | -.046 |
| | | | | | | | (.042) | (.054) | (.021) | (.053) | ( .061) | (.061) |
| Cat origin non catalan family | | | | | | | -.382*** | -.429*** | -.412*** | -.442*** | -.156*** | -.160*** |
| | | | | | | | (.045) | (.051) | (.051) | (.050) | (.048) | (.047) |
| YEARS OF EDUCATION | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| YEAR OF BIRTH | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| PROVINCES | NO | NO | NO | NO | NO | NO | YES | YES | YES | YES | YES | YES |
| No. obs. | 941 | 941 | 941 | 941 | 941 | 941 | 941 | 941 | 941 | 941 | 941 | 941 |
| Rsq | .057 | | .04 | | .054 | | .194 | | .194 | | .084 | |
| Pseudo-Rsq | | .046 | | .030 | | .042 | | .161 | | .153 | | .065 |

Note: The dependent variable is a Dummy variable which is equal to 1 if the respondent voted for a Catalanist party during the egional elections. In Columns 1-2, 7-8 we consider CIU, ERC and IC as Catalanist parties; in Columns 3-4, 9-10 we consider CIU, ERC as Catalanist parties; in columns 5-6, 11-12 we consider only CIU as a Catalanist party. We restrict the sample only to respondents who voted. The specifications reported in Columns 1-6 include years of education and year of birth fixed effects. The specifications reported in Columns 7-12 include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondent. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia (omitted category). Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%. Marginal effects evaluated at the average are reported for the logit regressions.

**Table 17: Reform and Ethnic voting (2)**

*Dependent variable:Probability of ethnic voting (among all citizens)*

| | CIU+ERC+IC *OLS* [1] | CIU+ERC+IC *Logit* [2] | CIU+ERC *OLS* [3] | CIU+ERC *Logit* [4] | CIU *OLS* [5] | CIU *Logit* [6] | CIU+ERC+IC *OLS* [7] | CIU+ERC+IC *Logit* [8] | CIU+ERC *OLS* [9] | CIU+ERC *Logit* [10] | CIU *OLS* [11] | CIU *Logit* [12] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| intensity | .049*** | .057*** | .053*** | .059*** | .048*** | .053*** | .054*** | .068*** | .057*** | .071*** | .050*** | .055** |
| | (.015) | (.017) | (.017) | (.019) | (.016) | (.091) | (.014) | (.018) | (.016) | (.020) | (.016) | (.019) |
| non Catalan origin | | | | | | | -.363*** | -.347*** | -.381*** | -.335*** | -.182*** | -.158*** |
| | | | | | | | (.067) | (.053) | (.064) | (.045) | (.063) | (.049) |
| Cat origin mixed family | | | | | | | -.176*** | -.190*** | -.207*** | -.204*** | -.077 | -.073 |
| | | | | | | | (.043) | (.044) | (.045) | (.040) | (.051) | (.045) |
| Cat origin non catalan family | | | | | | | -.368*** | -.379*** | -.377*** | -.370*** | -.164*** | -.157*** |
| | | | | | | | (.038) | (.035) | (.043) | (.037) | (.041) | (.035) |
| YEARS OF EDUCATION | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| YEAR OF BIRTH | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| PROVINCES | NO | NO | NO | NO | NO | NO | YES | YES | YES | YES | YES | YES |
| No. obs. | 1248 | 1248 | 1248 | 1248 | 1248 | 1248 | 1248 | 1248 | 1248 | 1248 | 1248 | 1248 |
| Rsq | .062 | | .053 | | .055 | | .174 | | .176 | | .087 | |
| Pseudo-Rsq | | .047 | | .004 | | .048 | | .134 | | .137 | | .075 |

Note: The dependent variable is a Dummy variable which is equal to 1 if the respondent voted for a Catalanist party during the egional elections. In Columns 1-2, 7-8 we consider CIU, ERC and IC as Catalanist parties; in Columns 3-4, 9-10 we consider CIU, ERC as Catalanist parties; in columns 5-6, 11-12 we consider only CIU as a Catalanist party. The specifications reported in Columns 1-6 include years of education and year of birth fixed effects. The specifications reported in Columns 7-12 include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondent. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia (omitted category). Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%. Marginal effects evaluated at the average are reported for the logit regressions.