



UNIVERSIDAD CARLOS III DE MADRID

working  
papers

Working Paper 07-77  
Economic Series 46  
November 2007

Departamento de Economía  
Universidad Carlos III de Madrid  
Calle Madrid, 126  
28903 Getafe (Spain)  
Fax (34) 91 6249875

## Identity and Language Policies\*

Oriol Aspachs-Bracons<sup>†</sup>, Irma Clots-Figueras<sup>‡</sup>, Paolo Masella<sup>§</sup>

### *Abstract*

---

The process of individual identity formation is still an enigma, as it is the capacity of public bodies to intervene on 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 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 be present also among individuals whose parents do not have Catalan origins. To the best of our knowledge, this is the first paper to analyze how policies affect individual identity.

---

---

\* The authors thank Oriana Bandiera, Robin Burgess, Luigi Guiso, Ramón Marimón and seminar participants at the London School of Economics and Universidad Carlos III de Madrid. Irma Clots-Figueras gratefully acknowledges financial support from the MEC grant SEJ2004-07861

<sup>†</sup> LSE and FMG

<sup>‡</sup> Universidad Carlos III de Madrid

<sup>§</sup> EUI

# 1 Introduction

Ethnic diversity has been shown to be correlated with economic growth and low-quality governments. Easterly and Levine (1997) claimed that high levels of ethnic fragmentation were at the root of the Africa's growth tragedy. The literature that followed confirmed these results. La Porta, Lopez de Silanes, Shleifer and Vishny (1999) showed that ethnic heterogeneity is negatively correlated with the quality of government and, more recently, Montalvo and Reynal (2005) 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.

Several social scientists (Anderson 1983, Bates 1983, Horowitz 1985), however, 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 be dependent on a number of economic and policy choices.

Thus, ethnic diversity can be influenced by precise political strategies aimed at promoting national or ethnic identities among the population. In fact, in ethnically diverse countries, politicians and political scientists have focused on very different policies as mechanisms of minority integration and conflict reduction. Nation building policies (for example the implementation of a unique national language 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.

Limited research, however, has examined the extent to which identity can be shifted by cultural policies and regulation. As the UN report for 2004 pointed out, language regulation influences the social environment that encourages identity formation. Since identity is typically formed in the early ages of life, language in schools turns out to be particularly important.

Our paper takes a step in this direction by analyzing the effect of a particular educational policy on the process of identity formation. 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 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 appears to be present also among individuals educated in Catalonia after the reform but whose parents do not have Catalan origins.

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 their effect on individual sentiments: multicultural policies tend to favour the development of regional identities while nation building policies tend to promote the expansion of a common national feeling. To the best of our knowledge, this is the first paper to analyze how policies affect individual identity.

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

A number of factors leads us to believe that the link between the linguistic reform and individual identity is causal. First, some robustness checks allow us to rule out the existence of education specific cohort trends for Catalan feelings. 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 does not present anomalous patterns. Moreover, since individual years of schooling are typically an individual and parental choice, we may be concerned that they are affected by the reform. Both another identification strategy that relies on the number of years during which students were taught in Catalan within compulsory education and comparisons with the pattern of

educational achievements in other Spanish regions suggests that this is not the case.

This paper is related to the literature previously mentioned on the relationship between ethnic fragmentation and economic outcomes and 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 propose a utility function that depends on the individual's assigned or chosen social category, on the match between (exogenous) "prescriptions" for that category and 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 have followed the work by Akerlof and Kranton, including that by Benabou and Tirole (2007), who endogenized identity payoffs and categorical prescription. However, the number of empirical studies analyzing the determinants of individual identity remain limited.

This paper is also connected to the literature on endogenous preferences. This literature (see Bowles 1998 for a review of the topic) 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.

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 after 1983. Then it discusses the data and the empirical strategy employed. Section 3 presents the basic empirical evidence and provides several robustness checks. Section 4 shows the results under a different identification strategy, while Section 5 discusses the role of individual origin. Finally, Section 6 concludes.

## 2 Background and empirical strategy

### 2.1 Background

The Catalan language developed by the 9th century from vulgar Latin on both sides of the eastern part of Pyrenees Mountains. Its territorial expansion went hand in hand with the expansion of the Catalano-Aragonese Crown, which was established in the "País Valencià", the Balearic Islands, the south of France and the town of Alghero (in Sardinia, Italy) during many centuries.

These are the regions where Catalan is present nowadays with more than 9 million speakers, making it the most important regional language of the European Union. Catalonia is the region that hosts most of them, with more than 6 million speakers. However, in order to achieve it, the Catalan Government has had to protect the Catalan Language, and sometimes even promote its usage, because it was banned during Franco's dictatorship (1940-1975).

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 "General Directory of Linguistic Policy (1983)" which allowed the educational system, which prior to that date was all in Spanish as an inheritance from the Franco's period, to become bilingual. The law states that:

*Art. 2: "The Certificate of Basic Educational Attainment will not be achieved without proving proficiency of the official languages of Catalonia (Catalan and Spanish)".*

*Art. 6: "The Catalan Language, as it is the own language of Catalonia, it is also the language used in the education system".*

*Art. 7.3: "The Language used in the education system will never separate the alumni due to linguistic differences".*

Since the Catalan Education system had been all in Spanish for such a long time period, the transition to a bilingual system was smooth. As stated in the "Order of Application of the General Directory of Linguistic Policy

(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 makes clear again that the increased presence of Catalan Language at 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.

## **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. This survey is performed relatively often, but the one done in 2001 is of special interest as individuals were asked not only about their national feelings, but they were also asked about their origin and their parents' origin.

To identify the individuals' national attachment to Spain or Catalonia, we rely on the following question: "With which of the following sentences do you feel more identified? (i) I feel only Spanish, (ii) I feel more Spanish than Catalan, (iii) I feel as Spanish as Catalan, (iv) I feel more Catalan than Spanish, (v) I feel only Catalan". That is, the closer is this number to 5, the more Catalan and less Spanish you feel, and vice versa.

In Spain, primary education lasts 8 years and starts at the age of 6, while secondary education lasts 4 years. Therefore, the cohorts that may have been treated are those born between 1966 and 1983, while the control cohorts are those born between 1948 and 1965. 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, an individual from the 1966 cohort that 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 that 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 only 1 year of treatment.

Figure (1) describes the relationship between cohorts, years of schooling and years of treatment. There are 4 major educational categories: (i) individuals that did not complete primary education, which we assume that received 4 years of schooling; (ii) individuals that completed only primary education, which received 8 years of education; (iii) individuals that started but did not complete secondary education or that received some professional training, for whom we assume that received 10 years of education; and (iv) individuals that completed secondary education or a higher level of professional training, who went through 12 years of education<sup>1</sup>.

Figure (2) plots the average time that each cohort has been exposed to Catalan at school. The average presence of Catalan at school is obtained from the questionnaire, where individuals are asked to classify between 1 and 5 how much their education was in Catalan and how much of it was in Spanish, 1 being only Spanish and 5 only Catalan. This graph makes clear that there has been a huge increase over time in the use of Catalan at school. However, the increase started slightly after the 1960s', and hence, earlier than 1966, the year that we define as the first cohort to be potentially treated. This may show (as the anecdotal evidence we collected confirms) that some schools started teaching some subjects in Catalan earlier at their own discretion.

What makes the change in the linguistic policy of Catalan schools a unique case study is the following: although Spanish immigrants saw that language of instruction was changed to a foreign one, they reached levels of education in Catalan very similar to those received by natives. More precisely, in terms of average years of education (see Table 1). The main reasons behind this are that primary and secondary education were publicly provided and schooling was mandatory until the age of 14 for most cohorts.

Our identification strategy relies on variation in Catalan instruction across cohorts and years of education. A first and simple analysis of the data shows interesting patterns in the variation of the Catalan sentiments be-

---

<sup>1</sup>We do not consider university education since the law we study did not affect it. Thus, individuals with university education are included in the 12 years of education group.

tween treated and control cohorts, and within each cohort (high vs. low educated). If there is a link between language policy and peoples' identity, the reform will boost the Catalan sentiment for those cohorts that were educated in Catalan, which will be younger (born after 1966) and more educated. Figures (3) to (15) show that this is indeed the case. Figures (3) and (4) plot, respectively, the evolution of national sentiments of high (with secondary education) and low (without secondary education) educated Catalans by cohort. The vertical lines divide control and treated cohorts. As seen in Figure (3), among control cohorts, younger and more educated cohorts are more likely to feel "as Catalan as Spanish" and less likely to feel "only Catalan": Catalan sentiments decrease with year of birth. The reform, however, seems to have reversed this trend in Catalan sentiments. A similar process took place among respondents without secondary education. Figure (4) shows that among control cohorts, the younger ones tend to feel "only Spanish". The linguistic policy seems to have altered this tendency, which is in part absorbed by the increase in the fraction of low educated Catalans feeling more Spanish than Catalan. Moreover, it is also possible to notice that prior to the reform, younger cohorts are more likely to feel "as Catalan as Spanish" and less likely to feel "more Catalan than Spanish". After the reform these trends change their sign, preventing the widening of the gap between the two.

The linguistic law is also expected to be accompanied by an increase in the within cohort variation, augmenting more the Catalan sentiments of those individuals that received higher levels of treatment, that is, those more educated. To check whether this is the case, Figure (5) plots the difference between the fractions of high and low educated people for each type of answer against year of birth. It is interesting to describe the pattern of the effect of the education reform on the category only Catalan. While prior to the reform the difference was pretty stable around zero, there is a clear change with the reform that leads the difference to increase significantly as the gap between the treatment intensity widens. This is partly reduced when the low educated begin to be affected by the reform and the fraction of them feeling only Catalan increases. Figures (6) to (10) and (11) to (15) show the between and within variation, respectively, in more detail.



Tables (3) provides the full set of descriptive statistics of the main variables 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. Cohorts who started their primary education on or after that academic year were completely affected, while cohorts who started primary education prior to this but were still in primary or secondary education after 1983-1984 were only partially affected by the reform. The students' exposure to the Catalan language varies according to the number of years they have been in education after the reform.

Thus, exposure to the reform varies according to year of birth and the number of years of education for each individual. 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) I feel only Spanish, (2) I feel more Catalan than Spanish, (3) I feel as Spanish as Catalan, (4) I feel more Catalan than Spanish, (5) I feel only Catalan.

$L_{ijm}$  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, individual's origin, parent's origin and province dummies<sup>2</sup>.

The identification strategy relies on the fact that there is no other variable that affects how some cohort-years of education group feels about the Catalan

---

<sup>2</sup>We exclude from the sample all the migrants who received some of their education outside Catalonia.

identity issue<sup>3</sup>. We run this specification, first as a linear regression and then as an ordered logit, due to the ordinal nature of the dependent variable<sup>4</sup>. 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.

## 3 Results

### 3.1 Difference-in-Difference Estimates

Table 3 reports estimates for equation (1). In columns 1 and 2 we show estimates in which we only control for years of education and year of birth fixed effects, for the OLS and the ordered logit specifications, respectively. OLS estimates show a positive and significant coefficient, suggesting that an increase in exposure to the reform increases the probability that an individual feels more Catalan. The ordered logit results go in the same direction and are significant. In columns 3 and 4 we add province dummies and controls for gender, family and individual origin. Coefficients are now slightly larger than before. Even if the coefficients for an OLS regression in which the dependent variable is a discrete variable with values from 1 to 5 are difficult to interpret, the results suggest that those who received Catalan-intensive instruction tend to have significantly more intense Catalan feelings. The intensity of Catalan sentiments is, as predictable, correlated with individual origin and with the origin of the parents of the interviewed. We classified respondents into 4 categories: 1) individuals that were not born in Catalonia 2) individuals that were born in Catalonia but whose parents were not 3) individuals that were born in Catalonia but with only one of parents born in Catalonia 4) individuals that were born in Catalonia whose parents were both born in Catalonia. We find that Catalan identity is strongest among respondents with Catalan origins and with parents with Catalan origins and it is weakest among respondents that were not born in Catalonia.

---

<sup>3</sup>In another section of the paper we discuss the fact that years of education may not be exogenous to the reform.

<sup>4</sup>Results for ordered probit are very similar and available on request

In order to interpret the ordered logit coefficients, in Figure 16, 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. This figure helps understanding the size of the effect of the exposure to 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 be strongly decreasing with years of treatment (from more than 0.5 with 0 treatment to just more than 0.1 with 12 years of treatment). The probability of choosing "Only Spanish" and "more Spanish than Catalan" decrease slightly and the probability of choosing "More Catalan than Spanish" shows a small increase with the length of exposure to the reform. Thus, results suggest 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".

### 3.2 Channel

The reform is expected to increase the amount of Catalan taught at schools and to increase as well the likelihood that affected individuals use Catalan as a normal language of expression. This change in the language of instruction can reinforce Catalan identity, both because Catalan becomes a 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. As it has been already mentioned in the data section, in the survey 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 between the index previously described

and years of education.

In the survey, there are also several questions that provide information about the use of the Catalan language in the everyday life: 1) when the interviewed meets with friends, 2) when he/she is at home, 3) when he/she goes shopping, 4) if he/she is asked for directions in the street, 5) when he/she answers the phone, 6) when he/she interacts with civil servants 7) or when he/she is with his/her 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 interviewed uses the Catalan language in all of these circumstances. As a first step, in the first and second columns of Table 4 we check whether the two indices are related to identity. As expected, there is a positive and strong relationship between the two: the greater the use of the Catalan language at school, the more intense the Catalan feelings and, at the same time, the greater the use of Catalan in the everyday life, the more intense are the Catalan feelings. As a second step, in columns 7 and 8, we run a 2SLS in which we regress two indices on our measure of the intensity of the reform and then regress our measure of identity on the predicted value obtained in the first stage. The coefficients are both positive and significant at the standard level, meaning that language was truly an important channel through which the reform affected individual feelings. First stage coefficients show the positive impact that the reform had on the number of years of education taught in Catalan and the extent to which Catalan is used in everyday life.

The fact that the 2SLS results are larger than the OLS ones indicates that, if there is an omitted variable that affects individual identity and the self-reported measure of Catalan at school, this will be negatively correlated with identity and positively correlated with the self-reported measure of Catalan at school (or vice versa). A possible explanation for this is that people who say they feel more Catalan may be more likely to perceive the amount of education they received in Catalan as being less or insufficient, while people who feel more Spanish might tend to perceive the amount of education they received in Catalan as excessive. There is no question in the survey that allows us to test this, but, since our self-reported measure of schooling in Catalan is based on recall, this remains a possibility.

### 3.3 Effect for each level of treatment

The marginal effect of the treatment 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 be 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. The effect of the reform may be different for a student who was only affected by one year than for a student who was affected for 12 years. In this section we wish to investigate how the effect increases with years of treatment. As a result, and similarly to Angrist et al. (2006), we decided to introduce a dummy for each possible level of treatment and run the following specification:

$$y_{ijm} = \alpha + \sum_{k=1}^{12} \beta_k I(L_{jm} = k) + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \quad (2)$$

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 10 report ordinary least square estimates of equation (2). 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. The results in both columns go in the same direction. Identity is affected by the reform, although 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 19 confirms the patterns proposed in the previous section. In this figure we plot how a 0 to 1 change in each treatment dummy 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 the years of treatment, while probability of answering "I am as Spanish as Catalan" appears to be strongly decreasing with the 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.

### 3.4 Heterogenous effects

The introduction of a Catalan based schooling system (even if mitigated by the rule that at least two subjects per year must be taught in Spanish) can be interpreted as an example of a nation building policy within the Catalan region. It is interesting then 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  $L_{ijm}$ <sup>5</sup>.

Columns 1-2 of Table 11 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. Moreover the coefficient of the interaction term between the variable that captures the length of exposure to teaching in Catalan and a dummy variable equal to one if the respondent was born outside Catalonia is only marginally insignificant in column 2. Overall, the results suggest that the new schooling system successfully increased the level of integration of minorities within Catalonia.

## 4 Robustness checks

### 4.0.1 Exogenous trends

The estimates in Table 3, which point to higher Catalan feelings among those who have been exposed to more intensive Catalan instruction, may be con-

---

<sup>5</sup>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.

founded by education-specific cohort trends in Catalan feelings. Such trends could have arisen through several channels: more educated people could feel more Catalan, and that younger people could be more educated than older people. Different cohorts have been raised by parents with different values and preferences (the younger parents being more pro-Catalan), and those parents with stronger Catalan preferences could have invested more in their children’s education.

In order to check 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 have not been affected by the reform. These are cohorts born between 1930 and 1965. Then, among those cohorts, we assigned a pseudo-treatment to the younger cohorts (the ones born between 1948 and 1965, which were used as controls in the main specification). This pseudo-treatment consists of the intensity of the reform received by individuals with the same number of years of education but born 18 years later. We run the specification proposed by equation (1). We use both an ordinary least squares specification and an ordered logit one. The results are reported in columns 1 and 2 of Table 5. The sign of the coefficient of the pseudo treatment variable is very small and not significant. This is reassuring, as it may indicate that our results are not spuriously capturing the effect of a trend.

As a second check, we consider cohorts born between 1921 and 1980 and we use the following specification. To correct the difference-in-difference-type estimates for the education-cohort trends, we use a triple differences-type strategy. Triple differences estimates are based on the following equation,

$$y_{ijm} = \alpha + \beta L_{ijm} + \theta l_{ijm} + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \quad (3)$$

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

$L_{ijm}$  are the years of real treatment for cohorts 1966-1980; while  $l_{ijm}$  equals years of pseudo treatment for cohorts 1936-1950 (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 differ-

ences 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 be capturing spurious trends is negative and not significant. We can conclude that, among cohorts affected by the reform, there are no education-specific cohort trends<sup>6</sup>.

As last step, we use individuals who arrived to Catalonia after they had finished education. Pseudo-treatment for these individuals are the years of education they received after 1983; since they received education outside Catalonia, they can not be affected by the reform. Results are shown in column 5 of Table 5. As they are very few observations, only OLS results are reported; however, the effect of the reform is very small and not significant for these individuals. This further confirms that the effect is not due to other policies that affected younger and more educated people in a different way.

#### 4.0.2 Taking the within variation seriously

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 (individuals of the same age that have different levels of education and so may have been exposed to different levels of treatment). However, the number of years of schooling is typically an individual choice (or a parental one) and may be affected by the introduction of the linguistic reform. One may imagine that individuals with very intense Spanish (and anti-Catalan) feelings, for example a migrant from another region of Spain, may have decided to drop out from school because they may feel less comfortable about receiving education in Catalan.

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 normally speaks Catalan would have experienced fewer difficulties learning Catalan at school and would less likely to drop out from education after the

---

<sup>6</sup>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



reform. The results of this specification are shown in columns 1 and 2 of Table 7 and are very similar to those obtained previously. This is reassuring, as the language spoken at home could be the main source of endogeneity.

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. 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 6 and 7. As predictable, the first stage of the 2SLS analysis (Table 6) 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, even if they are not very different in magnitude. 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.

Then, as an additional test, we do as in Angrist et al (2006) and we control for the cumulative distribution function for education using data from the Spanish Census 2001. That is, 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. This accounts for the possibility that the selection of students who reach higher levels of education changes over time, especially if the reform causes it. Our coefficients, shown in columns 6 and 7 increase slightly, which is reassuring.

Finally, instead of using years of education as intra-cohort variation in years of treatment, we use variation in the date of arrival as a determinant of more or less exposure to the language reform<sup>7</sup>. In order to perform this exercise we restrict the sample to individuals who were not born in Catalonia. These individuals arrived at different points in time and at different stages during their education. For each individual we compute the number of years, within compulsory education, for which they have studied in Catalonia after 1983. This is now our intensity variable, which we use to run:

---

<sup>7</sup>We include in the sample also migrants who received some education outside Catalonia and who were not included in any of the previous specifications.

$$y_{ijm} = \alpha + \beta L_{ijm} + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \quad (4)$$

Where the dependent variable  $y_{ijm}$  indicates whether individual  $i$ , from cohort  $j$  who arrived in Catalonia when he was  $m$  years old answers: (1) I feel only Spanish, (2) I feel more Catalan than Spanish, (3) I feel as Spanish as Catalan, (4) I feel more Catalan than Spanish, (5) I feel only Catalan.

$L_{ijm}$  is the length of exposure to teaching in Catalan.  $\gamma_j$  are the cohort dummies,  $\delta_m$  are age at arrival dummies and  $X_{ijm}$  is a vector of individual-level control variables, including gender, origin and years of education dummies.

This specification is particularly useful since it allows us to use both compulsory years of education, and cohort dummies. In this way we are able to control for any issue related with the endogeneity of the education decision and, at the same time, to rule out any potential problem of omitted variables correlated with both identity and age.

The results are shown in column 8 of Table 7 and are very similar to those obtained when the within-cohort variation used was given by differences in the number of years of education. Now we do not have enough observations to run an ordered logit, but the fact that the OLS results are very similar to those obtained previously is reassuring. This shows that our results are not driven by changes in years of education due to the reform.

Results in this section suggest that years of education are unlikely to have changed endogenously as a consequence of the reform. This is as well supported by the fact that schools provided facilities and extra time for students for whom Catalan was not their mother tongue. As a final check, we plot average years of education by cohort for Catalonia, the Madrid region (a region that has been at a similar level of development than Catalonia) and the rest of Spain. This is done using data from the Spanish Census in 2001. The graph (Figure 17) shows that years of education for cohorts affected by the reform (those born after 1966) evolved quite similarly in Catalonia and in Madrid and did not decline for the Catalan cohorts affected by the reform. Years of education increased more for the other communities in Spain, as a result of a process of convergence.

### **4.0.3 Restricting the sample**

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 but compare individuals aged 26 (cohort 1975) with individuals aged 45 (cohort 1956). As the cohorts we are comparing are more similar, the results should be more reliable in this sense. Table 8 shows that, when we restrict the sample and compare cohorts that have similar ages, results sensibly improve. This may be because we are comparing cohorts that are more similar among them and that have been affected by similar shocks. This allows us to increase the precision of our estimates and to capture better the effects of the reform.

### **4.0.4 Migration patterns**

The reform might have changed migration patterns into and out of Catalonia. If this is the case, our results could be capturing a change in the composition of the Catalan society rather than the effect of the reform. Schooling in Catalan could have implied an additional cost of migration to Catalonia for Spanish speakers. Potential migrants with very intense Spanish sentiments could decide not to migrate towards Catalonia and, vice-versa, while the additional migration cost could have been lower for open minded migrants. In addition, people with very intense Spanish sentiments could decide to leave Catalonia after the reform. If this is the case, our results could be the product of a different composition of the Catalan residents and not a consequence of the reform.

Since we do not know the year in which the parents of the interviewed migrated, we cannot exclude from the sample individuals who are potentially affected by this problem (with parents that migrated as a consequence of the new law). Then we decide to check if our results hold only for individuals born in Catalonia. Since the youngest cohort was born in 1983, we are then restricting the sample to individuals whose parents were already in Catalonia when the reform was implemented, so parents' decision to migrate towards Catalonia is less likely to be affected by the linguistic reform. Columns 1 and

2 of Table 9 shows 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 9 shows that results are robust to this robustness check.

Table 9 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 data from INE, in Figure 18 we compare the patterns of migration of Catalonia in the last 40 years with the patterns of migration of another Spanish region that is comparable in terms of income, i.e. Madrid. We plot net migration inflows as a fraction of migration outflows for Catalonia, Madrid and the other communities. It turns out there is no substantial difference, either before or after the reform, in fact Madrid seems to be losing more citizens after 1990.

## 5 Summary and conclusions

Much has been said about the possibility that policies affect individual identity, sentiments and preferences. However, to date, little case study research has been conducted on this matter. This paper presents evidence of a positive effect of the educational reform, by which the education system became bilingual and Catalan, together with Spanish, was taught at schools, on Catalan feelings. The positive effect extends also to individuals whose parents have no Catalan origins.

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 regional identities, nation building policies encourage the growth of a shared national sentiment.

Research of this sort together with the study of the effects of such reforms

on the 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.

## 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)
- 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 economic 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 Beckham': identity, socialization and assimilation" *CEPR Discussion paper* No 5662
- Bowles, S (1978) "Capitalist Development and Educational Structure" *World development*, 6, 783-796.

- Breton, A (1964) The Economics of Nationalism. *Journal of Political Economy*, 72, 4, 376-386.
- Breton, A (1978). Nationalism and Language Policies *Canadian Journal of Economics*, 11(4): 656-668
- Breton, A and Breton, M (1995) Nationalism Revisited. In Galeotti, G, Salomon, P and Wintrobe, R (eds). *Competition and structure: the political economy of collective decisions: essays in honour of Albert Breton*, Cambridge University Press, Cambridge.
- Coleman J and C. Rosberg (1964) "Political parties and national integration in tropical Africa" *Univ. of California Press, Los Angeles*
- Deutch, K. and W. Foltz (1963) "Nation building" *Atherton press, New York*
- Easterly, W and R. Levine (1997) "Africa's Growth Tragedy: Policies and Ethnic Divisions" *Quarterly Journal of Economics*, Vol. 112, Issue 4
- Fernandez, R, A Fogli and C. Olivetti "Marrying your mom: Preference transmission and Womens' Labor and Education choice" *NBER working paper*
- Gellner, E (1983), *Nations and Nationalism*, Oxford, Blackwell.
- Hoffman, C (2000). Balancing Language Planning and Language Rights Catalonia's Uneasy Juggling Act. *Journal of Multilingual and Multicultural Development*, 21 (5): 425-441.
- Janeba, E (2004) "International trade and cultural identity" *working paper*
- Johnson, H.G (1965). A theoretical model of economic nationalism in new developing states. *Political Science Quarterly*, 80: 169-85.
- Johnson, H.G (1975). *Technology and Economic Interdependence*, London McMillan.
- Konya I. "A dynamic model of cultural assimilation" *working paper*
- 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*

- Miguel T and Posner D. (2004) "Sources of ethnic identification in Africa" *working paper*
- Miguel E. "Tribe or nation? Nation-Building and Public goods in Kenya versus Tanzania" *World Politics*
- 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
- Pagano, U (1995). Can economics explain nationalisms? In Bretton, A , Galeotti, G, Salmon, P and Wintrobe, R (eds). *Nationalism and Rationality*, Cambridge University Press, Cambridge.
- Pagano, U (2003). Nationalism, development and integration: the political economy of Ernest Gellner. *Cambridge Journal of Economics*, 27: 623-646.
- Sen, A (1985). Goals, commitment and Identity. *Journal of Law, Economic and Organisation*, 1: 345-355.
- Tilly, C (1975) " The formation of national States in Western Europe" *Princeton University Press, Princeton*
- UN report (2004) "Building multicultural democracy"
- Watkins, M (1978). The Economics of Nationalism and the Nationality of Economics: A Critique of Neoclassical Theorizing. *Canadian Journal of Economics*, 11: 87-120.

Treatment by cohort and Education

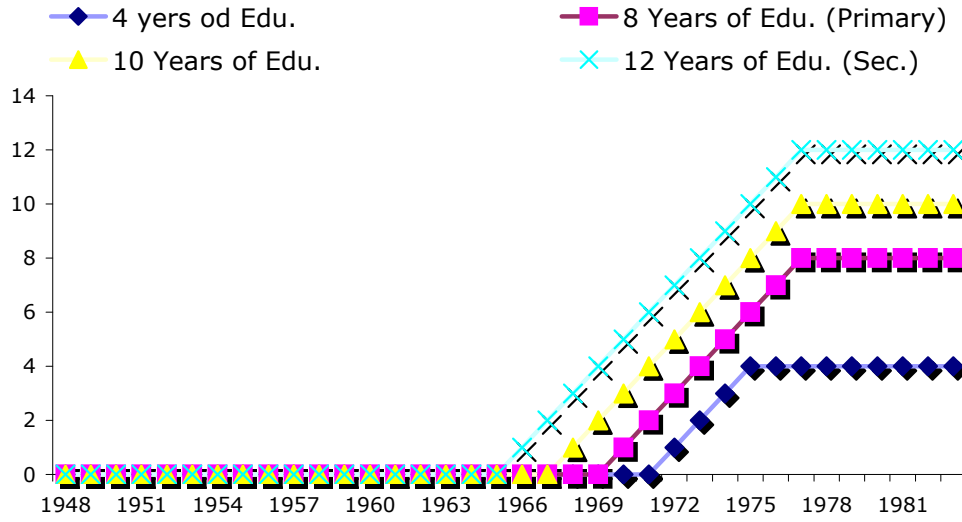


Figure 1: Treatment by Cohort and Years of Education

Presence of the Catalan language at school

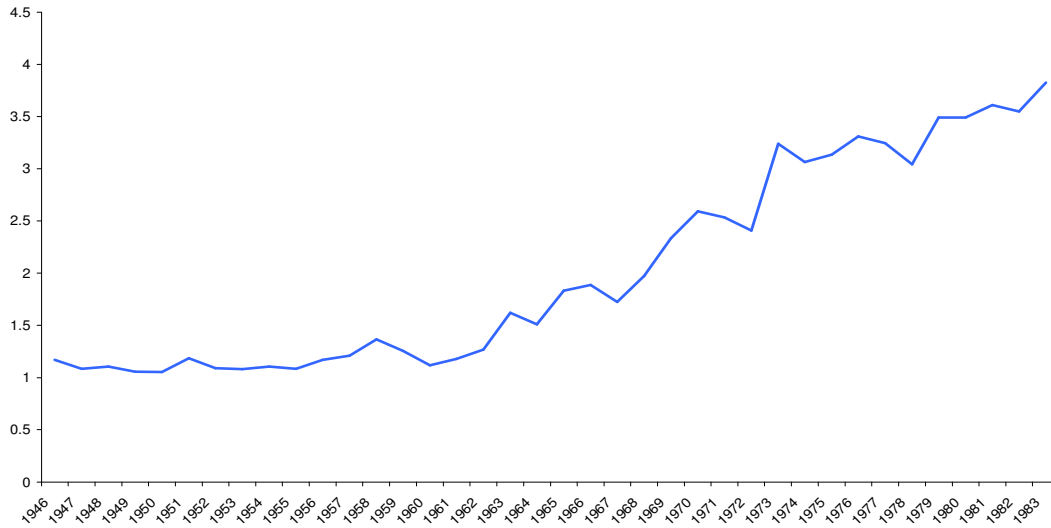


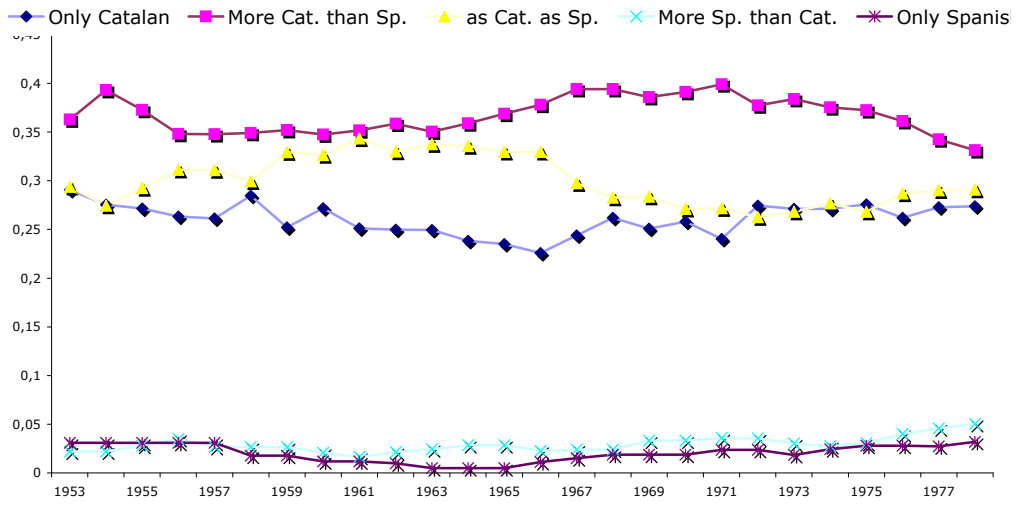
Figure 2: Catalan at School



		Average years of Education		
		Total Sample (1948-1983)	Control Cohorts (1948-1965)	Treated Cohorts (1966-1983)
Total	Average	10,22	9,81	10,53
	St. Dev.	1,90	1,94	1,82
	Observations	1490,00	626,00	864,00
Catalan Origin	Average	10,30	9,92	10,56
	St. Dev.	1,87	1,89	1,81
	Observations	1354,00	547,00	807,00
Cat. Origin with mixed Families	Average	10,34	9,90	10,55
	St. Dev.	1,81	1,86	1,76
	Observations	246,00	79,00	167,00
Cat. Origin and No Cat. Family	Average	10,04	9,62	10,23
	St. Dev.	1,93	1,89	1,92
	Observations	343,00	104,00	239,00
No Catalan Origin	Average	9,50	9,09	10,07
	St. Dev.	2,10	2,16	1,88
	Observations	136,00	79,00	57,00

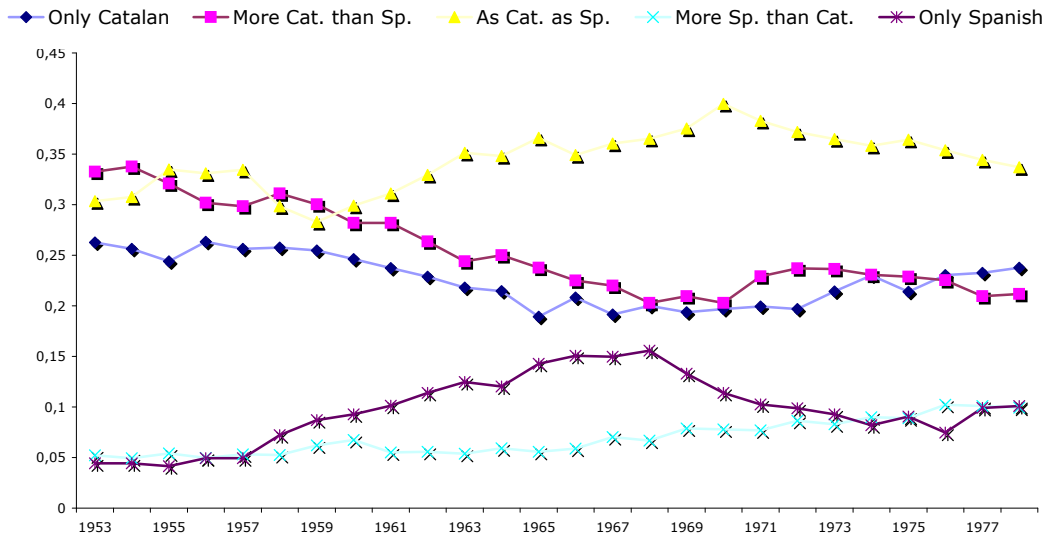
Table (1)

**Figure 3: National Sentiments of Catalans with Secondary Education by year of birth**



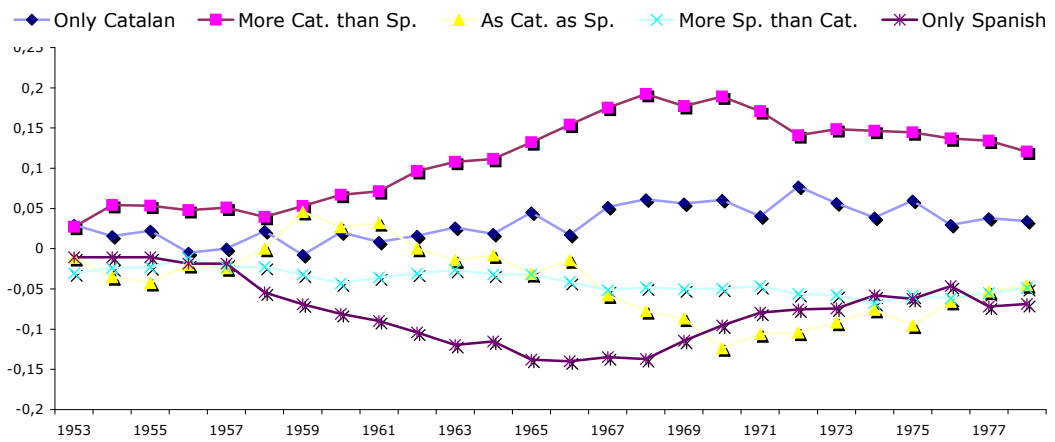
All series are (5,1,5) moving averages.

**Figure 4: National Sentiments of Catalans without Secondary Education by year of birth**



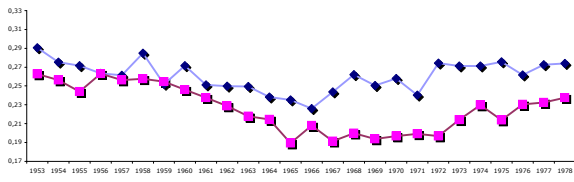
All series are (5,1,5) moving averages.

Figure 5: The effect of Education on the National Sentiments



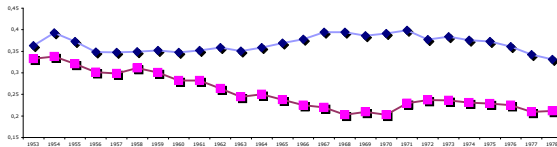
All series are (5,1,5) moving averages.

**Figure 6: Fraction of respondents that feel only Catalan by cohort and level of education**

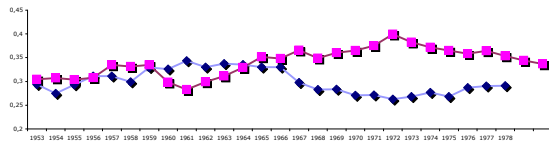


All series are (5,1,5) moving averages.

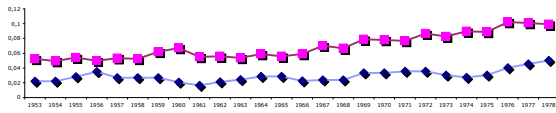
**Figure 7: Fraction of respondents that feel more Catalan than Spanish by cohort and level of Education**



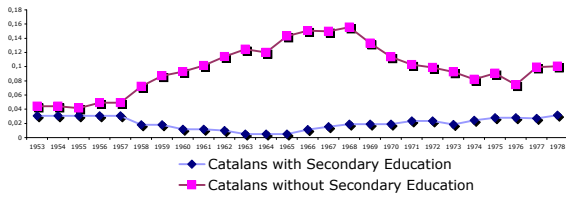
**Figure 8: Fraction of respondents that feel as Catalan as Spanish by cohort and level of Education**



**Figure 9: Fraction of respondents that feel more Spanish than Catalan by cohort and level of Education**

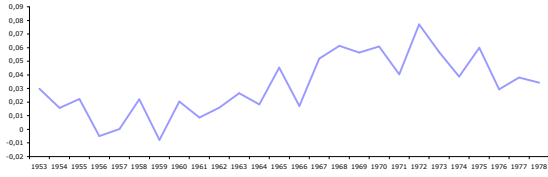


**Figure 10: Fraction of respondents that feel only Spanish by cohort and level of Education**

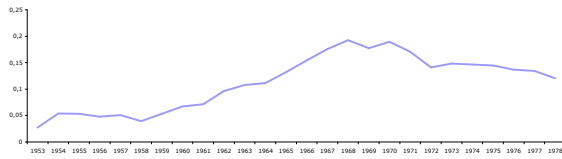


◆ Catalans with Secondary Education  
 ■ Catalans without Secondary Education

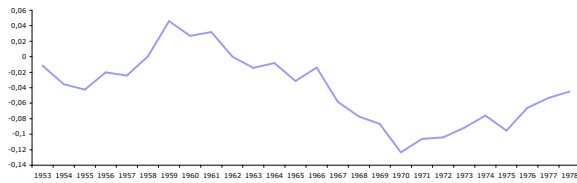
**Figure 11: The effect of education on feeling only Catalan by year of birth**



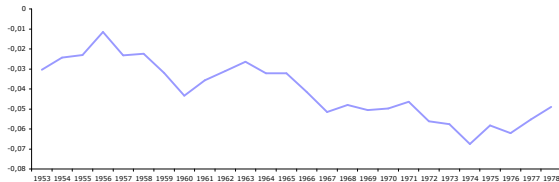
**Figure 12: The effect of education on feeling more Catalan than Spanish by year of birth**



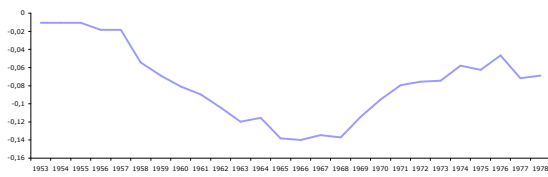
**Figure 13: The effect of education on feeling as Catalan as Spanish by year of birth**



**Figure 14: The effect of education on feeling more Spanish than Catalan by year of birth**



**Figure 15: The effect of education on feeling only Spanish by year of birth**



All series are (5,1,5) moving averages.

**Table 2**

		Cohorts (total sample)			Cohorts (with sec. edu)			Cohorts (without sec. edu)		
		1948-83	1948-65	1966-83	1948-83	1948-65	1966-83	1948-83	1948-65	1966-83
<b>Identity</b>	Average	3,51	3,56	3,47	3,7	3,73	3,69	3,29	3,42	3,16
	St. Dev.	1,04	1,03	1,05	0,91	0,9	0,92	1,12	1,1	1,13
	Observatio	1467	618	849	732	250	482	735	368	367
<b>Years of Education</b>	Average	10,33	9,92	10,61	12	12	12	8,45	8,29	8,61
	St. Dev.	1,88	1,97	1,77	0	0	0	0,97	0,96	0,95
	Observatio	1490	626	864	749	255		741	371	370
<b>Cat. Origin, Cat. Family</b>	Average	0,43	0,48	0,40	0,49	0,56	0,45	0,37	0,42	0,32
	St. Dev.	0,49	0,50	0,49	0,5	0,49	0,49	0,48	0,49	0,46
	Observatio	1490	626	864	749	255	494	741	371	370
<b>Cat. Origin, mixed Family</b>	Average	0,17	0,14	0,19	0,17	0,14	0,19	0,16	0,14	0,18
	St. Dev.	0,37	0,35	0,39	0,38	0,35	0,39		0,35	0,39
	Observatio	1490	626	864	749	255	494	741	371	370
<b>Cat. Origin, no Cat. Family</b>	Average	0,28	0,20	0,33	0,24	0,15	0,28	0,32	0,23	0,39
	St. Dev.	0,44	0,40	0,47	0,42	0,36	0,45	0,46	0,42	0,49
	Observatio	1490	626	864	749	255	494	741	371	370
<b>No Catalan Origin</b>	Average	0,11	0,16	0,07	0,08	0,13	0,06	0,13	0,19	0,08
	St. Dev.	0,31	0,37	0,26	0,28	0,33	0,24	0,34	0,39	0,28
	Observatio	1490	626	864	749	255	494	741	371	370
<b>Catalan at school</b>	Average	2,15	1,23	2,78	2,44	1,34	3,01	1,82	1,15	2,46
	St. Dev.	1,35	0,71	1,33	1,39	0,85	1,27	1,23	0,57	1,35
	Observatio	1487	625	862	747	255	492	740	370	370
<b>Cat. at school* years edu (Language exposure)</b>	Average	3,13	0,63	4,86	4,33	1,03	6,04	1,79	0,32	3,17
	St. Dev.	3,78	1,98	3,77	4,18	2,55	3,81	2,72	1,28	2,98
	Observatio	1487	625	862	747	255	492	740	370	370
<b>Catalan Social use</b>	Average	4,12	4,45	3,91	4,81	5,13	4,66	3,3	3,87	2,81
	St. Dev.	2,70	2,68	2,72	2,34	2,22	2,39	2,86	2,82	2,8
	Observatio	1340	534	806	702	230	472	638	304	334
<b>Language at home</b>	Average	0,44	0,49	0,40	0,5	0,56	0,47	0,37	0,44	0,29
	St. Dev.	0,49	0,50	0,49	0,5	0,49	0,49	0,48	0,49	0,45
	Observatio	1480	622	858	743	252	491	737	370	367
<b>Intensity of treatment</b>	Average	4,02	0,00	6,78	5,37	0	8,12	2,49	0	4,83
	St. Dev.	4,63	0,00	4,18	4,99	0	3,91	3,63	0	3,78
	Observatio	1490	626	864	749	255	494	741	371	370
<b>Intensity of (compulsory) treatment</b>	Average	2,68	0,00	4,52	3,02	0	4,57	2,29	0	4,45
	St. Dev.	3,51	0,00	3,53	3,53	0	3,43	3,45	0	3,67
	Observatio	1490	626	864	749	255	494	741	371	370
<b>Female</b>	Average	0,49	0,49	0,49	0,49	0,43	0,51	0,49	0,53	0,45
	St. Dev.	0,50	0,50	0,50	0,5	0,49	0,5	0,5	0,49	0,49
	Observatio	1490	626	864	749	255	494	741	371	370

**Table 3: Reform and Identity***Dependent variable: Identity*

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

\* Significant at the 10%, \*\* significant at the 5%, \*\*\* significant at the 1%. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. 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.

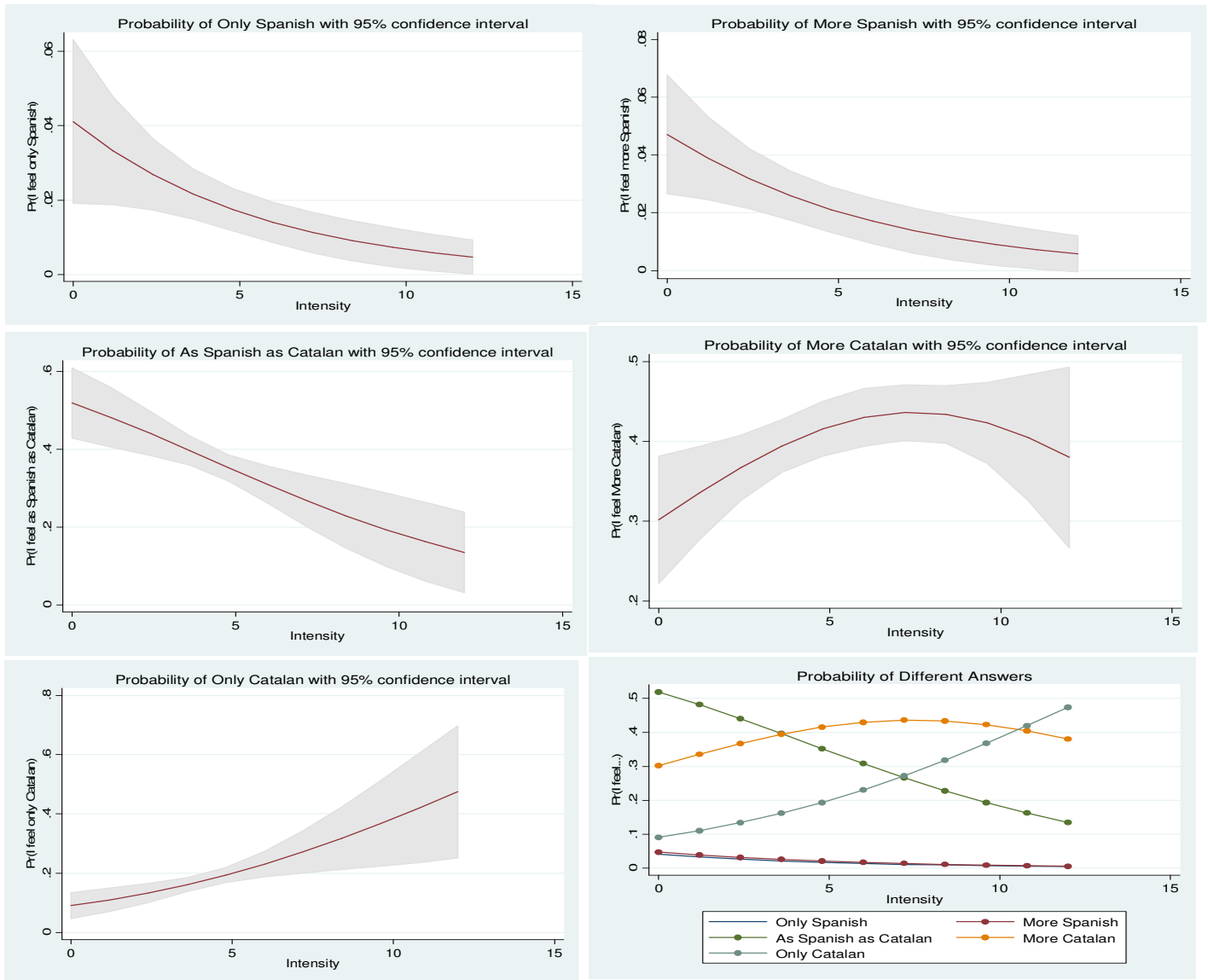


Figure 16



**Table 4: Channels***Dependent variable: Identity*

	Language exposure		Use of Catalan		2SLS		2SLS	
	OLS [1]	Ord. Logit [2]	OLS [3]	Ord. Logit [4]	First stage [5]	First Stage [6]	[7]	[8]
Language exposure	0.032*** (.01)	0.075*** (.024)					.113*** (.262)	
use of Catalan			.193*** (.013)	.515*** (.04)				.508*** (.161)
Intensity					.695*** (.08)	.179** (.08)		
INDIVIDUAL CONTROLS	YES	YES	YES	YES	YES	YES	YES	YES
YEARS OF EDUCATION	YES	YES	YES	YES	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES	YES	YES	YES	YES
PROVINCES	YES	YES	YES	YES	YES	YES	YES	YES
No. obs.	1464	1464	1321	1321	1464	1321	1464	1321
Rsq	0.329		0.457		0.524	0.519	0.287	
Pseudo Rsq		0.152		0.232				0.138

\* Significant at the 10%, \*\* significant at the 5%, \*\*\* significant at the 1%. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. 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.

**Table 5: Controlling for exogenous trends**

*Dependent variable: Identity*

	OLS [1]	Ord. Logit [2]	OLS [3]	Ord. Logit [4]	OLS [5]
intensity			.115*** (.037)	.232** (.93)	
pseudo-intensity	-.015 (.032)	-.07 (.075)	-.022 (.036)	-.024 (.094)	.023 (.196)
INDIVIDUAL CONTROLS	YES	YES	YES	YES	YES
YEARS OF EDUCATION	YES	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES	YES
PROVINCES	YES	YES	YES	YES	YES
	1930-1965	1930-1965	1921-1980	1921-1980	1948-1983
No. obs.	1008	1008	1860	1860	155
Rsq	.291		.295		.350
Pseudo Rsq		.131		.135	

\* Significant at the 10%, \*\* significant at the 5%, \*\*\* significant at the 1%. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. 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.

**Table 6: First stage**

Dependent variable: Intensity

	<u>OLS</u>
	<u>[1]</u>
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
YEARS OF BIRTH	YES
PROVINCES	YES
AGE and AGE SQUARED	YES
No. obs.	1467
Rsq	.951

\* Significant at the 10%, \*\* significant at the 5%, \*\*\* significant at the 1%. Robust standard errors are reported between parenthesis and are clustered at the cohort level. 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.

**Table 7: Controlling for endogeneity of years of education**

*Dependent variable: Identity*

	OLS [1]	Ord. Logit [2]	2SLS [3]	OLS [4]	Ord. Logit [5]	OLS [6]	Ord. Logit [7]	OLS [8]
intensity	<b>.066**</b> (.026)	<b>.158***</b> (.059)	<b>.054**</b> (.023)			<b>.099***</b> (.034)	<b>.249***</b> (.074)	
intensity compulsory				<b>.056**</b> -0,025	<b>.105**</b> (.049)			
language at home	.688*** (.096)	1.669*** (.235)						
intensity (people arriving)								<b>.069**</b> (.038)
AGE AT ARRIVAL	NO	NO	NO	NO	NO	NO	NO	YES
CONTROLLING FOR CDF	NO	NO	NO	NO	NO	YES	YES	NO
INDIVIDUAL CONTROLS	YES	YES	YES	YES	YES	YES	YES	YES
YEARS OF EDUCATION	YES	YES	OLS	YES	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	NO	NO	NO	YES	YES	YES
PROVINCES	YES	YES	YES	YES	YES	YES	YES	YES
AGE and AGE SQUARED	NO	NO	YES	YES	YES	NO	NO	NO
No. obs.	1467	1467	1467	1467	1467	1467	1467	361
Rsquared	.361		.299	.296		.325		.523
Pseudo Rsquared		.171			.135		.149	

\* Significant at the 10%, \*\* significant at the 5%, \*\*\* significant at the 1%. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level for columns 1 and 2, at the cohort level for columns 3-9 and at the cohort-year of arrival level incolumns 10-12. 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

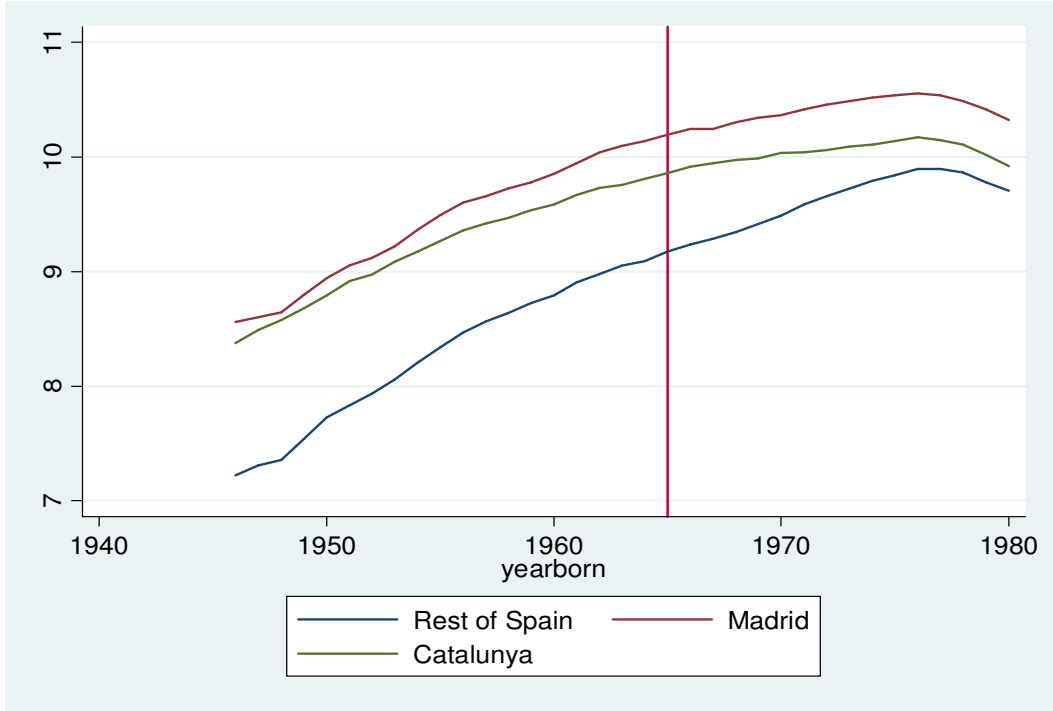


Figure 17

**Table 8: Other robustness check: Restricting the sample**

	<i>OLS</i> [1]	<i>Ord. Logit</i> [2]	<i>OLS</i> [3]	<i>Ord. Logit</i> [4]
intensity	0.108*** (.027)	0.249*** (.064)	0.115*** (.031)	0.289*** (.072)
SAMPLE	1951-1980	1951-1980	1956-1975	1956-1975
INDIVIDUAL CONTROLS	YES	YES	YES	YES
YEARS OF EDUCATION	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES
PROVINCES	YES	YES	YES	YES
No. obs.	1245	1245	853	853
Rsq	0.318		0.354	
Pseudo Rsq		0.149		0.166

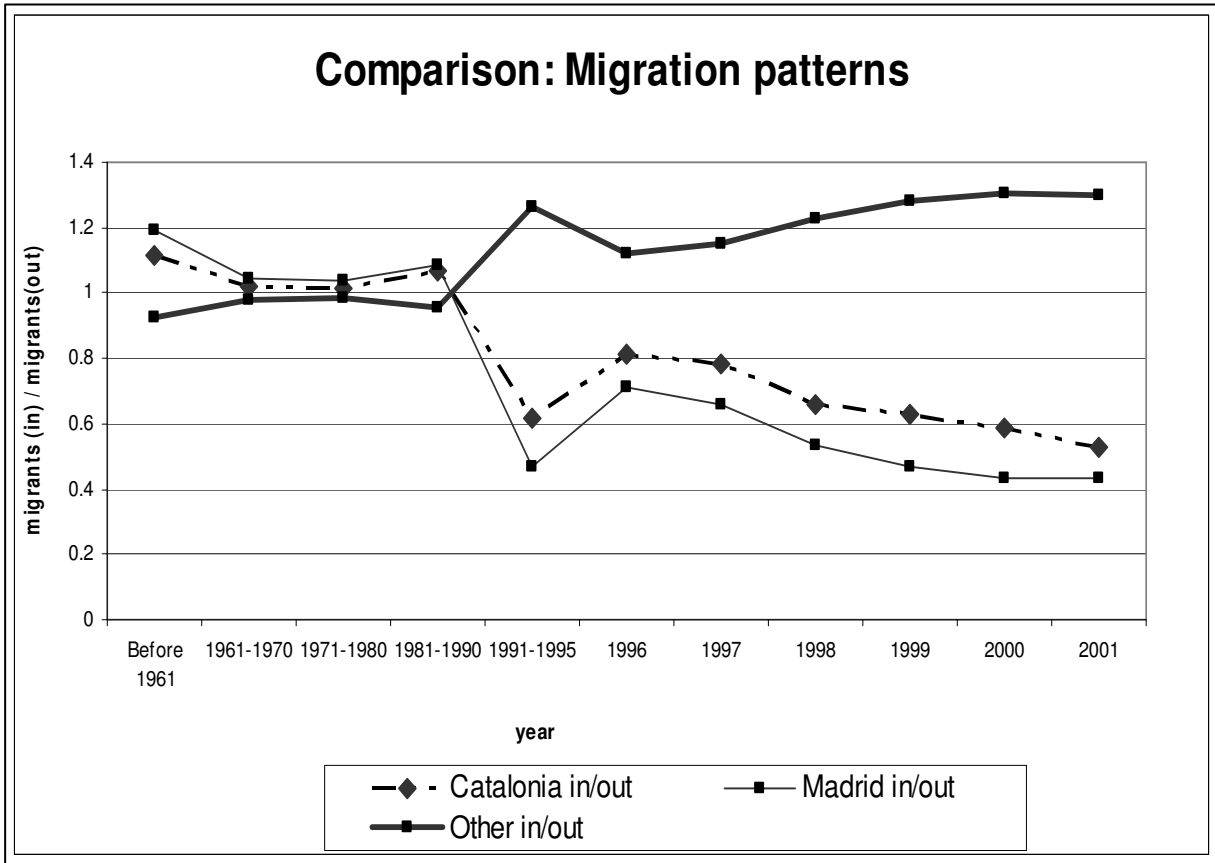
\* Significant at the 10%, \*\* significant at the 5%, \*\*\* significant at the 1%. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. Individual controls include a dummy variable for gender, for individuals not born in Catalonia, individuals born in Catalonia but with non-catalan parents and individuals born in Catalonia with mixed parents.

**Table 9: Other robustness checks: Migration.**

*Dependent variable: Identity*

	<i>OLS</i> [1]	<i>Ord. Logit</i> [2]	<i>OLS</i> [3]	<i>Ord. Logit</i> [4]
intensity	.069** (.027)	0.164*** (.063)	.07** (.033)	0.18*** (.079)
SAMPLE	1948-1983	1948-1983	1953-1978	1953-1978
	catorigin	catorigin	catorigin	catorigin
INDIVIDUAL CONTROLS	YES	YES	YES	YES
YEARS OF EDUCATION	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES
PROVINCES	YES	YES	YES	YES
No. obs.	1333	1333	994	994
Rsq	0.288		0.287	
Pseudo Rsq		0.135		0.138

\* Significant at the 10%, \*\* significant at the 5%, \*\*\* significant at the 1%. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. 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.



**Figure 18: Migration**

**Table 10: Effect for each level of treatment***Dependent variable: Identity*

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

\* Significant at the 10%, \*\* significant at the 5%, \*\*\* significant at the 1%. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. 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.



### Non-linear Effects of the Reform: Changes in probability

Legend:  
◆ Only Spanish  
■ More Spanish  
▲ As Spanish as Catalan  
✕ More Catalan  
✱ Only Catalan

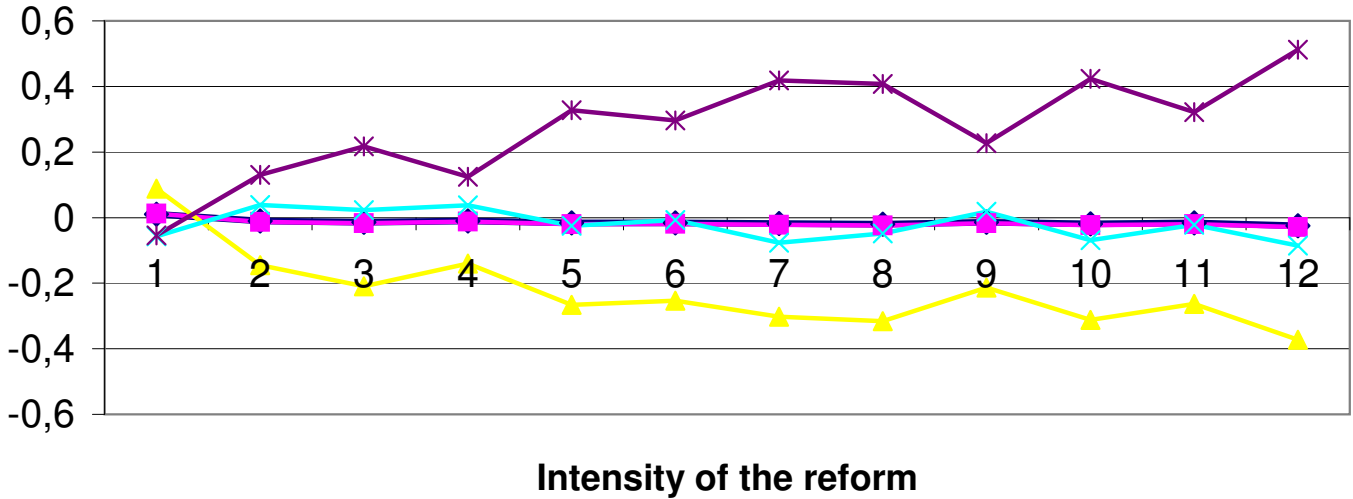


Figure 19

**Table 11: The role of individual origin***Dependent variable: Identity*

	<i>OLS</i> [1]	<i>Ord. Logit</i> [2]
intensity*non cat.origin	.062* (.035)	.115 (.078)
intensity*cat.family	0.086*** (.026)	0.197*** (.06)
intensity*mixed family	0.081*** (.029)	0.172*** (.065)
intensity*non cat.family	0.075*** (.028)	0.163** (.063)
1 catalan origin	-1.272*** -0,114	-2.835*** (.272)
mixed family	-.472*** (.086)	-.976*** (.184)
1 catalan family	-.913*** (.095)	-2.047*** (.210)
YEARS OF EDUCATION	YES	YES
YEARS OF BIRTH	YES	YES
PROVINCES	YES	YES
No. obs.	1467	1467
Rsq	.325	
Pseudo Rsq		.15

\* Significant at the 10%, \*\* significant at the 5%, \*\*\* significant at the 1%. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. 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.