

Bilingual education and identity

Ramon Caminal

Antonio Di Paolo

Ada Ferrer-i-Carbonell

Work in progress

Introduction

- Main theme: effect of the education reform (introduction of Catalan as a language of instruction) implemented in Catalonia in the 1980's on identity formation and political preferences
- Potential effects of the linguistic reform:
 - Language skills
 - Identity: nation building, interethnic conflict
- Skills and identity could be connected (Akerlof and Kranton, 2000): lower costs of acquiring a language that is a marker of identity stimulate the acquisition of such identity

Introduction

- Previous work: Cappellari and Di Paolo (2018), Caminal and Di Paolo (2019), Caminal et al. (2021): show that the exposure to the education reform of 1983 improved the oral skills of native Spanish speakers with implications for labor market outcomes, mixed marriages, and intergenerational transmission of languages.
- This paper: are skills and identity connected?

Introduction

- The construction of nation states in the XIX and XX century: homogenization of a diverse population: Alesina et al, 2000, 2020), Bandiera (2019)
 - state-controlled mass education
 - Exclusive use of the "national" language in school
 - compulsory military service
 - building infrastructures
- Language shift (one country, one language)

Introduction

- Hauk and Ortega (2019): the XIX century Spanish authorities did not invest enough in mass education, as the interests of the country elite and the regional elites diverged, which possibly explain the relative vitality of regional languages in Spain.
- Clots-Figueras and Masella (2013): the introduction of Catalan as medium of instruction in Catalonia fostered (?) the adoption of the Catalan identity.

Introduction

- Catalonia: education reform of 1983 (LNA): shift from a monolingual system (Spanish only) to a bilingual one, where Catalan has become the main language of instruction (at least in primary education).
- Context:
 - Broad political support to the reform in Catalonia
 - In the 1980's more than one half of the population were native Spanish speakers (hardly a minority)
 - Regional government under the supervision of the central government (including the judicial system)
 - Effect of the school can be counterbalanced by the reaction of the families (Fouka, 2020)

Outline

- Main focus: Effect of the 1983 reform on language identity (EULP, IDESCAT)
- Additional evidence on: effect on national identity (feeling of belonging) and political preferences (support for independence), BoPO,

Linguistic identity

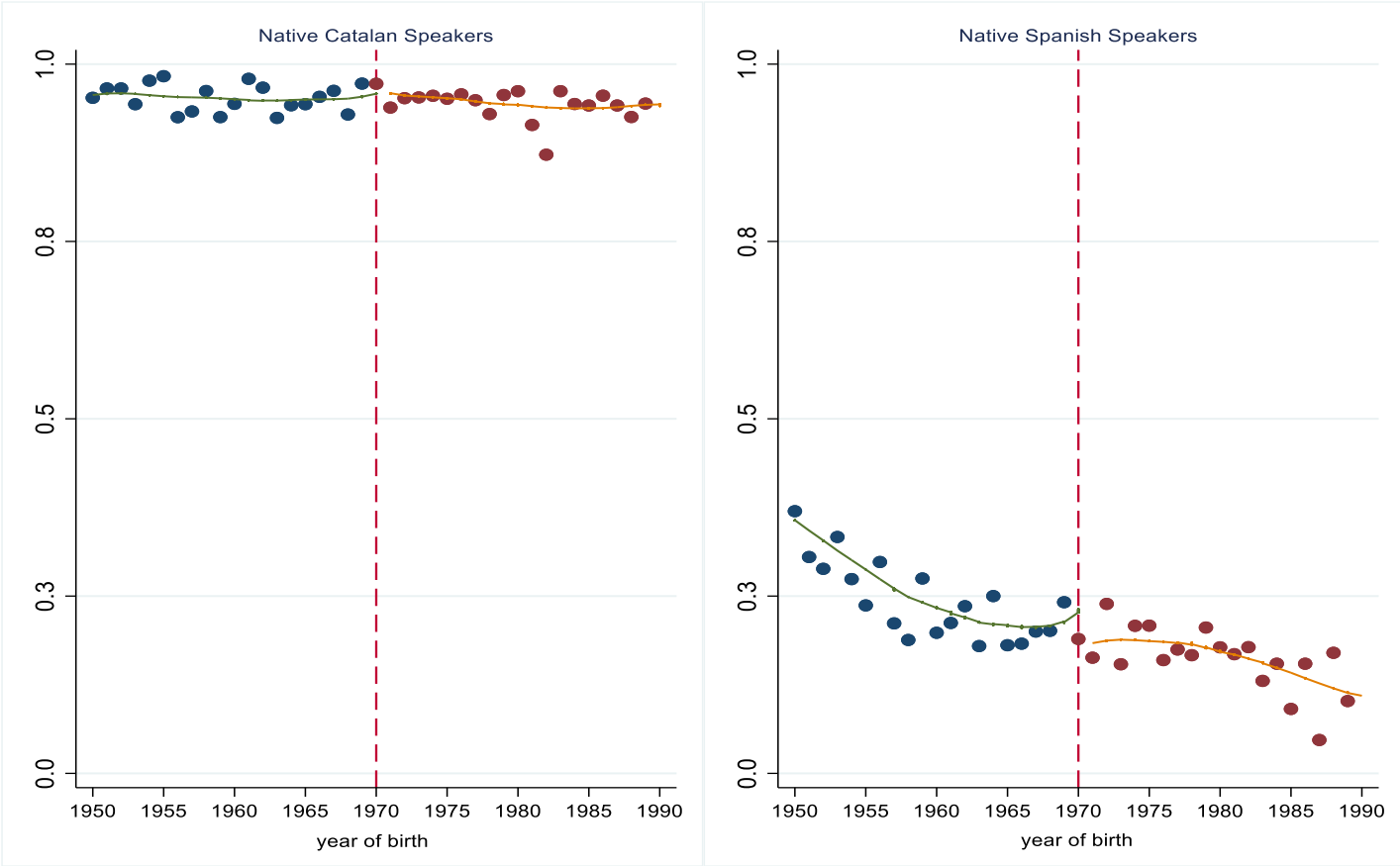
- Survey of Language Use of the Catalan Population (Catalan Institute of Statistics)
- Three cross-sections: waves 2008, 2013 and 2018.
- Language of self-identification (which language you feel more identified with)
- Restricted sample:

Descriptive statistics (1)

Variable	Pooled Sample		Native Catalan Speakers		Native Spanish Speakers		wave 2008		wave 2013		wave 2018	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
	Catalan as self-identification language	0.578	0.494	0.950	0.218	0.198	0.399	0.581	0.494	0.599	0.490	0.558
Native Spanish speaker	0.494	0.500					0.495	0.500	0.478	0.500	0.508	0.500

Descriptive statistics (2)

Y axis: Catalan as self-identification language (0-1)



Empirical methodology (1)

- (Potential) exposure to bilingualism during compulsory education:

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

→ The duration of compulsory education increased until age 16 (10 years) for cohorts born since 1983.

- Main equation (as in CF-M):

$$Y_i = \alpha + \beta'X_i + \pi SPA_i + \delta E_{\tau(i)} \times SPA_i + \theta_{\tau(i)} + \varepsilon_i$$

→ $SPA_i = 1$, if native Spanish speaker; θ_{τ} = year of birth fixed effect

→ Baseline specification: X_i includes year-wave effects, a 2nd order polynomial in age and gender. We additionally include parents' and individual's control variables.

→ Standard errors clustered at the year of birth level.

Empirical methodology (2)

- Robustness checks:
 - Different specification of the age polynomial: 0, 1, 2, 3, 4
 - Different definitions of exposure to the LNA reform
 - Inclusion of year of birth fixed effects specific by parental origins (at least one born in Catalonia vs both born outside Catalonia):
 - It relaxes the main underlying assumption that birth cohort effects are common for Catalan and Spanish speakers.
 - Use of alternative proxies of native language: language of the interview and parents' language (at least one Catalan vs both Spanish)-
 - Exclusion of individuals with Catalan and Spanish as mother tongue and mixed couples.

Empirical methodology (3)

- Analysis of heterogeneous effects: wave/year, gender, place of residence (BMA vs rest of Catalonia), individual and parents place of birth (BMA, rest of Catalonia, rest of Spain), plus others (not significant).

Results (1)

Dependent Variable: Catalan as self-identification language (0-1)

	(1)	(2)	(3)	(4)
native Spanish speaker	-0.716*** (0.012)	-0.579*** (0.018)	-0.647*** (0.013)	-0.549*** (0.018)
exposure × native Spanish speaker	-0.008*** (0.002)	-0.011*** (0.002)	-0.010*** (0.002)	-0.012*** (0.002)
parents' control variables	no	yes	no	yes
individual's control variables	no	no	yes	yes
adjusted R-squared	0.582	0.595	0.598	0.608
number of observations	10045	10045	10045	10045

Note: OLS, dependent variable: Catalan as self-identification language (vs Spanish or Spanish & Catalan). Standard errors clustered by year of birth in parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. Native Spanish speaker = 1 if Spanish or Spanish and Catalan as mother tongue (vs only Catalan). All regression contains wave dummies, year of birth dummies, gender and age (2nd order polynomial) as baseline controls. Parents' control: dummies for father's and mother's place of birth, highest parental education. Individual's controls: dummies for place of residence and place of birth, completed education.

Results (2)

Dependent Variable: Catalan as self-identification language (0-1)

Robustness to the choice of the age polynomial

	(1)	(2)	(3)	(4)	(5)
order of the age polynomial:	0	1	2	3	4
native Spanish speaker	-0.549*** (0.018)	-0.548*** (0.018)	-0.549*** (0.018)	-0.549*** (0.018)	-0.549*** (0.018)
exposure × native Spanish speaker	-0.012*** (0.002)	-0.012*** (0.002)	-0.012*** (0.002)	-0.012*** (0.002)	-0.012*** (0.002)
age		-0.001 (0.006)	0.011 (0.007)	0.008 (0.014)	-0.030 (0.044)
age ²			-0.000** (0.000)	-0.000 (0.000)	0.001 (0.002)
age ³				-0.000 (0.000)	-0.000 (0.000)
age ⁴					0.000 (0.000)
adjusted R-squared	0.608	0.608	0.608	0.608	0.608
number of observations	10045	10045	10045	10045	10045

Note: OLS, dependent variable: Catalan as self-identification language (vs Spanish or Spanish & Catalan). Standard errors clustered by year of birth in parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. Native Spanish speaker = 1 if Spanish or Spanish and Catalan as mother tongue (vs only Catalan). All regression contains the full set of control variables (as in column 4 of table 1).

Results (3)

Dependent Variable: Catalan as self-identification language (0-1)

Robustness to different definition of exposure to the LNA

	(1)	(2)	(3)
native Spanish speaker	-0.549*** (0.018)	-0.553*** (0.017)	-0.506*** (0.027)
exposure × native Spanish speaker	-0.012*** (0.002)	-0.011*** (0.002)	
year birth 1950-1960		reference category	
year birth 1961-1965		-0.010 (0.015)	0.021 (0.013)
year birth 1966-1969		-0.013 (0.018)	0.045** (0.018)
year birth 1970-1976 (partial exposure)		0.031 (0.023)	0.081*** (0.022)
year birth 1977-1990 (full exposure)		0.055* (0.032)	0.118*** (0.033)
year birth 1961-1965 × native Spanish speaker			-0.049* (0.027)
year birth 1966-1969 × native Spanish speaker			-0.086*** (0.025)
year birth 1970-1976 × native Spanish speaker			-0.100*** (0.026)
year birth 1977-1990 × native Spanish speaker			-0.151*** (0.028)
adjusted R-squared	0.608	0.608	0.609
number of observations	10045	10045	10045

Note: OLS, dependent variable: Catalan as self-identification language (vs Spanish or Spanish & Catalan). Standard errors clustered by year of birth in parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. Native Spanish speaker = 1 if Spanish or Spanish and Catalan as mother tongue (vs only Catalan). All regression contains the full set of control variables (as in column 4 of table 1).

Results (4)

Dependent Variable: Catalan as self-identification language (0-1)

Robustness to year of birth FEs specific by parental origins

	(1)	(2)
native Spanish speaker	-0.549*** (0.018)	-0.532*** (0.018)
exposure × native Spanish speaker	-0.012*** (0.002)	-0.016*** (0.003)
adjusted R-squared	0.608	0.611
number of observations	10045	10045

Note: OLS, dependent variable: Catalan as self-identification language (vs Spanish or Spanish & Catalan). Standard errors clustered by year of birth in parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. Full set of controls. Column (1): common year of birth dummies. Column (2) year of birth dummies specific by parents' place of birth (at least one born in Catalonia vs both born outside Catalonia).

Results (5)

Dependent Variable: Catalan as self-identification language (0-1)

Robustness to alternative proxies of native language, the exclusion of individual with Catalan & Spanish as mother tongue and to the exclusion of mixed couples

	baseline	interview	par. lang.	mix. lang.	mix couple
native Spanish speaker	-0.549*** (0.018)			-0.603*** (0.017)	-0.703*** (0.017)
exposure × native Spanish speaker	-0.012*** (0.002)			-0.011*** (0.002)	-0.008*** (0.002)
interview in Spanish or Spanish & Catalan		-0.292*** (0.008)			
exposure × interview in Spanish or Spanish & Catalan		-0.004*** (0.001)			
both parents Spanish speakers			-0.230*** (0.016)		
exposure × both parents Spanish speakers			-0.012*** (0.002)		
adjusted R-squared	0.608	0.504	0.492	0.641	0.732
number of observations	10045	10035	10045	9628	7484

Heterogeneous Effects (1)

Dependent Variable: Catalan as self-identification language (0-1)

Heterogeneous effects by year/wave

	(1)	(2)
native Spanish speaker	-0.549*** (0.018)	-0.552*** (0.018)
exposure × native Spanish speaker	-0.012*** (0.002)	-0.016*** (0.003)
wave 2008	reference category	
wave 2013	-0.016 (0.032)	-0.015 (0.033)
wave 2018	-0.037 (0.062)	-0.021 (0.063)
wave 2013 × native Spanish speaker		0.024 (0.024)
wave 2018 × native Spanish speaker		-0.014 (0.027)
wave 2013 × exposure × native Spanish speaker		0.001 (0.004)
wave 2018 × exposure × native Spanish speaker		0.009* (0.005)
adjusted R-squared	0.608	0.608
number of observations	10045	10045

Heterogeneous Effects (2)

Dependent Variable: Catalan as self-identification language (0-1)

Heterogeneous effects by gender

	(1)	(2)
native Spanish speaker	-0.549*** (0.018)	-0.551*** (0.018)
exposure × native Spanish speaker	-0.012*** (0.002)	-0.009*** (0.003)
male	-0.006 (0.006)	0.006 (0.007)
male × native Spanish speaker		0.003 (0.014)
male × exposure × native Spanish speaker		-0.007*** (0.002)
adjusted R-squared	0.608	0.609
number of observations	10045	10045

Heterogeneous Effects (3)

Dependent Variable: Catalan as self-identification language (0-1)

Heterogeneous effects by place of residence (Metropolitan Area vs rest of Catalonia)

	(1)	(2)
native Spanish speaker	-0.549*** (0.018)	-0.477*** (0.022)
exposure × native Spanish speaker	-0.012*** (0.002)	-0.018*** (0.003)
living in BMA		0.080** (0.030)
living in BMA × native Spanish speaker		-0.126*** (0.021)
living in BMA × exposure × native Spanish speaker		0.009*** (0.003)
adjusted R-squared	0.608	0.610
number of observations	10045	10045

Heterogeneous Effects (4)

Dependent Variable: Catalan as self-identification language (0-1)

Heterogeneous effects place of birth (Metropolitan Area vs rest of Catalonia vs rest of Spain)

	(1)	(2)
native Spanish speaker	-0.549*** (0.018)	-0.461*** (0.030)
exposure × native Spanish speaker	-0.012*** (0.002)	-0.020*** (0.004)
born in Catalonia (outside BMA)	reference category	
born in BMA		-0.017 (0.046)
born in other Spanish regions		0.089 (0.232)
born in BMA × native Spanish speaker		-0.126*** (0.032)
born in other Spanish regions × native Spanish speaker		-0.097 (0.098)
born in BMA × exposure × native Spanish speaker		0.011** (0.004)
born in other Spanish regions × exposure × native Spanish speaker		0.013** (0.006)
adjusted R-squared	0.608	0.610
number of observations	10045	10045

Heterogeneous Effects (5)

Dependent Variable: Catalan as self-identification language (0-1)

Heterogeneous effects by parents' place of birth (Metropolitan Area vs rest of Catalonia vs rest of Spain)

	(1)	(2)
native Spanish speaker	-0.549*** (0.018)	-0.461*** (0.030)
exposure × native Spanish speaker	-0.012*** (0.002)	-0.020*** (0.004)
parents born in Catalonia (outside BMA)	reference category	
parents born in BMA		-0.017 (0.046)
parents born in other Spanish regions		0.089 (0.232)
parents born in BMA × native Spanish speaker		-0.126*** (0.032)
parents born in other Spanish regions × native Spanish speaker		-0.097 (0.098)
parents born in BMA × exposure × native Spanish speaker		0.011** (0.004)
parents born in other Spanish regions × exposure × native Spanish speaker		0.013** (0.006)
adjusted R-squared	0.608	0.610
number of observations	10045	10045

National identity and political preferences

- Barometer of Political Opinion, conducted by the Catalan Public Opinion Center
- All waves between 2014 and 2020 + 2022 (multiple waves per year).
- New identity variables (feeling of belonging, national identity) and political preferences (preference about the organization of the state, independence)
- Correlation between language and national identity is at most 0.54.

Additional Results (1)

Additional results using BOP data:

Dependent Variables: column (1) Catalan as self-identification language (vs Spanish or Spanish & Catalan); column (2) national identity feelings (1-5 scale); column (3) feeling only Catalan; column (4) feeling only Catalan or more Catalan than Spanish; column (5) feeling at least as Catalan than Spanish; column (6) national identity feeling (0-10 scale)

	(1)	(2)	(3)	(4)	(5)	(6)
native Spanish speaker	-0.715*** (0.010)	-0.982*** (0.021)	-0.408*** (0.010)	-0.452*** (0.010)	-0.081*** (0.004)	-0.553*** (0.058)
exposure × native Spanish speaker	-0.006*** (0.001)	-0.011*** (0.003)	-0.003* (0.002)	-0.008*** (0.002)	-0.001 (0.001)	-0.011 (0.012)
adjusted R-squared	0.581	0.273	0.254	0.254	0.038	0.157
number of observations	0.581	0.273	0.199	0.254	0.038	0.157

Note: OLS, dependent variables: column (1) Catalan as self-identification language (vs Spanish or Spanish & Catalan); column (2) national identity feelings (1-5 scale); column (3) feeling only Catalan; column (4) feeling only Catalan or more Catalan than Spanish; column (5) feeling at least as Catalan than Spanish; column (6) national identity feeling (0-10 scale). Standard errors clustered by year of birth in parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. Native Spanish speaker = 1 if Spanish or Spanish and Catalan as mother tongue (vs only Catalan). All regression contains wave dummies, year of birth dummies, gender and age (2nd order polynomial) as controls.

Additional Results (2)

Additional results using BOP data:

Dependent Variables: column (1) desire of independence of Catalonia; column (2) preference for Catalonia as Spanish region; column (3) preference for Catalonia as Autonomous Community; column (4) preference for Catalonia as Federal State; column (5) preference for Catalonia as Independent State

	(1)	(2)	(3)	(4)	(5)
native Spanish speaker	-0.427*** (0.010)	0.059*** (0.004)	0.228*** (0.007)	0.133*** (0.010)	-0.420*** (0.010)
exposure × native Spanish speaker	-0.006*** (0.002)	-0.001* (0.001)	0.006*** (0.001)	0.000 (0.002)	-0.005*** (0.002)
adjusted R-squared	0.216	0.022	0.106	0.031	0.201
number of observations	0.581	0.273	0.199	0.254	0.038

Note: OLS, dependent variables: column (1) desire of independence of Catalonia; column (2) preference for Catalonia as Spanish region; column (3) preference for Catalonia as Autonomous Community; column (4) preference for Catalonia as Federal State; column (5) preference for Catalonia as Independent State. Standard errors clustered by year of birth in parenthesis. *** significant at 1%, ** significant at 5%, * significant at 10%. Native Spanish speaker = 1 if Spanish or Spanish and Catalan as mother tongue (vs only Catalan). All regression contains wave dummies, year of birth dummies, gender and age (2nd order polynomial) as controls.

Conclusions

- Catalonia: Bilingual education has not been an effective tool to promote the Catalan identity or the political goals of the Catalanist movement.
- Exposure to the reform has been roughly neutral for native Catalan speakers and has actually reduced the propensity of native Spanish speakers to adopt Catalan as the language of self-identification or to feel Only Catalan or More Catalan than Spanish. Similarly, native Spanish speakers exposed to the reform are less likely to support the independence of Catalonia.
- Alternatively, the reform has improved Catalan skills among native Spanish speakers, with implications in the labor market, patterns of social interactions, etc.

Conclusions

- Thus, skills and identity are dissociated (which runs against the theories of endogenous identity formation)
- Interpretation 1: designers of the education reform never had the will or the means to promote the Catalan identity among native Spanish speakers
- Interpretation 2: a policy shift (education reform) perceived as a threat to own identity may trigger a reinforcement of alternative mechanisms of identity formation (intergenerational transmission of identity). Fouka (2020)