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

Document de Treball 2015/23 1/47
Working Paper 2015/23 1/47

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

Document de Treball 2015/11 1/47
Working Paper 2015/11 1/47

"Bilingual Schooling and Earnings: Evidence from a Language-in-Education reform"

Lorenzo Cappellari and Antonio di Paolo

Research Institute of Applied Economics

Regional Quantitative Analysis Research Group

Abstract

We exploit the 1983 language-in-education reform that introduced Catalan alongside Spanish as medium of instruction in Catalan schools to estimate the labour market value of bilingual education. Identification is achieved in a difference-in-differences framework exploiting variation in exposure to the reform across years of schooling and years of birth. We find positive wage returns to bilingual education and no effects on employment, hours of work or occupation. Results are robust to education-cohort specific trends or selection into schooling and are mainly stemming from exposure at compulsory education. We show that the effect worked through increased Catalan proficiency for Spanish speakers and that there were also positive effects for Catalan speakers from families with low education. These findings are consistent with human capital effects rather than with more efficient job search or reduced discrimination. Exploiting the heterogeneous effects of the reform as an instrument for proficiency we find sizeable earnings effects of skills in Catalan.

JEL classification: J24, J25, I28.
Keywords: bilingual education, returns to schooling, language-in-education reform, Catalonia

Lorenzo Cappellari. Department of Economics and Finance. Universitá Cattolica Milano, Largo Gemelli 1, 20123 Milano, Italy. E-mail: lorenzo.cappellari@unicatt.it

Antonio di Paolo. AQR Research Group-IREA. Department of Econometrics. University of Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain. E-mail: antonio.dipaolo@ub.edu

Acknowledgements

Funding from Càtedra Pasqual Maragall at Universitat de Barcelona, the MEC grant ECO2013-41022-R and the PRIN grant no. 2010T8XAXB is gratefully acknowledged. Thanks are due to Frederic Udina, Josep Maria Martínez and other IDESCAT staff for their assistance with the data. Ada Ferrer-i-Carbonell, Ramon Caminal, Albert Costa, Xavier Vila-i- Moreno, Germà Bel, Joan-Ramon Borrell and seminar participants at URV, UdG, UB, IRVAPP-Trento, SFI Copenhagen, IDESCAT, EALE-SOLE, ESPE, IWAEE, SAEe and the Lisbon Workshop on Economics and Econometrics of Education provided useful comments. The usual disclaimers apply.

1. Introduction

Language of instruction is a key input in the educational production function. While most students still receive education only in their native language, bilingual education (i.e. education taught in two languages) is becoming increasingly popular both among families and educationalists. Bilingualism affects the quality of education, favours the development of cognitive skills and its effectiveness in fostering educational outcomes has been found to increase with early exposure (Bleakley and Chin, 2008). Evidence in cognitive psychology and related disciplines shows that bilingualism has positive impacts on several dimensions of cognition, including a reduction of the rate of skill depletion in the adult population (see, for example, Adesope et al., 2010; Costa and Sebastian-Gallés, 2014). Policies introducing bilingual education are also a means for building a common identity in regions or countries where different nationalities coexist (Clots-Figueras and Masella, 2013). Despite the relevance of bilingual education for both school quality and skill formation, little is still known about its effects on economic outcomes. In this paper we contribute new evidence on this, by providing the first quasi-experimental estimates of the earnings effects of bilingual schooling.

We estimate the labour market value of bilingual education using a reform of the language of instruction in the Spanish region of Catalonia, the so-called 'Linguistic Normalization'. Until 1983, Spanish was the only official language in Catalan schools. Starting from that academic year, Catalan became a co-official language of instruction, which means that both Catalan and Spanish were used in education. Ours is not the first paper studying this reform, as other studies have investigated its effects on Catalan proficiency and on feelings of Catalan identity (see Rendon, 2007; Aspachs-Bracons et al., 2008; Clots-Figueras and Masella, 2013). We estimate the earnings effects of the 1983 reform for the first time. Our empirical framework is built around a difference-in-difference (DiD) estimator that compares the returns to schooling across cohorts of students that were differentially exposed to bilingual schooling as a consequence of the reform.

Catalan has been spoken in Catalonia since the Middle Ages, but was progressively dismissed with the Spanish domination of the eighteenth century and officially banned during Franco's dictatorship (1939–1975). The return of democracy in the late 1970s provided the political framework for a reintroduction of Catalan, also as a means of promoting the political autonomy of the region and its internal social cohesion. Increasing Catalan proficiency in the population became an important target for the new regional government, because a substantial share of the population was non-Catalan after the internal migration flows of the 1950s and 1960s, especially in the more industrialised areas of the region. More in general, after forty years of prohibition, proficiency in written Catalan was problematic among prime age and young individuals, irrespective of regional origins. Catalan became a compulsory subject in schools immediately after the end of dictatorship, but it was only with the 1983 law that a broader reform of the educational system towards bilingualism was put in place, through changes of the language of instruction in various subjects, teachers' turnover and immersion programmes in specific target areas.

There are two main channels through which exposure to bilingualism at school might be rewarded in the labour market. First, bilingualism can favour the development of cognitive skills, in this way increasing the stock of human capital accumulated per year of education – resulting in higher wage returns. Related to this, it may increase the quality of education if more skilled teachers are needed for bilingual teaching, inducing higher returns to schooling. Second, bilingual knowledge might increase search efficiency and increase the rate of arrival of job offers in a local labour market in which Catalan-speaking employers represent a relevant share of overall labour demand; therefore, leading to better job matches and higher wages. A specific example of this channel is provided by public sector jobs, for which, after the reform of 1983, the Catalan government gradually established proficiency in Catalan as a prerequisite. Effects through better search might also derive from better access to Catalan social networks, which might be particularly relevant for individuals with non-Catalan origins that increasingly perceived themselves as Catalans after the reform (see Aspachs-Bracons et al., 2008; Clots-Figueras and Masella, 2013), or might even mitigate employers' discrimination.

Our analytical approach is in the spirit of Angrist and Lavy (1997), who studied the wage effects of the 1983 'Arabization' policy in Morocco, which replaced French with Arabic as the language of instruction of post-primary education. They develop a DiD estimator across levels of education and birth cohorts and conclude that returns to schooling were substantially lowered for individuals exposed to the reform, a likely consequence of the loss of skills in French induced by 'Arabization'. A similar strategy is used by Angrist et al. (2008), who evaluated the effect of English-intensive instruction on English skills in Puerto Rico, exploiting the reform that in 1949 substituted English with Spanish as the language of instruction in all grades, finding no effect on English skills. In both cases, identification is achieved by comparing returns to education (in terms of either wages or language skills) between cohorts that were differentially exposed to the reforms. We apply this estimating framework to wage returns in Catalonia. Differently from the reforms studied in Angist and Lavy (1997) and Angrist et al. (2008), where policy changes were aimed at substituting one language with another, the Catalan reform made education bilingual, with the aim of fostering full proficiency in both Catalan and Spanish regardless of pupils' language background.

Using data from the Survey on the Living Conditions of the Catalan Population for 2006 and 2011, we find that one year of bilingual education generates on average an extra earnings return of

1.1 percent on top of the yearly return to education of 5.8 percent. The effect of bilingual education is not constant across levels of education, but is higher during the first five years of education (around 3.5 percent per year) and declines afterwards, becoming lower than 1 percent after 15 years of education. Conversely, we do not find any significant effect of the reform on employment, working hours or occupation. We subject these findings to a number of sensitivity checks. We control for selection into employment and for differences in life-cycle earnings trajectories across birth cohorts. We perform placebo analyses using both individuals belonging to the same birth cohorts from different non-bilingual Spanish regions and Catalan individuals belonging to older birth cohorts not affected by the reform. Evidence from all of these sensitivity checks points towards the robustness of our findings. We additionally consider whether the estimated positive effects are due to the endogenous choice of years of education and focus only on returns to compulsory education, still finding that bilingualism induces positive effects on earnings. Moreover, by allowing bilingual education to have differential effects across educational levels, we show that exposure during compulsory education is the driver of the overall effect of bilingual education.

In the last part of the paper we explore possible channels of the effects of bilingual education. We show that the reform did not affect the chances to work in the public sector. This is consistent with the evidence of no effects on employment or on occupations, suggesting that search is not the main mechanism through which the reform operated. We find a significant effect of the reform on proficiency in Catalan for the sub-population of non-Catalan origins, one of the main targets of the reform, which generated a small earnings premium. However, the baseline earnings effect is not accounted for by differential effects for non-Catalans, which does not support interpretations based on the reform reducing discrimination. Instead, the effects were concentrated among Catalans with low parental education, which is a possible effect of increased quality of human capital because of the reform. Exploiting the heterogeneous effects of the reform between Spanish and Catalan speakers as an instrument for proficiency, we find sizeable earnings effects of skills in Catalan.

The rest of the paper proceeds as follows. In Section 2 we review the literature on language reforms, bilingual education and the 1983 reform of the Catalan schooling system. Section 3 provides a detailed account of the institutional setting, which is key in the construction of the treatment variable for the analysis. Section 4 describes the data, while in Section 5 we lay out the empirical framework. Section 6 reports the main results, together with a number of sensitivity checks. Section 7 investigates the possible channels through which the reform had an effect, while in Section 8 we use the exogenous variation generated by the reform to construct an instrumental variable for the earnings effects of language proficiency. We conclude in Section 9.

2. Related literature

There exists a growing literature in economics that studies bilingualism and its impacts on socio-economic outcomes. Several recent studies in this literature report evidence of negative effects of bilingual education on school performance. Chin et al. (2013) analyse the impact of Bilingual Education Programmes for primary schools in Texas, which affected all school districts where the enrolment of Limited English Proficiency (LEP) Spanish-speaking students was above a certain threshold. Exploiting this feature of the policy in a quasi-experimental setting, they find no effect of bilingual schooling on test scores among LEP students, while some effect is found among non-LEP students. Anghel et al. (2012) evaluate the effects of a bilingual education policy on academic performance in some primary schools in the Madrid region. This is a *language-in-education* policy, i.e. not only was the language of instruction changed by the policy, but also other aspects of the educational process were affected such as training and wage incentives for teachers. The policy introduced English as medium of instruction in subjects such as Science, History and Geography. Results obtained using quasi-experimental methods point to a negative effect on exam performance in the subjects that were taught in English, concentrated among students from less-advantaged social origins. A similar negative effect has been found by Ivles and King (2014) for a *language-of-instruction* reform (i.e. limited to the medium of instruction) that took place in Russian secondary schools in Latvia. Bilingualism was introduced in these schools in the form of a 60/40 percent Latvian/Russian language of instruction mix. Results, again derived using quasi-experimental approaches, point to a short-term negative effect on exam performance among students belonging to the Russian minority.

These negative effects are apparently in contradiction with findings in cognitive psychology literature. There, the main message is that bilingual education – especially early in life – has a positive impact on cognitive development, mostly working through improved executive functions and their corresponding brain structures (Costa and Sebastián-Gallés, 2014). However, the evidence on the literacy skills of bilinguals is also mixed and some studies point towards a negative effect on literacy skills measured right after exposure to bilingual education (Bialystock, 2007). This suggests that there might be some longer-term benefits of bilingual education that are not captured when focusing on early outcomes in schools, and that longer-term outcome measures, such as labour market performance, may tell a different story. However, to the best of our knowledge, there is no evidence on the labour market effects of bilingual education, a gap we try to fill with this paper.

The labour market effects of bilingual education are still unexplored, but there exists evidence on the effect of policies that changed the language used in schools from one to another. In many cases, these changes occurred in former colonies switching from colonial to own language. This literature has an important methodological content which informs the empirical strategy of our paper. A seminal contribution is provided by Angrist and Lavy (1997), who analyse a 1983 language-of-instruction reform in Morocco. The reform switched the language of instruction at secondary schools from French to Arabic and affected all incoming students beginning secondary education in 1983, while leaving incumbent students unaffected. They develop a DiD estimator of the wage effect of the policy using years of education in the new regime as the running variable, i.e. they compare wages across educational levels for students enrolled before and after the policy took place. Their results point towards a negative effect on earnings of changing the language of instruction from French to Arabic. Using the reform as an instrumental variable, they estimate significant positive earnings effects of French proficiency.

The effect of a switch in the language of instruction from English to Spanish in Puerto Rican schools is the topic of Angrist et al. (2008), who consider English proficiency as the main outcome of interest. Differently from the Moroccan case, the reform in Puerto Rico affected all individuals in school at the time of the reform, and not only incoming students, which generates partial exposure to treatment. They apply a logic similar to Angrist and Lavy (1997) and use years of education in the new regime to derive a DiD estimator, but allow for partial exposure to the treatment exploiting variation across birth cohorts and educational levels, which is also the strategy that we adopt in the present paper. Focusing on English proficiency as the outcome of interest, they find a negative effect of the language reform, which, however, is not robust to allowance for flexible time trends by birth cohorts and educational levels.

Lleras-Muney and Shertzer (2015) estimate the effect of English-only schooling for migrants in the US within the so-called Americanization Process (1910–1930), exploiting variation in the timing of schooling legislations across states. They consider several short- and long-term outcomes, such as literacy, English proficiency schooling, employment, income and social assimilation. Their findings indicate only a modest positive effect of English-only schooling on literacy among low-background foreign-born children, but the reform seemed to be ineffective in terms of labour market and social integration outcomes.

The paper by Kuziemko (2014) is instead focused on the effect of children's English skills on parents' proficiency, considering the possibility of positive and negative spillovers within migrant families in California. The switch from bilingual education to English-only schooling for LEP students that took place in 1998 represented an exogenous shock in English proficiency among school-age children of migrants. Using territorial variation in the degree of compliance with this *language-in-education* policy, Kuziemko (2014) finds that children exposed to English-only as medium of instruction speak English better, but these increased language skills generated a negative effect on their parents' fluency in English.

While there is no study of the labour market effects of bilingual education – at least that we are aware of – there is evidence on civic outcomes, which have been investigated focusing on the same Catalan reform that we exploit in this paper for labour market outcomes. Aspachs et al. (2008) consider the Catalan reform in conjunction with a similar reform implemented in the Basque Country in the same year (1983). These reforms affected all individuals in school at the time of the reform and the authors exploit a research design that is similar to that of Angrist et al. (2008), i.e. a DiD across schooling levels and years of birth with partial exposure to treatment. They find an impact only in the case of Catalonia, because in the Basque Country two different language regimes (Basque-Spanish and Spanish-only) were allowed to coexist and parents were free to choose the language of instruction for their children, while in Catalonia the bilingual regime was compulsory in all primary and secondary schools. Clots-Figueras and Masella (2013) analyse the effect of the Catalan reform on feelings of Catalan identity and use only years of exposure at compulsory school to cope with the fact that total years of education in the new regime partly reflect possibly endogenous choices of educational attainment. In this way, they exploit only variation across cohorts before and after the reform implementation and not across educational levels, as years of compulsory education in the new regime vary only across birth cohorts and not within them. They confirm the findings of Aspach et al. (2008) that substituting Spanish-only with Spanish-Catalan education increased feelings of Catalan identity.

3. Institutional background

Catalan belongs to the family of romance languages (together with French, Italian, Occitan, Portuguese and Spanish) and has been the local language of the Spanish region of Catalonia since the early eleventh century. Starting with the War of Spanish Succession (1701–1714) and the subsequent incorporation of Catalonia within the Spanish Crown, the use of Catalan was progressively limited to domestic use and the language lost much of its social prestige. The trend reverted during the second half of the nineteenth century (the so-called '*Renaixença*'), when Barcelona became one of the cultural capitals of Europe. The dramatic political events of the 1930s represented a major negative shock for Catalan society: during the Franco dictatorial regime, Catalan was banned in the public milieu, its private use was prosecuted and Spanish became the only official language. After Franco's death in 1975 the country went through democratic transition. The decentralization process that took place with the Democratic Constitution of 1978 recognized the co-officiality of Spanish and local languages in bilingual regions, which were allowed to recover and stimulate the public and private use of their own languages. The Catalan government pursued the use of Catalan through language policies, culminating with the Language Normalization Act (LNA) of 1983. This law represented a sharp change of language policy that we detail later in this section, and enjoyed full support from all political parties and society at large. It was conceived to provide the institutional and legal basis for a complete transition towards a bilingual society, favouring the return of Catalan as the (co-)official language of the region.

The socio-demographic landscape of Catalonia after the Franco regime represented the main challenge in pursuing the aims of the reform. Due to mass migration from the Spanish-speaking areas of the country towards Catalonia since the 1950s, a substantial share of Catalan residents (i.e. migrants and their offspring) were Spanish native speakers, with limited or zero knowledge of Catalan, particularly in the periphery of the city of Barcelona, where most migrants were located. Instead, in Catalan-speaking families Catalan represented the native language even for new generations born during the dictatorship. This means that individuals of Catalan origin were fluent, at least orally, in their native language. It was against this background of linguistic segmentation that the local government used language-in-education policies as the main instrument of 'language normalization' for the new generations.

Immediately after the 1978 Constitution came into effect, Catalan language became a compulsory subject (for at least three hours per week) in non-tertiary education. With the LNA of 1983, Catalan became a medium of instruction in primary and secondary schools, alongside Spanish, making the education system effectively bilingual. The reform established that, by the end of compulsory school, all pupils must have achieved complete proficiency in the four basic competences (understanding, speaking, reading and writing) in both Catalan and Spanish. Under the new system, the two official languages were taught as subjects in a similar number of hours. Catalan had to be used as language of instruction in at least one area of study (over eight, including languages) from grade 3 to 5, and in two areas from grade 6, while Spanish had to be used as language of instruction in at least one area throughout the course of studies. Beyond these minimum mandatory requirements, the exact amount of teaching in each of the two co-official languages was determined by students' composition in terms of language background and by teachers' language skills – not all teachers were initially proficient in Catalan.

The 1983 reform introduced language immersion programmes in primary and pre-primary schools (Arnau and Vila, 2013; Artigal, 1997). These were targeted to schools whose students predominantly (more than 70 percent) came from Spanish-speaking families, which tended to be settled in areas where Catalan had very little presence. Schools in immersion programmes used Catalan as the only language of instruction during the first years of education and followed a specific methodology to stimulate second language (L2, i.e. Catalan) acquisition. Spanish was introduced as additional language of instruction only at a later stage (normally grade 3).

The LNA also regulated the language regime at post-compulsory secondary school. However, the introduction of Catalan as medium of instruction at secondary schools was less intense and more gradual compared to primary education. The reform established that secondary schools, besides mandatory language courses in Spanish and Catalan, had to employ Catalan as medium of instruction in at least two subjects. In practice, however, the choice between Catalan and Spanish was left to teachers and taken on the basis of the linguistic composition in the classroom.

The language regime of universities was not explicitly regulated by the LNA, as universities were already endowed with special autonomy when the language policy was implemented. The law only established the right to use any of the two official languages and required Catalan universities to offer Catalan courses to students and teachers with limited knowledge of Catalan.[1]

The LNA promoted bilingualism among teaching staff (Arenas, 1990, p. 28). This was especially relevant to implementing the reform, given the lack of skills to teach in Catalan among teachers, even among Catalan native speakers, particularly in the area of Barcelona and in secondary schools. Several 'specialized' teachers were hired to guarantee the minimum level of staff proficiency required by the law. These new teachers were allocated on the basis of the linguistic composition of the school in terms of both teachers and students. At the same time, less Catalan-proficient teachers were given the opportunity to develop adequate skills through special courses promoted by the SEDEC (*Servei d'Ensenyament del Català*, Catalan Teaching Service) and the Institutes of Education of the three Catalan public universities of that period. While in 1978–79 only 52 percent of teachers in Catalonia were able to speak and write in Catalan, this percentage rose to 87 percent in 1986–87 (see Arenas and Muset, 2007, p. 71).

In the first years after the LNA reform, teachers were assigned to subjects on the basis of their prior knowledge and skills in Catalan, which limited turnover. On the other hand, new positions were filled by certified 'Catalan Teachers', either through the special courses mentioned above or because they were new university graduates who developed the required skills in college. Until 1989, both new hires and incumbent personnel applying for promotions in primary and secondary schools had to either pass a specific test of Catalan knowledge or complete the special language courses. However, tests were not strictly eliminatory and those who failed had the obligation to demonstrate sufficient knowledge of Catalan in the following years. Only a subsequent norm of 1989 established the mandatory and eliminatory character of the language test for teachers.

[1] Anecdotal evidence suggests that the use of Catalan in tertiary education was already quite widespread during the early 1980s as almost half of university courses were taught in Catalan.

The existing evidence regarding the implementation of the LNA reform highlights a smooth transition towards a bilingual system. According to a school language census conducted by the SEDEC, the share of primary schools using Catalan as the main medium of instruction – alongside Spanish used in Spanish language and literature courses – rose from 42 percent in 1986 to 73 percent in 1992, while those employing both Catalan and Spanish decreased from 33 percent to 24 percent in the same years (see Vila-i-Moreno, 2000; Vila-i-Moreno and Galindo-Solé, 2009). Thence, after 10 years since enactment of the LNA, the system of compulsory education was essentially fully bilingual. The use of Catalan in these years was also substantial in post-compulsory secondary education, although less extensive: in 1991, 43 percent of secondary schools used Spanish as the main vehicular language, while this percentage fell to 30 percent and 24 percent in 1993 and 1996 respectively.

In 1994 the Spanish Constitutional Court acknowledged the constitutional validity of the so-called 'single model' for primary and pre-primary education, in which Catalan is the main language of instruction, while Spanish remains compulsory but limited to intensive Spanish language courses and a minimum number of subjects taught in Spanish. Language use at secondary schools was not affected by these changes, but the subsequent Language Policy Law (LPL) of 1998 established the single model as the mandatory language regime in all non-tertiary education, at least in public schools. The LPL also implemented several relevant changes regarding the relevance of Catalan in the labour market. First, it settled the C level of proficiency in Catalan as the prerequisite to enter public sector jobs.[2] Second, it increased the incentives to foster the use of Catalan in private business, especially among those firms who have direct commercial contacts with the Catalan public sector and/or service firms with a strong contact with the public (e.g. restaurant and hotel industry). That is, the LPL introduced the institutional basis for the creation of a bilingual labour market.

It seems worth highlighting that the progressive transition towards a school system in which Catalan is the main language did not come at the expense of proficiency in Spanish. Evidence from centralized tests by Spanish and Catalan education authorities indicates that at the end of compulsory school the level of proficiency in Catalan and Spanish are similar, and Spanish skills of Catalan students are not different from the average for the whole of Spain.[3]

[2] This corresponds to the C1 level (Effective Operational Proficiency) of the Common European Framework of Reference for Languages (CEFR) and is automatically awarded to individuals who completed lower-secondary education and had Catalan as compulsory subject from primary education (i.e. those who started primary education from the academic year 1978–79).

[3] Detailed information can be found in Instituto de Evaluación (2011) and Consell Superior d'Avaluació del Sistema Educatiu (2013).

4. Data and descriptive statistics

We use data from the Survey on Habits and Living Conditions of the Catalan Population (‘*Enquesta de Condicions de Vida i Hàbits de la Població*’, ECVHP), waves 2006 and 2011. The ECVHP provides information on labour market outcomes (net monthly earnings, hours of work and occupation), socio-demographic characteristics (province of birth of individuals and their parents, and parental education), educational attainment, language (Catalan, Spanish, both Catalan and Spanish, other languages) and self-reported Catalan proficiency (both oral and written). ECVHP data are representative of the Catalan population and the only data containing the information needed for the analysis of this paper.

Our sample consists of individuals who completed education, were born in Catalonia or migrated to Catalonia from other Spanish regions when they were aged 6 at the latest, thus excluding individuals who were, at least in part, educated outside Catalonia. We begin by selecting individuals born between 1965 and 1977 because they were aged 6 to 18 and thence potentially attending either primary or secondary school when the reform came into effect in 1983. We complement these cohorts by extending the birth year limit backwards and forwards, including in the sample individuals that were subject to the same compulsory schooling law as those born between 1965 and 1977. The first and last years of birth satisfying this requirement are 1961 and 1982 which, therefore, define the birth year limits for inclusion in the estimating sample.

[TABLE 1]

Apart from the year of birth, the main variables in our analysis are years of schooling, years of bilingual schooling and earnings. Years of schooling are imputed from the very detailed information about completed levels of education available in the database. We impute years of exposure to bilingual education on the basis of years of schooling and year of birth. The exact imputation of bilingual education is reported in Table 1. Individuals born in 1977 or after received all their schooling under the bilingual system and for them reform exposure coincides with years of schooling, while those born between 1966 and 1976 were partially exposed to the reform and individuals born before 1966 were never affected by reform. Bilingualism was not explicitly regulated at college and whether one should consider college years as years of exposure to bilingual education is inherently undetermined. To get around the indeterminacy, we consider years of college education as years of exposure only for cohorts that have not yet started college by 1983 (cohorts 1966 and onwards). Our treatment variable, therefore, would deliver a lower bound of the true effect if individuals that had already completed secondary school by 1983 (cohorts born in

1965 or earlier) actually received some bilingual education during college years. In summary, the treatment variable illustrated in Table 1 is obtained by variation in exposure to bilingual education across birth cohorts and educational levels, generating a DiD framework.

We exclude the self-employed and focus on wage earners who are regularly employed at the time of the survey.[4] As the effects of interest are identified by earnings variation across years of education and birth cohorts, our model will control for birth cohort effects. In order to avoid confounding the latter with life-cycle effects, we use repeated cross-sections (2006 and 2011) and control for quadratic experience profiles in the baseline specification, augmented by age dummies in some of the sensitivity checks (see Section 5). Clearly, the business cycle was very different between 2006 and 2011, and we will provide sensitivity checks on whether changing selection into employment between the two years affects our results.

The ECVHP collects information about net monthly earnings in both waves, but in 2006 information is reported in brackets rather than continuously. We harmonise the earnings information across waves by deflating 2011 earnings to 2006 terms and by discretizing them into the same intervals that categorize earnings in 2006. We analyse the resulting variable by means of interval regression models. Detailed descriptive information about net monthly earnings is displayed in Table A1 in the Appendix, while remaining descriptive statistics are reported in Appendix Tables A2 and A3.

In order to run falsification exercises on the validity of the underlying common trend assumption in our DiD strategy, we complement the main ECVHP sample with two auxiliary sources of information that we use to conduct placebo analyses. First, to test for the existence of contemporaneous cohort-education trends in earnings, we use data from the Spanish component of the European Survey on Income and Living Conditions, (EU-SILC, waves 2006 and 2011) and retain observations of individuals born in Spain and residing in non-bilingual regions. Second, we use older cohorts from the ECVHP, namely individuals born between 1945 and 1960 who had already completed education before the introduction of the reform and, therefore, were not affected by bilingual education. This 'placebo cohort' enables considering the presence of pre-existing cohort-education specific earnings trends in the Catalan labour market.

5. Empirical strategy

We identify the labour market returns to bilingual education by exploiting variation in exposure across cohorts and years schooling. As shown in Table 1, the LNA reform generates full, partial or

[4] We run sensitivity checks including in the analysis self-employed and entrepreneurs, finding that results were robust to this sample selection (these results are not shown and are available upon request).

null exposure to the new language regime at school depending on the year of birth and the level of education, inducing differential treatment intensity. The treatment is described by the following variable:

$$e_i = \begin{cases} s_i & if \quad b_i \geq 1977 \\ max\{0, s_i - (1977 - b_i)\} & if \quad 1965 < b_i < 1977 \\ 0 & if \quad b_i \leq 1965 \end{cases} \qquad (1)$$

where e_i are years of *potential* exposure of person *i*, s_i are years of schooling, and b_i is the year of birth. The treatment variable e_i may differ from *actual* exposure because of grade repetitions and differential intensity in the use of Catalan as medium of instruction across schools. Therefore, our treatment variable can be interpreted as capturing an Intention to Treat (ITT) effect.

5.1 Baseline model

Following Angrist et al. (2008), our baseline specification for estimating the effect of the reform on labour market outcomes is:

$$w_i = \alpha + \beta' x_i + \gamma s_i + \delta e_i + \theta_{c(i)} + \varepsilon_i \qquad (2)$$

where w_i denotes labour market outcomes, x_i is a vector of controls (wave, gender, potential experience and its square) and $\theta_{c(i)}$ is a birth cohort fixed effect, $c(i)$ being the cohort of person *i*. Our main outcome of interest is monthly net earnings from employment, but we also consider other outcomes such as employment, hours and occupation.[5] In our baseline specification birth cohorts are defined as birth-year ranges formed on the basis of reform exposure: 1961–1965 (Spanish-only schooling), 1966–1969 (exposure only after compulsory education), 1970–1976 (partial exposure also during compulsory schooling) and 1977–1982 (full exposure). The coefficient δ is the additional return per year of education generated by the reform, and parameterises a DiD estimator – across cohorts and schooling levels – that allows for partial exposure.

Equation (2) imposes linearity of the return to bilingual schooling; alternatively we allow for non-linearities through the following model which substitutes s_i and e_i with dummies for years of schooling and years of exposure. The equation that enables for differential treatment effects across schooling levels takes the form,

[5] Besides providing evidence on labour market outcomes other than earnings, looking at the effect of the reform on working hours is particularly relevant in our case since in the ECHVP 2006 monthly earnings are reported in brackets, preventing us from deriving hourly earnings measures.

$$w_i = \alpha + \beta' x_i + \sum_j [\gamma_j I(s_i = j) + \delta_j I(e_i = j)] + \theta_{c(i)} + \varepsilon_i \qquad (3)$$

where $I(\quad)$ is a dichotomous indicator variable. Equations (2) and (3) provide consistent estimates of the effects of the reform as long as any other trends in labour market outcomes (besides the ones induced by the reform) were common across birth cohorts and schooling levels.

5.2 Accounting for potential confounders

Life-cycle earnings trends may confound the effects of the reform because treated and non-treated cohorts are observed at different stages of their labour market trajectories, a form of life-cycle bias. Life-cycle effects are controlled for in our baseline model through the quadratic experience profile. We check the robustness of the baseline specification to life-cycle trends in earnings by exploiting the availability of two cross-sections. This enables adopting a more demanding specification for the cohort effects ($\theta_{c(i)}$), namely year of birth fixed; we further saturate the specification with age dummies (on top of potential experience). However, the use of two cross-sections comes with a cost, since the second one (2011) comes from a period of slack in the Spanish and Catalan economies, which has substantially increased unemployment, especially among the young. The potential change in selection into employment might undermine the consistency of our estimation; to check if this is an issue, as a sensitivity we estimate a joint model for earnings and employment and correct for selection.[6]

An additional confounding factor that we take into account is the contemporaneous expansion of education that took place in Spain over the period in which the LNA reform was implemented. Following Angrist et al. (2008), we augment the specification with a measure of the educational cumulative density function (CDF). For each survey respondent, we use data from the Spanish Census of 2001 and define the educational CDF as the fraction of individuals with a lower level of education who were born in the same birth cohort and were residing in the same province. The CDF would capture the changing selectivity into schooling levels due to education expansion, which might have reduced the average ability of individuals with higher schooling attainments.

[6] We jointly estimate an interval regression earnings equation and a linear probability for selection into employment, using as exclusion restriction the unemployment rate at the province level when the individual was 16, the legal minimum working age. We also show that the estimated effects are not very different if we estimate the earnings equation separately by year.

5.3 Placebo experiments and triple difference estimates

We pay special attention to the existence of spurious relationships between our exposure variable and the outcome due to education-cohort specific trends. The main assumption upon which validity of the identification strategy rests is the absence of differential trends in labour market outcomes across years of birth and years of schooling. We address these concerns by means of two different falsification exercises. First, we run a placebo analysis that uses data on cohorts of individuals born in the same years as those in the baseline estimating sample (1961–1982), but not residing in Catalonia or in any other bilingual region of Spain. We use the Spanish component of EU-SILC (waves 2006 and 2011) and impute reform exposure 'as if' it had been implemented also in those regions.

Second, we go back to ECHVP data and consider cohorts of Catalans born between 1945 and 1960, to whom we impute placebo exposure to bilingual schooling ($\pi^*{}_i$), pretending that the reform was rolled out in 1963 rather than in 1983, thus generating partial placebo exposure for birth cohorts 1946–1956 and full placebo exposure for individuals born after 1956:

$$\pi^*{}_i = \begin{cases} s_i & if \quad 1957 \leq b_i < 1961 \\ max\{0, s_i - (1957 - b_i)\} & if \quad 1946 \leq b_i < 1957 \\ 0 & if \quad \qquad b_i < 1946 \end{cases} \tag{4}$$

We estimate the baseline model on these cohorts using $\pi^*{}_i$ as the treatment variable. Since these cohorts were not exposed to the reform, obtaining a significant effect of the placebo variable would suggest that our treatment variable is capturing pre-existing education-cohort specific trends in earnings rather than the impact of bilingual schooling exposure.

Finally, we pool all cohorts born between 1945 and 1982 to derive a triple-difference estimator that corrects the baseline estimator for any potential pre-existing trends in earnings. We capture such spurious trends with a pseudo-exposure variable (π_i) that is equal to actual exposure for the younger cohorts ($\pi_i = e_i$ if $1961 \leq b_i \leq 1982$) and to placebo exposure for their older counterparts ($\pi_i = \pi^*{}_i$ if $1945 \leq b_i \leq 1960$). The estimating equation for the triple-difference estimator is:

$$w_i = \alpha + \beta' x_i + \eta' z_i + \gamma s_i + \lambda \pi_i + \delta^{3D} e_i + \theta_{c(i)} + \varepsilon_i \tag{5}$$

where z_i is a vector including a never-treated cohort indicator and its interactions with x_i and s_i.[7] Equation (5) provides a triple-difference estimator of the effect of the reform (δ^{3D}), which removes from the DiD any spurious earnings trend across cohorts and educational levels that is common to all cohorts. In other words, δ^{3D} is equal to the real effect of bilingual schooling exposure among the baseline cohorts (1961–1982) minus the pseudo-exposure for the never-treated cohort (1944–1960), assuming that the same trends in earnings observed among older cohorts applied to younger ones.

5.4 Selection into education

There is a final non-trivial concern with the derivation of the treatment variable, namely that it is constructed on the basis of completed schooling, which is a choice variable. This would not be an issue if the unobservables driving selection into education did not change after the implementation of the LNA reform. This assumption might not hold if, for example, before the reform some individuals with intrinsically high ability were not enrolling into post-compulsory education because school programmes were taught only in Spanish, and the introduction of bilingualism induced their younger counterparts to continue education. In such an instance, we would be erroneously attributing to the reform what in effect is a change in the unobservables.

In order to gauge the relevance of this potential issue, we first split both years of schooling and years of exposure by (observed) level of completed education, considering three levels: compulsory schooling, secondary post-compulsory schooling and tertiary schooling. We then re-estimate the equations of interest (both linear and dummies specifications) allowing for separate effects of exposure to bilingualism at school by education level. If the reform changed selection in post-compulsory education in the way described above, then we should observe the effects of the reform to be strongest at the post-compulsory level.

Secondly, we apply the strategy of Clots-Figueras and Masella (2013) and define the treatment as years of potential exposure at *compulsory* education:

$$\tilde{e}_i = \begin{cases} 8 & if \quad b_i \geq 1977 \\ max\{0,8-(1977-b_i)\} & if \quad 1970 < b_i < 1977 \\ 0 & if \quad b_i \leq 1970 \end{cases} \qquad (6)$$

As suggested by Clots-Figueras and Masella (2013), using compulsory education means that the corresponding exposure variable only depends on year of birth, which would free from any selection bias the estimated effect of exposure to bilingual schooling. Following their intuition, as

[7] We also estimate a non-linear specification of the triple difference model with dummies for years of schooling, years of actual exposure and years of pseudo-exposure.

an alternative to our baseline specification we estimate the baseline equation substituting total years of exposure with years of exposure at compulsory school.

[TABLE 2]

6. Results

6.1 Bilingual schooling and labour market outcomes

We begin the presentation of results in Table 2 reporting selected parameter estimates for the effects of the reform on various labour market outcomes.[8] Each column shows estimates from two alternative specifications. Panel A refers to the linear specification of the treatment (equation (2)), while Panel B refers to the specification with dummy variables allowing the impact to vary by degree of exposure (equation (3)). In each case estimates are net of birth cohort effects ($\theta_{c(i)}$), cohorts being formed by groups of birth years defined on the basis of reform exposure: 1961–65 (individuals who received education only in Spanish); 1966–69 (individuals who received bilingual education exposure only at secondary school); 1970–76 (individuals who received some bilingual education at primary school); 1977–82 (individuals who were entirely educated in the bilingual system). Column (1) reports estimates from the earnings model. The DiD estimates show a sizeable positive effect of bilingual education on earnings. According to the model with linear exposure, one additional year of bilingual education increases earnings by 1.1 percent. This incremental return to schooling comes on top of a yearly return of 5.8 percent (see Table A4, Column 5) and represents a proportional increase of about 20 percent. When we relax the linearity assumption in Panel B, returns vary with the amount of exposure and initial exposure matters more, suggesting that the effect is concave in exposure to the treatment. Indeed, receiving five years of bilingual schooling increases earnings by 18.1 percent relative to the benchmark case of no exposure, but having 10 more years of exposure raises the difference from the benchmark by only an additional 4 percent.

In Columns (2) to (4) of Table 2 we apply the same specifications of Column (1) to labour market outcomes other than earnings. Column (2) focuses on employment. In Column (3) we consider whether the effect observed on monthly earnings could stem from differences in hours of work between treated and non-treated individuals and use weekly hours as the relevant outcome. Such exercise is important because we are unable to construct an hourly wage measure consistently in both waves. Finally, we consider whether the earnings effect could stem from better access to

[8] In this and the following tables we report only key parameter estimates; the full set of estimated coefficients of the earnings regression are reported in Table A4 in the Appendix for models with linear exposure. In the same table, we also report separate estimates by wave, as well as estimates on hourly wages (only available in 2011), with and without the exposure variable.

more prestigious occupations and estimate a linear probability model for being in highly skilled white-collar jobs (Column (4)). For none of these additional labour market outcomes do we find a significant effect of the reform (the exposure dummies are jointly significant on hours only at the margins of the 10 percent confidence level, but none of them is individually significant). These results indicate that the earnings effects of bilingual education operate through earning capacity rather than through working time or better employment prospects. This is a first piece of evidence suggesting that the reform operated by affecting the process of accumulation of productive skills, rather than by favouring more or better matches in the labour market (say via increased search efficiency or reduced employer discrimination). In the light of this evidence, we focus on earnings returns in the remainder of this section, and we will return in the next section to the analysis of possible mechanisms behind the relationship linking bilingual schooling and earnings.

6.2 Sensitivity checks

Table 3 reports the results of sensitivity checks on the earnings effects. In Column (1) we augment the baseline model to correct for selection into employment, using unemployment rate in the province of birth at the time the individual was 16 plus all the regressors of the earnings equation to model individual employment probabilities. This is relevant since our data come from two pooled cross-sections from different phases of the business cycle. In Columns (2) and (3) we account as much as possible for life-cycle differences in earnings. First, in Column (2) we augment the baseline specification (that already includes a quadratic trend in experience) with age dummies. Second, we saturate the model by also substituting the birth cohort dummies (that group several birth years depending upon reform exposure) with birth year dummies (in Column (3)). Finally, in Column (4), to shed additional light on the threat to identification that may come from the contemporaneous educational expansions, we include the educational CDF (and its square) as additional control in the earnings equation (as in Angrist et al., 2008).

Accounting for selection into employment barely affects the estimate of the treatment effect. As shown in Column (1) of Table 3, the estimates from both specifications with linear and non-linear exposure effects are only slightly higher than in the baseline specification (Column (1) of Table 2). The overall picture remains the same when we include age dummies in the model, which generates a slight increase in the treatment effect estimates (especially the linear specification). The effect goes back to its baseline size when we saturate the model with birth year dummies, and marginally loses significance. When controlling for the educational CDF in Column (4) the earnings effect remains sizeable and significant, suggesting that our treatment variable is not confounded by the expansion of education that took place in Spain during the same years the LNA reform was implemented.[9] Overall, these sensitivity checks point to a substantial robustness of the estimated earnings effect of exposure to bilingual schooling.[10]

[TABLE 3]

6.3 Falsification analysis

In this section we run several falsification experiments that are aimed at ruling out the existence of spurious relationships between the treatment variable and earnings. The main assumption underlying the consistency of our DiD estimator is that any (education/cohort specific) earnings trend unrelated to the reform is common between treated and control cohorts. To assess the validity of this non-trivial assumption, in Table 4 we report results of placebo regressions and contrast the resulting estimates with the baseline results shown in Column (1).

[TABLE 4]

First, we check for the existence of contemporaneous spurious relationships by using EU-SILC data and imputing placebo variable to individuals born in the same cohorts as our main sample (1961–1982) but in other non-bilingual Spanish regions, who, therefore, were not exposed to the reform. The results are displayed in Column (2). The estimates of placebo exposure among contemporaneous cohorts of individuals from non-bilingual Spanish region are not significant and their point estimates are substantially lower than the coefficient for true exposure. This evidence suggests that our exposure variable is not capturing contemporaneous trends in earnings that apply to other Spanish regions.

Second, we consider the presence of a pre-existing trend within the Catalan labour market, by assigning placebo exposure (π^{*}_{i}) to an older cohort (1945–1960) from our main sample (EHCVP) 'as if' the LNA reform had been applied in 1963 instead of 1983. The results obtained from this cohort of never-treated Catalan individuals are displayed in Column (3) of Table 4. Neither the coefficient from the linear specification nor any of the coefficients on the dummies for years of exposure is statistically significant at conventional levels of confidence. The estimates on the dummies are jointly statistically significant but they tend to float around quite substantially without

[9] The estimated coefficients on the educational CDF and its square (available upon request) are very precisely estimated and indicate that the relationship between the measure of educational expansion and earnings is convex.

[10] In Table A4, columns (7) and (8), we report estimates of the linear specification separately by wave. By construction, these estimates cannot fully control for life-cycle effects, but anyway point towards the robustness of the results. In 2006, the baseline return to education was 5.4 percent and the treatment increased it by 0.8 percent, while in 2011 the baseline return was 6.5 percent and the effect of the reform was 1.8 percent. In both waves, the effect of the reform is less precisely estimated than in the baseline model due to smaller sample sizes.

revealing any clear pattern. In any event, both the coefficient from the linear specification and the coefficients from the specifications with exposure dummies are negative, suggesting that our baseline estimates would be, at worst, a lower bound for the true effect. Taken together, the evidence from Column (3) rules out any threat to identification coming from pre-dating earnings trends across cohorts and schooling levels.

Third, we combine the baseline (1961–1982) and the never-treated (1945–1960) cohorts to build a triple-difference estimator, which removes any pre-existing trend from the treatment effect estimator. The results from the triple-difference equation (5) are reported in Column (4), which displays the estimates of real exposure (δ^{3D}) and pseudo-exposure (λ), where the latter removes any pre-existing trend from the former. Consistent with the results obtained for the never-treated cohort only, the earnings effect of (real) exposure to bilingual education is somewhat higher when estimated applying a triple difference. This again suggests that if there exists any pre-dating trend unrelated to the reform, this seems to be weak and to generate a modest downward bias in the baseline estimates of the returns to bilingual schooling.

6.4 Endogenous selection into education

Another fundamental concern with our DiD estimator is selection into schooling, which is an issue if the estimated earnings effects reflect changing unobserved ability of more educated individuals before and after the reform, rather than the true effect of bilingual schooling. We begin dispelling these doubts in Table 5, which reports the effects of reform exposure by levels of education: if the baseline effect is an artefact of selection into education, we should observe it to be the strongest after compulsory education. In Columns (1) to (3) we present returns to years of bilingual education separately by segment of the educational system in which they are received (compulsory, post-compulsory, tertiary). The evidence definitely points to the fact that the predominant effect stems from compulsory schooling. The linear specification indicates that one year of compulsory bilingual education increases earnings by 1.7 percent when we account only for exposure at compulsory school (Column 1) and by 1.8 percent when we simultaneously control for exposure at all schooling levels. The earnings effect of exposure during compulsory schooling is clearly non-linear, as shown in the lower panel of Table 5. In fact, the estimates show patterns similar to those obtained for whole exposure, highlighting that the earnings return is increasing with the amount of compulsory schooling received in both languages, but initial years of exposure during compulsory schooling matter more than late exposure. On the contrary, the effects of exposure at higher educational levels are smaller in size and imprecisely estimated.

[TABLE 5]

This goes against the hypothesis that the earnings effect is reflecting changes in endogenous selection into schooling, as there is no element of choice in years of compulsory schooling. This is the argument used by Clots-Figueras and Masella (2013), who estimated the effects of the 1983 reform on feelings of Catalan identity using the *potential* number of years of compulsory education (the variable $\tilde{e}_i$ in Section 5) to capture the exogenous amount of exposure to the reform. Their measure differs from the one used in Table 5 due to compulsory school drop outs, which are assigned eight years of compulsory schooling in the potential measure of Clots-Figueras and Masella (2013) and actual years of compulsory schooling in Columns (1) of Table 5.

[TABLE 6]

Using the same measure as Clots-Figueras and Masella (2013) delivers an effect of compulsory exposure that is essentially the same as the one we obtain with actual exposure at compulsory school, as shown in Table 6. These results confirm that changing selection into schooling does not affect our results, mostly because what really matters is exposure to bilingualism during compulsory education (which is not driven by individual choices and related unobservable characteristics). Nevertheless, there is a final important issue that deserves additional attention. By only depending on year of birth, the potential exposure variable rules out selection on unobservables, but, similarly to the full exposure variable, could be biased by differential earnings trends across cohorts. The issue is in fact more serious for compulsory exposure, as the corresponding estimator is simply based on a before-after comparison, not a DiD one, since there is no difference in exposure between members of the same birth cohort. We address the relevance of these perils with a falsification analysis based on the contemporaneous cohorts of individuals from non-bilingual Spanish regions, which we assume to receive pseudo years of compulsory exposure.[11] The results are reported in Column (3) of Table 6. The estimates obtained from the linear specification are low and not statistically significant. A similar picture emerges from the non-linear specification of the placebo, estimates being generally close to zero and not significant.

Combining all the evidence that we presented so far enables us to draw some initial conclusions. Individuals exposed to bilingualism at compulsory school obtain a sizeable earnings premium,

[11] Notice that we are unable to use the never-treated cohort of Catalans to falsify the effect of exposure during compulsory schooling, since individuals belonging to the 1945–1960 cohort were not subject to the same compulsory schooling rules. Specifically they were not affected by the "*Ley General de Educación* (LGE)", approved in 1970 and enforced in 1974, which increased the duration of compulsory education to eight years.

which is concave in the amount of exposure. The positive return to bilingual education that we have found is not contaminated by confounders, spurious relationship of unobserved heterogeneity affecting schooling progression. The last claim is motivated by the fact that exposure during compulsory education – that is plausibly exogenous – represents the main driver of the results. We now investigate possible channels of these effects.

7. Channels and heterogeneous effects

Results from the previous section show that the effects of the reform are limited to earnings, while no impact can be detected on employment, hours of work or occupation. That evidence supports a human capital interpretation of our results, as alternative mechanisms such as search or discrimination would have produced effects also on outcomes other than earnings. Focusing on the potential exposure at compulsory education to rule out endogeneity of schooling decisions, we now provide further evidence about possible channels of those effects.

[TABLE 7]

7.1 Possible channels

We consider several possible 'first-stages' of hypothetical 'structural models' that are compatible with the earnings effects shown so far. We focus on three candidate-mediating factors: sector of occupation (public or private), the language of the respondent, and Catalan proficiency (being able to speak and write in Catalan). In Spain, public sector workers enjoy a wage premium (Ramos et al., 2014) and, in the specific case of Catalonia, the 1998 Language Policy Law established level C of Catalan proficiency as a prerequisite for accessing public sector jobs. Individuals educated in the bilingual system (specifically birth cohorts 1972 and onwards), were automatically certified in level C, which makes their access to better-paid public sector jobs easier compared with non-treated cohorts. This would explain the positive earnings effects consistently with a search and matching mechanism, because the reform would widen the set of job offers.

Second, we consider respondents' language as a possible mediator of the earnings effect of the reform, as it may serve as an indicator for the access to Catalan social networks.[12] The results by Aspachs et al. (2008) and Clots-Figueras and Masella (2013) indicate that language exposure at school fostered the feeling of Catalan identity, which could in turn improve the chances for non-Catalans of having contact with Catalans and of joining their networks. We do not observe proxies

[12] Respondents are asked which alternative better describes their language among Catalan only, Spanish only, Catalan and Spanish, other languages. Only three cases indicated other languages and are excluded throughout the analysis.

for social networks or identity feelings; therefore, we focus on respondents' language, given the strong interrelations between language and identity (Ginsburgh and Weber, 2011). If treated individuals are more likely to see themselves as Catalan, we should find among them a reduction in Spanish usage, which in turn could foster access to better job opportunities provided by Catalan social networks.

Third, language proficiency represents a natural channel for the effectiveness of bilingual education. Before the reform proficiency in Spanish was homogeneously high for both treated and control cohorts irrespective of their origins, while Catalan proficiency was more heterogeneous. Individuals of Catalan origin could speak Catalan but still had issues with writing, and individuals of Spanish origin – mainly second-generation immigrants – were also lacking spoken proficiency. Therefore, we consider whether individuals exposed to the reform are more likely to be proficient in Catalan and, later, how this interacts with their regional origins.

Results are reported in Panel A of Table 7, together with the linear estimate of the earnings returns to exposure at compulsory schooling for comparison. The coefficients of the treatment variable are close to zero and insignificant when the dependent variable is either having a public sector job or being Spanish usual speaker. This evidence shows that the reform does not affect employment sector or language choices, which does not support a search-and-matching interpretation of the earnings effects. On the contrary, compulsory exposure has a positive effect on the language proficiency indicator.[13] This time the effect of the reform is small but statistically significant, increasing the probability of reporting themselves as proficient by 1.6 percent, which provides further support for a human capital interpretation of our results.

7.2 Heterogeneous treatment effects

In Panel B of Table 7 we allow for heterogeneous effects by regional origins, which are captured by an indicator for having both parents born outside Catalonia. This is interesting not only because second-generation immigrants were the main target of the reform, but also because it could shed light on the relevance of labour market discrimination as a channel for the earnings effect. If workers with non-Catalan origins are discriminated against in the Catalan labour market, the observed earnings effect might reflect a reduction in the 'social distance' between the two groups brought about by the reform, and a consequent reduction in discrimination, rather than a human capital effect as we have argued. Prospective employers cannot always observe parental origins, but

[13] Self-reported language proficiency is prone to be over-reported, which may give rise to non-classical measurement error issues, see Dustmann and van Soest (2004). We experimented using as an alternative 'objective' proficiency indicator the language of the interview (about 30 percent of the respondents gave their interview in Spanish and this information is directly recorded by the interviewer) and found results in line with those discussed in the text.

in Catalonia there is a close mapping between origins and surnames, the latter being perfectly observable by employers (Güell et al., 2015). According to the discrimination interpretation, we would then expect a positive and significant interaction between compulsory language exposure and the indicator for non-Catalan origins, while the main effect should evaporate. The results in Panel B do not support the interpretation based on discrimination. We only find mild evidence compatible with labour market discrimination against non-Catalans, the associated coefficient in the earnings equation being negative (-2.5 percent) but imprecisely estimated (s.e. = 0.19). Most importantly, there is only a very small differential effect of the reform for non-Catalans (0.5 percent, with s.e. of 0.003) while the main effect of exposure remains sizeable and significant. Compulsory language exposure has again no effect on either public sector employment or language used, regardless of regional origins. Note also that the effect of the reform on proficiency in Catalan is concentrated among the group with non-Catalan origins, who were the main target of the reform.

[TABLE 8]

As discussed in Section 3, with the migrations of the 1950s and the 1960s many low-skilled workers arrived in Catalonia from the rest of Spain. Finding, as we do, that non-Catalan origins are associated – though only mildly – with the effects of the reform may be consistent with the idea that the reform is beneficial for individuals from less-advantaged parental backgrounds. If the Catalan language-in-education reform increased the quality of education (say because it promoted teachers' turnover) or favoured the development of cognitive skills, we expect low-background individuals to benefit the most as school quality could compensate for low parental education, irrespective of regional origins. To shed further light on the heterogeneous effects of the reform and the possible channels of its effectiveness, in Table 8 we interact treatment exposure with an indicator for having parents with at most compulsory education, splitting the sample according to regional origins. Indeed, for individuals with Catalan roots, all the effect is coming from low-background individuals, which is consistent with the interpretation based on educational quality. On the other hand, among individuals with non-Catalan origins, there is no differential earnings effect of the reform for those with low parental background, suggesting that for them what matters is the effect coming through proficiency irrespective of background.

In summary, the analysis of possible channels and heterogeneous effects suggests that mechanisms like search-and-matching or labour market discrimination are not consistent with the evidence, which instead supports human capital interpretations. Specifically, for individuals with non-Catalan origins (the main target of the reform) bilingual schooling resulted in increased proficiency in Catalan. For the general population, the reform increased the quality of education, which was beneficial for individuals from disadvantaged parental backgrounds.

8. Compulsory exposure, language proficiency and earnings

We have just shown that proficiency in Catalan appears to be one relevant channel through which exposure to bilingualism at school fostered earnings potential among the target population. We now exploit this finding to derive an instrumental variables (IV) estimator for the earnings effects of Catalan proficiency, using an identification strategy similar to that of Bleakley and Chin (2004). They estimate the return to English proficiency for US migrants using the interaction between age at arrival (which proxies for differential timing of language exposure for children of different ages) and country of origin (whether English-speaking or not) as an instrument for proficiency.[14] While age at arrival itself may have direct wage effects (say via adaptation to institutions in the host country), its interaction with the country of origin provides an additional (and exogenous) source of variation in proficiency, since language proficiency among individuals from English-speaking countries is the least affected by age at arrival. Indeed, assuming that the direct wage effects of age of arrival are the same irrespective of country of origin, their instrument isolates the proficiency effects of English exposure from the direct wage effects of age at arrival.

We recover a similar setup using the interaction between years of compulsory exposure ($\tilde{e}_i$) and the indicator for respondents' language introduced in the last section, which we use as a proxy for *native* language. This interaction affects proficiency as long as exposure is the most beneficial for Spanish speakers. Instrument validity relies upon the assumption that any direct earnings effects of the reform are common between Spanish and Catalan speakers. Using this instrument, we obtain a Local Average Treatment Effect (LATE) estimator for the earnings effects of proficiency, namely the effect of the treatment on the sub-population of Spanish speakers. Let ES_i be an indicator for whether individual *i* reports Spanish being her language. The first stage regression of the IV-LATE estimator is:

$$p_i = \alpha_p + \beta_p{}' x_i + \gamma_p s_i + \delta_p \tilde{e}_i + \varphi_p ES_i + \phi \tilde{e}_i \times ES_i + \lambda_{c(i)} + v_i \qquad (7)$$

where p_i is a dummy variable for proficiency in Catalan. The 'structural' equation that estimates the earnings effect of proficiency is the following:

[14] Bleakley and Chin (2010) used the same strategy to identify the causal effect of English proficiency on social assimilation of migrants, considering as outcomes marital status and partner's characteristics, fertility and residential decisions.

$$w_i = \alpha + \beta' x_i + \gamma s_i + \delta \tilde{e}_i + \varphi ES_i + \psi p_i + \theta_{c(i)} + \varepsilon_i \quad (8)$$

[TABLE 9]

Results from the estimation of the IV-LATE model are in Table 9. The first column simply adds Catalan proficiency to the baseline model of equation (2), finding a positive effect of 5.8 percent and statistically significant. This estimate may reflect both the causal earnings effect of proficiency and selection bias stemming from unobserved ability, but may also suffer from measurement/misclassification errors. While unobserved ability bias would imply that the estimate is an upward bias of the effect of interest, the presence of errors in the self-reported proficiency variable is likely to generate a bias in the opposite direction (Dustmann and van Soest, 2004).

In column (2) we further augment equation (2) with the Spanish language indicator to account for any possible extra earnings effect of Spanish-speaking, besides the one operating through proficiency. Discriminatory employer behaviour or search efficiency may represent an example of those effects and, similarly to the evidence on the effects of non-Catalan origins in Table 7, there is an earnings penalty for Spanish speakers (-1.6 percent), albeit imprecisely estimated (s.e. = 0.013). The estimates of the reduced form model that excludes proficiency from the earnings equation but includes the instrument are in column (3), showing that while Spanish speakers have a sizeable and significant earnings penalization (-5.8 percent), that disadvantage is reduced by bilingual compulsory schooling (0.8 percent for each additional year of exposure, with s.e. = 0.003).

The last column reports the estimates of the IV-LATE model. Because earnings are in brackets, we estimate by maximum likelihood a two-equations system, one for earnings (interval regression model) and one for proficiency (linear probability model). The proficiency equation (on the right) shows that the instrument captures the proficiency effect of the reform and does so in a powerful enough way. Being a Spanish-only speaker is associated with a substantially lower probability of reporting oneself as proficient, but the gap with Catalan speakers is significantly reduced by exposure to bilingualism, by 3.7 percentage points for each additional year of exposure. The IV-LATE estimator of the effect is in the left part of column (4), showing a sizeable and significant proficiency premium of 22 percent. This effect is rather large compared with the OLS, and is in the range of the IV estimates of Bleakley and Chin (2004), who report a 33 percent premium associated with a unit increase of self-assessed proficiency on a four-point scale. Also similar to Bleakley and Chin (2004) is the increase in size compared with the model with exogenous proficiency, though in their case the difference is less pronounced (they estimate approximately a 50 percent increase between OLS and IV, while in our case the increase is more than fourfold). A likely reason for this discrepancy is that while the Bleakley and Chin (2004) measure is on a four-point scale, our measure is binary and thus the coefficient parameterises the earnings effect of a larger shift in language fluency. Bleakley and Chin (2004) exclude that their evidence is compatible with the IV estimates reflecting heterogeneous effects between groups whose proficiency is differentially affected by the instrument, because simple OLS earnings regressions by subgroups do not reveal differential returns to proficiency. Following their strategy we also find no evidence of differential returns to Catalan proficiency according to whether or not respondents report Spanish as their language. Thence, similarly to them, we suggest that presence of errors in self-reported proficiency is the reason for the size increase characterizing IV estimates.

In Panel B of Table 9 we propose an alternative identification strategy that uses Catalan origins in place of respondents' language to form the exclusion restriction. The instrument in this case is less powerful, and results in an imprecise estimate of the effect of proficiency on earnings, but the point estimate is similar to the one obtained using our preferred exclusion restriction. As both instruments operate in the same direction, using them jointly in Panel C of Table 9 offers the opportunity to run an over-identification test, whose estimated p-value is 0.76. This is a piece of evidence that supports the identification assumption that any direct earnings effects of the reform do not depend on whether Spanish is reported as the respondent's language, or on Catalan origins. We also experimented by including the interaction between compulsory language exposure and the non-Catalan origin indicator (and the base effect of the latter variable) in the earnings equation, without finding any significant effect on the included interaction or any alteration in the instrumented proficiency effect. This result is important, since adding this interaction as control would implicitly relax the underlying hypothesis that any remaining effect of the reform is homogeneous within the population of interest, and is thus reassuring about the reliability of our IV-LATE estimates.

9. Concluding remarks

The 1983 introduction of Catalan alongside Spanish as a medium of instruction in the Catalan schooling system provides a unique opportunity for evaluating the effects of bilingual schooling. In this paper we exploit the differential exposure to reform across birth cohorts and levels of education to provide the first evaluation of the earnings effects of bilingual education, complementing strands of literature that have focused either on the effects of bilingualism on educational outcomes, or the labour market effects of changing one language of instruction with another. We find positive earnings effects: one year of bilingual education raising earnings by on average 1 percent, representing one-fifth of the baseline return to education. These effects are robust to alternative specifications, and placebo analyses exclude that they could reflect spurious trends. Conversely, we do not find effects on employment, hours of work or occupation, nor do we find the earnings effect to stem exclusively from non-Catalans, all of which leads us to favour human capital interpretations of the findings rather than explanations based on search-and-matching mechanisms or discrimination.

Coming after four decades of a totalitarian regime that banned Catalan from the public milieu, one of the main motivations of the reform was to level the playing field in Catalonia, promoting an effective integration of non-Catalans, many of whom were first- or second-generation immigrants from poorer Spanish regions. Our results indeed show that the non-Catalan labour force enjoyed the highest earnings benefits due to its increased language proficiency. In this respect, the reform was arguably successful in reducing segregation and favouring the development of a unitary society. However, the positive effects spread to all the cohorts exposed to bilingual schooling irrespective of their origins, particularly individuals from unfavourable family backgrounds. This is consistent with an increased quality of the overall educational process brought about by the reform and with direct positive effects of bilingualism on skill formation.

References

Adesope O. O., Lavin T., Thompson, T. and Ungerleider C. (2010). "A systematic review and meta-analysis of the cognitive correlates of bilingualism," Review of Educational Research 80 (2): 207–245.

Akerlof, G.A. and Kranton, R.E. (2010). “Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being”. Princeton University Press, Princeton: NJ.

Anghel, B., Cabrales, A. and Carro, J. (2012). “Evaluating a bilingual education program in Spain: the impact beyond foreign language learning,” CEPR Discussion Papers (No. 8995).

Angrist, J. D. and Lavy, V., (1997). "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco," Journal of Labor Economics, 15(1): 48-76.

Angrist, J. D., Chin, A. and Godoy, R., (2008). "Is Spanish-Only Schooling Responsible for the Puerto Rican Language Gap?" Journal of Development Economics, Elsevier, 85(1-2): 105-128.

Arenas, J. (1990). “*Llengua i Educació a la Catalunya d’Avui* (Language and Education in Today’s Catalonia)”. La llar del llibre: Barcelona.

Arenas, J., Muset, M. (2007). “*La immersió lingüística. Una obra de govern, un projecte compartit* (Language immersion: a governance work, a shared project)”. Centre d'Estudis Jordi Pujol. Barcelona.

Arnau, J., and Vila, F. X. (2013). “Language-in-education policies in the Catalan language Area.” In “Reviving Catalan at school. Challenges and instructional approaches”, by Arnau, J. (Ed.), pages 1-28. Multilingual Matters, Bristol, UK.

Artigal, J.M. (1997). “The Catalan Immersion Program. In Johnson, R.K. and Swain, M. (Eds.), Immersion Education: International Perspectives.” Cambridge University Press: Cambridge.

Aspachs-Bracons, O., Clots-Figueras, I., Costa-Font, J. and Masella, P. (2008). "Compulsory Language Educational Policies and Identity Formation," Journal of the European Economic Association, 6(2-3): 434-444.

Bialystok, E. (2007). “Acquisition of literacy in bilingual children: A framework for research,” Language learning, 57(s1): 45-77.

Bleakley, H. and Chin, A. (2004). "Language Skills and Earnings: Evidence from Childhood Immigrants," The Review of Economics and Statistics, 86(2): 481-496.

Bleakley, H. and Chin, A. (2008). “What Holds Back the Second Generation? The Intergenerational Transmission of Language Human Capital among Immigrants,” Journal of Human Resources, 43(2): 267-298

Bleakley, H. and Chin, A. (2010). "Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants," American Economic Journal: Applied Economics, 2(1): 165-92.

Chin, A., Daysal, N. M. and Imberman, S. A. (2013). "Impact of bilingual education programs on limited English proficient students and their peers: Regression discontinuity evidence from Texas," Journal of Public Economics, 107: 63-78.

Clots-Figueras, I. and Masella, P. (2013). "Education, Language and Identity," The Economic Journal, 123(570): 332-357.

Consell Superior d'Avaluació del Sistema Educatiu (2013). "*Sistema d'indicadors d'ensenyament de Catalunya*", No. 17, Generalitat de Catalunya.

Costa, A. and Sebastián-Gallés, N. (2014). "How does the bilingual experience sculpt the brain?" Nature Reviews Neuroscience, 15(5): 336-345.

Dustmann, C. and Van Soest, A. (2004). "An analysis of speaking fluency of immigrants using ordered response models with classification errors," Journal of Business and Economic Statistics, 22(3), 312-321.

Güell, M., Mora, J. V. R. and Telmer, C. I. (2015). "The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating," The Review of Economic Studies, 82(2), 693-735.

Instituto de Evaluación (2011). "*Evaluación General de Diagnóstico 2010: Educación Secundaria Obligatoria, Segundo Curso*", Informe de Resultados, Ministerio de Educación.

Ivlevs, A. and King, R. M. (2014). "2004 Minority Education Reform and pupil performance in Latvia," Economics of Education Review, 38, 151-166.

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

Lleras-Muney, A. and Shertzer, A. (2015). "Did the Americanization Movement Succeed? An Evaluation of the Effect of English-Only and Compulsory Schools Laws on Immigrants," American Economic Journal: Economic Policy, 7(3): 258-90.

Ramos, R., Sanromá, E. and Simón, H. (2014). "Public-Private Sector Wage Differentials by Type of Contract: Evidence from Spain", Hacienda Pública Española/Review of Public Economics, 208 (1): 107-141.

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

Vila-i-Moreno, X. (2000). "*Les polítiques lingüístiques als sistemes educatius dels territoris de llengua catalana* (Language policies in the education systems of Catalan language territories)," Revista de Llengua i Dret, 34: 169-208.

Vila-i-Moreno, X. Galindo-Solé, M. (2009). “*El sistema de conjunció en català en l’educació primària a catalunya: impacte sobre els usos* (the Catalan conjunction system of primary education in Catalonia; impact on usage),” Treballs de Sociolingüística Catalana, 20: 21-69.

Table 1: Imputed years of exposure

year of birth	years of schooling							
	3	6	8	11	12	14	15	17-20
1961-65	0	0	0	0	0	0	0	0
1966	0	0	0	0	1	3	4	6
1967	0	0	0	1	2	4	5	7
1968	0	0	0	2	3	5	6	8
1969	0	0	0	3	4	6	7	9
1970	0	0	1	4	5	7	8	10
1971	0	0	2	5	6	8	9	11
1972	0	1	3	6	7	9	10	12
1973	0	2	4	7	8	10	11	13
1974	0	3	5	8	9	11	12	14
1975	1	4	6	9	10	12	13	15
1976	2	5	7	10	11	13	14	16
1977-82	3	6	8	11	12	14	15	17

Table 2: Effects of the reform on labour market outcomes

	(1)	(2)	(3)	(4)
	Earnings	Employment	Hours	White collar High skilled
Panel A: Linear specification				
years of exposure	0.011**	-0.001	0.068	-0.001
	(0.006)	*(0.005)*	*(0.132)*	*(0.007)*
Panel B: Dummy variables specification (reference = 0 years of exposure)				
years of exposure = 1	0.030	-0.015	-1.557	-0.025
	(0.053)	*(0.043)*	*(1.196)*	*(0.045)*
years of exposure = 2	0.084*	-0.032	1.315	0.003
	(0.051)	*(0.058)*	*(1.260)*	*(0.041)*
years of exposure = 3	0.109**	-0.019	-0.066	-0.010
	(0.047)	*(0.048)*	*(1.481)*	*(0.052)*
years of exposure = 4	0.131***	-0.034	0.331	-0.015
	(0.048)	*(0.045)*	*(1.160)*	*(0.038)*
years of exposure = 5	0.181***	-0.040	-0.395	0.075
	(0.056)	*(0.048)*	*(1.222)*	*(0.052)*
years of exposure = 6	0.206***	-0.066	0.893	0.028
	(0.046)	*(0.045)*	*(1.317)*	*(0.040)*
years of exposure = 7	0.119***	-0.058	0.653	-0.036
	(0.044)	*(0.047)*	*(1.240)*	*(0.045)*
years of exposure = 8	0.185***	-0.026	1.156	0.009
	(0.049)	*(0.054)*	*(1.517)*	*(0.044)*
years of exposure = 9	0.184***	-0.057	1.399	-0.037
	(0.060)	*(0.059)*	*(1.555)*	*(0.056)*
years of exposure = 10	0.191***	-0.046	1.322	-0.008
	(0.058)	*(0.063)*	*(1.586)*	*(0.057)*
years of exposure = 11	0.155**	-0.009	1.644	-0.025
	(0.062)	*(0.062)*	*(1.630)*	*(0.058)*
years of exposure = 12	0.209***	-0.028	0.157	-0.004
	(0.066)	*(0.068)*	*(1.794)*	*(0.062)*
years of exposure = 13	0.237***	-0.038	2.314	-0.017
	(0.071)	*(0.069)*	*(1.841)*	*(0.078)*
years of exposure = 14	0.236***	-0.047	1.647	0.038
	(0.074)	*(0.074)*	*(2.016)*	*(0.072)*
years of exposure = 15	0.217***	0.001	0.093	-0.043
	(0.079)	*(0.078)*	*(2.108)*	*(0.077)*
years of exposure = 16	0.295***	-0.004	0.263	-0.025
	(0.073)	*(0.080)*	*(2.049)*	*(0.090)*
years of exposure = 17	0.176*	-0.056	1.362	-0.087
	(0.090)	*(0.093)*	*(2.405)*	*(0.099)*
F-test of joint significance (p-value)	0.000	0.416	0.099	0.524
Number of observations	3,323	4,189	3,290	3,323

****, **, * denote significance at the 1, 5 and 10 percent level. Robust standard errors in parentheses are clustered by year of birth, years of schooling and wave. Additional controls shown in Table A4: wave, gender, years of schooling, potential experience and its square, birth-cohort dummies. Column 1 uses interval regression, columns 2 and 4 use linear probability model, column 3 uses OLS.*

Table 3: Earnings effects of the reform: sensitivity checks (N=3,323)

	(1)	(2)	(3)	(4)
Panel A: Linear specification				
years of exposure	0.012**	0.015**	0.011	0.015***
	(0.005)	*(0.007)*	*(0.008)*	*(0.005)*
Panel B: Dummy variables specification (reference = 0 years of exposure)				
years of exposure = 1	0.021	0.034	0.027	0.033
	(0.050)	*(0.051)*	*(0.049)*	*(0.050)*
years of exposure = 2	0.080	0.087*	0.086*	0.087*
	(0.056)	*(0.049)*	*(0.047)*	*(0.052)*
years of exposure = 3	0.110***	0.112**	0.083	0.112**
	(0.043)	*(0.049)*	*(0.053)*	*(0.044)*
years of exposure = 4	0.127***	0.137***	0.135***	0.131***
	(0.043)	*(0.045)*	*(0.048)*	*(0.045)*
years of exposure = 5	0.175***	0.210***	0.192***	0.184***
	(0.055)	*(0.052)*	*(0.058)*	*(0.056)*
years of exposure = 6	0.207***	0.226***	0.209***	0.206***
	(0.050)	*(0.047)*	*(0.055)*	*(0.046)*
years of exposure = 7	0.121**	0.149***	0.149***	0.121***
	(0.049)	*(0.045)*	*(0.056)*	*(0.047)*
years of exposure = 8	0.183***	0.223***	0.205***	0.185***
	(0.050)	*(0.051)*	*(0.060)*	*(0.051)*
years of exposure = 9	0.183***	0.247***	0.220***	0.186***
	(0.061)	*(0.063)*	*(0.072)*	*(0.058)*
years of exposure = 10	0.194***	0.262***	0.247***	0.193***
	(0.059)	*(0.064)*	*(0.078)*	*(0.058)*
years of exposure = 11	0.151**	0.233***	0.217**	0.158**
	(0.060)	*(0.072)*	*(0.086)*	*(0.063)*
years of exposure = 12	0.207***	0.315***	0.289***	0.215***
	(0.067)	*(0.078)*	*(0.095)*	*(0.069)*
years of exposure = 13	0.235***	0.334***	0.323***	0.243***
	(0.069)	*(0.083)*	*(0.104)*	*(0.070)*
years of exposure = 14	0.240***	0.373***	0.344***	0.239***
	(0.076)	*(0.092)*	*(0.116)*	*(0.077)*
years of exposure = 15	0.210***	0.384***	0.353***	0.236***
	(0.079)	*(0.106)*	*(0.130)*	*(0.081)*
years of exposure = 16	0.295***	0.426***	0.419***	0.304***
	(0.093)	*(0.099)*	*(0.132)*	*(0.094)*
years of exposure = 17	0.190**	0.387***	0.357**	0.198**
	(0.091)	*(0.122)*	*(0.156)*	*(0.092)*
F-test of joint significance (p-value)	0.017	0.000	0.003	0.014
employment selection correction	*yes*	*no*	*no*	*no*
cohort dummies	*yes*	*yes*	*no*	*yes*
age dummies	*no*	*yes*	*yes*	*no*
year of birth dummies	*no*	*no*	*yes*	*no*
educational CDF	*no*	*no*	*no*	*yes*

****, **, * denote significance at the 1, 5 and 10 percent level. Robust standard errors in parentheses are clustered by year of birth, years of schooling and wave. Additional controls not shown: wave, gender, years of schooling, potential experience and its square, birth-cohort dummies. Column (1) is the earnings equation of a model with endogenous selection into employment that uses the provincial unemployment rate when the individual was 16 as the exclusion restriction, and clusters standard errors also at the province level. Column (4) clusters standard errors also at the province level because of repeated observations in the educational CDF.*

Table 4: Falsification exercises for the assumption of parallel trends in monthly earnings

	(1)	(2)	(3)	(4)	
	Baseline	Placebo: Other Spanish regions	Placebo: Never treated Catalan cohorts	Triple difference	
				Real	Pseudo
Panel A: Linear specification					
(linear) years of exposure	0.011^{**}	0.000	-0.008	0.020	-0.008
	(0.006)	*(0.005)*	*(0.014)*	*(0.015)*	*(0.014)*
Panel B: Dummy variables specification (reference = 0 years of exposure)					
years of exposure = 1	0.030	0.050	-0.037	0.070	-0.044
	(0.053)	*(0.040)*	*(0.079)*	*(0.095)*	*(0.078)*
years of exposure = 2	0.084^{*}	0.022	-0.100	0.194^{*}	-0.113
	(0.051)	*(0.040)*	*(0.100)*	*(0.107)*	*(0.094)*
years of exposure = 3	0.109^{**}	0.023	0.122	0.013	0.094
	(0.047)	*(0.041)*	*(0.110)*	*(0.116)*	*(0.106)*
years of exposure = 4	0.131^{***}	-0.024	-0.040	0.174^{*}	-0.047
	(0.048)	*(0.043)*	*(0.093)*	*(0.103)*	*(0.091)*
years of exposure = 5	0.181^{***}	0.027	-0.076	0.262^{**}	-0.087
	(0.056)	*(0.048)*	*(0.106)*	*(0.118)*	*(0.104)*
years of exposure = 6	0.206^{***}	0.026	-0.020	0.226^{**}	-0.030
	(0.046)	*(0.053)*	*(0.103)*	*(0.111)*	*(0.101)*
years of exposure = 7	0.119^{***}	0.075	-0.053	0.183	-0.069
	(0.044)	*(0.054)*	*(0.107)*	*(0.113)*	*(0.103)*
years of exposure = 8	0.185^{***}	0.047	-0.073	0.272^{**}	-0.092
	(0.049)	*(0.059)*	*(0.126)*	*(0.132)*	*(0.122)*
years of exposure = 9	0.184^{***}	0.072	-0.144	0.340^{**}	-0.162
	(0.060)	*(0.069)*	*(0.144)*	*(0.152)*	*(0.140)*
years of exposure = 10	0.191^{***}	0.047	-0.095	0.298^{*}	-0.113
	(0.058)	*(0.071)*	*(0.152)*	*(0.159)*	*(0.148)*
years of exposure = 11	0.155^{**}	-0.018	-0.098	0.264	-0.116
	(0.062)	*(0.076)*	*(0.152)*	*(0.160)*	*(0.147)*
years of exposure = 12	0.209^{***}	0.013	-0.040	0.261	-0.059
	(0.066)	*(0.079)*	*(0.165)*	*(0.173)*	*(0.160)*
years of exposure = 13	0.237^{***}	0.009	-0.083	0.331^{*}	-0.101
	(0.071)	*(0.086)*	*(0.177)*	*(0.186)*	*(0.172)*
years of exposure = 14	0.236^{***}	-0.014	-0.247	0.490^{**}	-0.262
	(0.074)	*(0.097)*	*(0.189)*	*(0.199)*	*(0.184)*
years of exposure = 15	0.217^{***}	-0.032	-0.069	0.293	-0.084
	(0.079)	*(0.100)*	*(0.196)*	*(0.207)*	*(0.190)*
years of exposure = 16	0.295^{***}	-0.035	-0.157	0.458^{**}	-0.170
	(0.073)	*(0.107)*	*(0.200)*	*(0.208)*	*(0.194)*
years of exposure = 17	0.176^{*}	0.013	-0.322	0.499^{**}	-0.331
	(0.090)	*(0.111)*	*(0.223)*	*(0.236)*	*(0.217)*
F-test of joint significance (p-value)	*0.000*	*0.578*	*0.009*	*0.121*	*0.013*
Number of observations	3,323	8,086	1,010	4,333	

***, **, * denote significance at the 1, 5 and 10 percent level. Robust standard errors in parentheses are clustered by year of birth, years of schooling and wave. Additional controls not shown: wave, gender, years of schooling, potential experience and its square, birth-cohort dummies. The model of Column 2 uses EU-SILC excluding observations from Catalonia and other bilingual regions. The model of Column 4 also includes a placebo cohort dummy and its interactions with the other controls.

Table 5: Reform effects on monthly earnings by schooling level (N=3,323)

	(1)	(2)	(3)	(4)
	Compulsory	Post-comp. Secondary	Tertiary	All
Panel A: Linear specification				
years of exposure at compulsory edu.	0.017**			0.018**
	(0.008)			*(0.008)*
years of exposure at post-comp secondary edu.		0.010		0.011
		(0.009)		*(0.009)*
years of exposure at tertiary edu.			0.011	0.011
			(0.009)	*(0.009)*
Panel B: Dummy variables specification (reference = 0 years of exposure)				
years of exposure at compulsory edu. = 1	-0.012			0.008
	(0.039)			*(0.041)*
years of exposure at compulsory edu. = 2	0.052			0.077*
	(0.038)			*(0.040)*
years of exposure at compulsory edu. = 3	0.064*			0.091***
	(0.033)			*(0.034)*
years of exposure at compulsory edu. = 4	0.067*			0.099**
	(0.039)			*(0.040)*
years of exposure at compulsory edu. = 5	0.091**			0.126***
	(0.043)			*(0.044)*
years of exposure at compulsory edu. = 6	0.101**			0.141***
	(0.044)			*(0.044)*
years of exposure at compulsory edu. = 7	0.118***			0.162***
	(0.044)			*(0.044)*
years of exposure at compulsory edu. = 8	0.121**			0.178***
	(0.049)			*(0.052)*
years of exposure at post-comp second. edu. = 1		0.033		0.048
		(0.043)		*(0.046)*
years of exposure at post-comp second. edu. = 2		-0.021		0.007
		(0.062)		*(0.067)*
years of exposure at post-comp second. edu. = 3		0.018		0.045
		(0.036)		*(0.041)*
years of exposure at post-comp second. edu. = 4		-0.026		-0.019
		(0.029)		*(0.037)*
years of exposure at post-comp second. edu. = 5		0.119		0.169*
		(0.085)		*(0.087)*
years of exposure at post-comp second. edu.= 6		0.060		0.078
		(0.045)		*(0.050)*
years of exposure at tertiary edu.= 3			-0.017	0.020
			(0.044)	*(0.054)*
years of exposure at tertiary edu.= 5			0.015	0.054
			(0.044)	*(0.049)*
F-test (p-value) of joint significance	*0.209*	*0.178*	*0.863*	*0.057*

****, **, * denote significance at the 1, 5 and 10 percent level. Robust standard errors in parentheses are clustered by year of birth, years of schooling and wave. Additional controls not shown: wave, gender, years of schooling, potential experience and its square, birth-cohort dummies.*

Table 6: Earnings effects of exposure at compulsory schooling

	(1)	(2)	(3)
Exposure at compulsory schooling	Observed	Potential	Placebo
	Panel A: Linear specification		
years of exposure	0.017**	0.018***	-0.002
	(0.008)	*(0.006)*	*(0.008)*
	Panel B: Dummy variables specification (reference = 0 years of exposure)		
years of exposure = 1	-0.012	-0.006	-0.015
	(0.039)	*(0.042)*	*(0.031)*
years of exposure = 2	0.052	0.047	-0.078*
	(0.038)	*(0.031)*	*(0.040)*
years of exposure = 3	0.064*	0.061*	-0.005
	(0.033)	*(0.032)*	*(0.046)*
years of exposure = 4	0.067*	0.065	-0.021
	(0.039)	*(0.045)*	*(0.039)*
years of exposure = 5	0.091**	0.082*	0.015
	(0.043)	*(0.045)*	*(0.057)*
years of exposure = 6	0.101**	0.097**	-0.018
	(0.044)	*(0.042)*	*(0.049)*
years of exposure = 7	0.118***	0.120***	0.006
	(0.044)	*(0.046)*	*(0.058)*
years of exposure = 8	0.121**	0.121**	0.001
	(0.049)	*(0.057)*	*(0.056)*
F-test of joint significance (p-value)	*0.209*	*0.064*	*0.604*
Number of observations	3,308	3,308	8,086

****, **, * denote significance at the 1, 5 and 10 percent level. Robust standard errors in parentheses are clustered by year of birth, years of schooling and wave in Columns (1), by year of birth and wave in Column (2) and by year of birth, wave and region in Column (3). Additional controls not shown: wave, gender, years of schooling, potential experience and its square, birth-cohort dummies. Column (3) uses EU-SILC data excluding observations from Catalonia and other bilingual regions.*

Table 7: Mechanisms and heterogeneous effects by regional origins (N=3,323)

	(1)	(2)	(3)	(4)
	Earnings	Public sector	Spanish only	Catalan proficiency
Panel A: Homogeneous effects				
years of exposure at compulsory school	0.018*	-0.004	0.000	0.016**
	(0.009)	*(0.008)*	*(0.008)*	*(0.007)*
Panel B: Heterogeneous effects by regional origins				
years of exposure at compulsory school	0.016**	-0.004	0.004	0.003
	(0.006)	*(0.009)*	*(0.007)*	*(0.007)*
non-Catalan origins	-0.025	-0.020	0.399***	-0.228***
	(0.019)	*(0.020)*	*(0.028)*	*(0.020)*
years of exposure at compulsory school × non-Catalan origins	0.005	-0.001	0.002	0.023***
	(0.003)	*(0.005)*	*(0.005)*	*(0.004)*

****, **, * denote significance at the 1, 5 and 10 percent level. Robust standard errors in parentheses are clustered by year of birth and wave. Additional controls not shown: wave, gender, years of schooling, potential experience and its square, birth-cohort dummies. The treatment is assigned according to potential exposure.*

Table 8: Heterogeneous earnings effects by parental education and regional origins

	(1)	(2)
	Catalan origins	Non-Catalan origins
years of exposure at compulsory school	0.014	0.009
	(0.014)	*(0.008)*
low parental background	-0.108*	-0.090***
	(0.058)	*(0.021)*
years of exposure at compulsory school× low parental background	-0.002	0.018***
	(0.010)	*(0.005)*
	1,276	1,978

****, **, * denote significance at the 1, 5 and 10 percent level. Robust standard errors in parentheses are clustered by year of birth, years of schooling and wave. Additional controls not shown: wave, gender, years of schooling, potential experience and its square, birth-cohort dummies. The treatment is assigned according to potential exposure. Information on parental background is missing in 69 cases.*

Table 9: Compulsory exposure, Catalan proficiency and earnings (N=3,323)

	(1)	(2)	(3)	(4)	
Model	Baseline with proficiency	Baseline with proficiency and reform shifter	Reduced form	IV-LATE	
				Earnings	Catalan Proficiency
Panel A: Interaction with respondent's language					
Catalan Proficiency	0.058**	0.050*		0.219**	
	(0.026)	*(0.029)*		*(0.088)*	
years of exposure at comp. school	0.017***	0.017***	0.015**	0.015**	0.004
	(0.006)	*(0.006)*	*(0.006)*	*(0.006)*	*(0.006)*
Spanish-only speaker		-0.016	-0.058***	0.024	-0.371***
		(0.013)	*(0.013)*	*(0.025)*	*(0.026)*
years of exposure at comp. school × Spanish-only			0.008***		0.037***
			(0.003)		*(0.005)*
Panel B: Interaction with regional origins					
Catalan Proficiency	0.058**	0.058**		0.214	
	(0.026)	*(0.026)*		*(0.153)*	
years of exposure at comp. school	0.017***	0.017***	0.016**	0.015**	0.003
	(0.006)	*(0.006)*	*(0.006)*	*(0.006)*	*(0.007)*
non-Catalan origins		0.001	-0.025	0.024	-0.228***
		(0.014)	*(0.019)*	*(0.028)*	*(0.020)*
years of exposure at comp. school× non-Catalan			0.005		0.023***
			(0.003)		*(0.004)*
Panel C: Interactions with respondent's language and regional origins					
Catalan Proficiency	0.058**	0.051*		0.222**	
	(0.026)	*(0.029)*		*(0.088)*	
years of exposure at comp. school	0.017***	0.017***	0.015**	0.015**	-0.000
	(0.006)	*(0.006)*	*(0.006)*	*(0.006)*	*(0.006)*
non-Catalan origins		0.008	-0.002	0.019	-0.098***
		(0.016)	*(0.022)*	*(0.018)*	*(0.018)*
years of exposure at comp. school× non-Catalan			0.002		0.011***
			(0.004)		*(0.003)*
Spanish-only speaker		-0.020	-0.057***	0.016	-0.322***
		(0.014)	*(0.016)*	*(0.024)*	*(0.029)*
years of exposure at comp. school× Spanish-only			0.007**		0.031***
			(0.004)		*(0.005)*

***, **, * denote significance at the 1, 5 and 10 percent level. Robust standard errors in parentheses are clustered by year of birth, years of schooling and wave. Additional controls not shown: wave, gender, years of schooling, potential experience and its square, birth-cohort dummies. The treatment is assigned according to potential exposure

Appendix

Table A1: Net monthly earnings by wave, Euros 2006

	wave 2006	**wave 2011**	**waves 2006-2011**
w< 300	0.92	0.79	0.87
300 <w < 451	2.20	2.99	2.47
450 < w < 601	5.17	2.46	4.24
600 < w < 751	8.15	5.71	7.31
750 < w < 901	14.42	15.73	14.87
900 < w < 1051	14.51	10.46	13.12
1050 < w < 1201	19.08	16.34	18.15
1200 < w < 1501	17.12	19.33	17.88
1500 < w < 1801	9.66	15.99	11.83
1800 < w < 2401	6.00	7.56	6.53
2400 < w < 3001	1.83	1.85	1.84
3000 < w < 3601	0.55	0.44	0.51
w >3600	0.41	0.35	0.39
Total	100	100	100
Estimated average net monthly earnings	1,155	1,228	1,180
Observed average net monthly earnings		1,232	
Observed average hourly wage— 2011 prices		8.48	

Note: average monthly earnings are estimated from reported monthly earnings in brackets by interval regression on a constant.

Table A2: descriptive statistics by subsamples

	employed 2006		employed 2011		employed 2006-2011		not employed 2006		not employed 2011	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
wave 2011	0	--	1	--	0.34	*0.47*	0	--	1	--
male	0.52	*0.50*	0.47	*0.50*	0.50	*0.50*	0.20	*0.40*	0.37	*0.48*
age	34.3	*5.95*	39.3	*5.99*	36.0	*6.41*	35.9	*5.89*	39.9	*6.19*
years of schooling	12.3	*3.27*	12.7	*3.53*	12.4	*3.36*	10.8	*3.37*	10.2	*3.59*
years of compulsory schooling	7.94	*0.39*	7.88	*0.58*	7.92	*0.47*	7.84	*0.68*	7.62	*1.07*
years of post-comp. secondary schooling	3.26	*2.13*	3.26	*1.95*	3.26	*2.07*	2.41	*2.34*	2.07	*2.25*
years of tertiary schooling	1.14	*1.93*	1.59	*2.19*	1.30	*2.04*	0.62	*1.57*	0.60	*1.53*
years of exposure	6.90	*5.49*	7.29	*5.77*	7.03	*5.59*	4.69	*4.98*	5.28	*5.28*
years of exposure at compulsory schooling (pot.)	3.54	*3.37*	3.59	*3.37*	3.56	*3.37*	2.65	*3.21*	3.28	*3.38*
years of exposure at compulsory schooling (obs.)	3.51	*3.35*	3.53	*3.35*	3.52	*3.35*	2.58	*3.19*	3.13	*3.33*
years of exposure at post-comp. secondary schooling	2.47	*2.27*	2.51	*2.15*	2.48	*2.23*	1.66	*2.17*	1.60	*2.16*
years of exposure at tertiary schooling	0.92	*1.77*	1.25	*1.98*	1.04	*1.85*	0.45	*1.34*	0.55	*1.47*
potential experience (age - years of schooling - 6)	16.0	*6.95*	20.5	*7.28*	17.5	*7.39*	19.0	*7.00*	23.6	*7.78*
low parental background	0.72	*0.45*	0.73	*0.45*	0.72	*0.45*	0.82	*0.38*	0.80	*0.40*
non-Catalan origins (both parents born outside Catalonia)	0.39	*0.49*	0.39	*0.49*	0.39	*0.49*	0.48	*0.50*	0.49	*0.50*
Catalan proficiency (speak and write in Catalan)	0.84	*0.36*	0.94	*0.24*	0.88	*0.33*	0.73	*0.44*	0.89	*0.31*
Spanish-only speaker	0.31	*0.46*	0.27	*0.44*	0.30	*0.46*	0.46	*0.50*	0.44	*0.50*
weekely hours of work	39.2	*9.29*	38.5	*8.70*	38.9	*9.09*	--	--	--	--
white-collar high-skilled occupation	39.2	*9.29*	38.5	*8.70*	38.9	*9.09*	--	--	--	--
public sector employment	0.20	*0.40*	0.26	*0.44*	0.22	*0.42*	--	--	--	--
local unemployment rate at age 16	15.6	*6.58*	15.2	*6.76*	15.5	*6.64*	15.9	*6.85*	15.6	*6.59*
LNA Reform exposure by birth cohort										
Spanish-only at school (1961-65)	0.20	*0.40*	0.20	*0.40*	0.20	*0.40*	0.26	*0.44*	0.26	*0.44*
exposure only from secondary education (1966-69)	0.18	*0.38*	0.17	*0.37*	0.17	*0.38*	0.24	*0.43*	0.15	*0.36*
partial exposure at all levels (1970-76)	0.37	*0.48*	0.38	*0.48*	0.37	*0.48*	0.32	*0.47*	0.34	*0.47*
fully exposed to LNA (1977-1982)	0.26	*0.44*	0.26	*0.44*	0.26	*0.44*	0.18	*0.39*	0.25	*0.43*
Number of observations	2,185		1,138		3,323		359		449	

Source: Survey on Habits and Living Conditions of the Catalan Population (ECVHP).

Table A3: Selected descriptive statistics of the placebo samples (never-treated cohorts from ECVHP and contemporaneous cohorts from EU-SILC from Spanish non-bilingual regions)

	ECVHP (2006-2011)				EU-SILC (2006-2011)	
	Baseline sample		**Placebo sample: never treated**		**Placebo sample: other Spanish regions**	
Birth-cohort	*1961-1982*		*1945-1960*		*1961-1982*	
	Mean	***S.D.***	***Mean***	***S.D.***	***Mean***	***S.D.***
net monthly earnings (in 2006 prices)	1,180*	--	1,437*	--	1335	*809.5*
wave 2011/2010	0.34	*0.47*	0.35	*0.48*	0.49	*0.50*
male	0.50	*0.50*	0.55	*0.50*	0.53	*0.50*
age	36.05	*6.41*	53.07	*4.41*	37.39	*6.78*
years of schooling	12.48	*3.36*	11.40	*3.95*	12.34	*4.23*
potential experience (age - years of schooling - 6)	17.58	*7.39*	35.67	*6.17*	19.05	*8.24*
years of exposure (real/placebo)	7.03	*5.59*	8.52	*5.01*	6.77	*5.87*
years of exposure at compulsory schooling (pot.)	3.56	*3.37*	--	--	3.22	*3.41*
birth cohorts (by real/placebo LNA Reform exposure)						
never-treated cohort (1945-60)	--	--	1.00	*0.00*		
Spanish-only at school (1961-65)	0.20	*0.40*	--	--	0.25	*0.43*
exposure only from secondary education (1966-69)	0.17	*0.38*	--	--	0.19	*0.39*
partial exposure at all levels (1970-76)	0.37	*0.48*	--	--	0.31	*0.46*
fully exposed to LNA (1977-1982)	0.26	*0.44*	--	--	0.25	*0.43*
Number of observations	3,323		1,010		8,086	

Note: summary statistics reported in the first two columns refer to the data from the Survey on Habits and Living Conditions of the Catalan Population (ECVHP), as in table 1. Data in the second column are taken from the same survey, using only observations of individuals born between 1945 and 1960 (never-treated cohorts) who meet the other selection criteria. Summary statistics reported in the last column refer to the EU-SILC, waves 2006 and 2011, excluding individuals from Catalonia and other bilingual regions or born outside Spain. Average monthly earnings in ECVHP *are estimated from reported monthly earnings in brackets by interval regression on a constant.*

Table A4: Coefficient estimates from regressions of monthly and hourly net earnings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
waves:	**2006-2011**	**2011**	**2011**	**2006**	**2006-2011**	**2011**	**2011**	**2006**
Earnings definition	**monthly**	**hourly**	**monthly**	**monthly**	**monthly**	**hourly**	**monthly**	**monthly**
constant	5.657***	0.931***	5.697***	5.673***	5.502***	0.410	5.294***	5.592***
	(0.063)	*(0.101)*	*(0.113)*	*(0.080)*	*(0.199)*	*(0.291)*	*(0.354)*	*(0.272)*
wave 2011	-0.003				-0.034			
	(0.018)				*(0.032)*			
male	0.297***	0.102***	0.274***	0.310***	0.298***	0.104***	0.274***	0.311***
	(0.015)	*(0.021)*	*(0.024)*	*(0.019)*	*(0.015)*	*(0.021)*	*(0.024)*	*(0.019)*
years of schooling	0.060***	0.061***	0.067***	0.057***	0.058***	0.067***	0.065***	0.054***
	(0.003)	*(0.003)*	*(0.004)*	*(0.004)*	*(0.005)*	*(0.007)*	*(0.008)*	*(0.007)*
years of exposure					0.011**	0.011	0.018*	0.008
					(0.006)	*(0.008)*	*(0.010)*	*(0.008)*
potential experience	0.040***	0.021**	0.027***	0.043***	0.052***	0.035**	0.049***	0.054***
	(0.005)	*(0.008)*	*(0.009)*	*(0.007)*	*(0.009)*	*(0.015)*	*(0.018)*	*(0.014)*
potential experience^2	-0.001***	-0.000	-0.000*	-0.001***	-0.001***	-0.000	-0.001**	-0.001***
	(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
Spanish-only at school (1961-65)						*reference category*		
exposure only from secondary education (1966-69)					-0.048	0.043	-0.037	-0.057
					(0.037)	*(0.053)*	*(0.070)*	*(0.044)*
partial exposure at all levels (1970-76)					-0.047	0.091	-0.037	-0.057
					(0.059)	*(0.080)*	*(0.106)*	*(0.073)*
fully exposed to LNA (1977-1982)					-0.055	0.092	-0.057	-0.061
					(0.083)	*(0.118)*	*(0.147)*	*(0.099)*
R^2	0.275	0.260	0.291	0.265	0.277	0.264	0.296	0.266
Number of observations	3,323	1,119	1,138	2,185	3,323	1,119	1,138	2,185

*Note: *** significant at 0.01, ** significant at 0.05, * significant at 0.1; standard errors adjusted for year of birth-years of schooling-wave clusters are shown in parenthesis. Earnings in pooled 2006-2011 regressions are expressed in 2006 prices. The R^2 reported for interval regressions is the McKelvey & Zavoina's R, while for other outcomes (estimated by OLS) it corresponds to the adjusted R^2.*