Economic Analysis and Policy 80 (2023) 429–449

Contents lists available at ScienceDirect

Economic Analysis and Policy

journal homepage: www.elsevier.com/locate/eap

Modelling economic policy issues

# Does geographical exposure to language learning centres affect language skills and labour market outcomes in a bilingual city?

Antonio Di Paolo *, Bernat Mallén

*Bernat Mallén: AQR-IREA, Department of Econometrics, Statistics and Applied Economics Faculty of Economics & Business, University of Barcelona, Campus Diagonal, Torre 4, Avinguda Diagonal 690, 08034 Barcelona, Spain*

ARTICLE INFO

*Article history:*
Received 9 February 2023
Received in revised form 27 July 2023
Accepted 1 August 2023
Available online 1 September 2023

*JEL classification:*
I38
J24
R23
Z13

*Keywords:*
Local language training
Language skills
Bilingualism
Labour market outcomes

ABSTRACT

In this paper, we investigate the effects of geographical exposure to local language training centres in a bilingual urban labour market, the Metropolitan Area of Barcelona, exploiting the implementation of a language policy that provided publicly subsidized language courses for adults. Our variable of interest consists in a measure of spatial availability of language schools that captures potential exposure and its expansion over time. First, we focus on the formation of local language skills, adopting a reduced-form approach. Our results indicate that exposure to language learning opportunities has a positive but quantitatively modest effect on local language skills, since individuals residing in neighbourhoods with a higher supply of language centres are more likely to be able to speak and write in Catalan. The effect is very robust to falsification exercises and several sensitivity checks and is driven by younger individuals, and especially by those who were born in Catalonia. Second, we analyse whether accessibility to language centres also affects employment, working hours, employment sector, and occupation. Our results indicate that, although skills in Catalan are positively and significantly associated with improved labour market outcomes, spatial accessibility to language centres is not directly related to any of them. This lack of effect is possibly due to the fact that the impact of geographical exposure on language skills is too small in size to improve performance in the local labour market.

© 2023 Economic Society of Australia, Queensland. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

This work analyses whether geographical exposure to language learning centres affects the formation of language skills and labour market outcomes in a bilingual urban labour market, the Metropolitan Area of Barcelona. The empirical setup exploits the implementation of a language policy that provided publicly subsidized language courses to adult individuals, in the bilingual region of Catalonia, aimed at fostering proficiency in Catalan, the local language. Existing research in language economics provides compelling evidence in favour of the general idea that language matters, since better language proficiency improves several dimensions of individuals' economic and social wellbeing, similarly to other forms of human capital such as schooling (Chiswick and Miller, 2007). The majority of existing works focus on the behaviour of

* Corresponding author.
*E-mail addresses:* antonio.dipaolo@ub.edu (A. Di Paolo), bernatmallen@ub.edu (B. Mallén).

https://doi.org/10.1016/j.eap.2023.08.001
0313-5926/© 2023 Economic Society of Australia, Queensland. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

migrants, based on the underlying hypothesis that language proficiency in the host country language affects the economic and social assimilation of migrants positively (Bleakley and Chin, 2004, 2010). However, investigating the formation of language skills and their relationship with socioeconomic outcomes is not only relevant for the case of migrants learning the language of the host country or for the case of natives who acquire skills in foreign languages. Indeed, studying the social and economic returns to linguistic competences, and most importantly, the role of language policies, is of extreme importance in multilingual societies where different languages coexist and are entitled to an official status. Actually, the issue of linguistic diversity and the vitality and preservation of local or minority languages in multilingual countries or regions is gaining importance over time among policymakers and in society in general. In Europe alone, examples include some of the former soviet republics (like Ukraine, Estonia, and Kazakhstan), Belgium, Cyprus, Wales, the region of South Tyrol in northern Italy, and Spain. In this last country, for example, since the adoption of the Democratic Constitution of 1978, four distinct local languages (Basque, Catalan, Galician, and Aranese, a variety of Occitan) are entitled to co-official status in specific regions, together with Spanish, the official language of the whole country.

Consistently, a growing number of papers point out that skills in local languages are rewarded in the labour market and positively affect other non-monetary outcomes. Moreover, language policies aimed at promoting the knowledge and use of local languages tend to have longstanding impacts on many dimensions of life. This is relevant not only for foreign migrants but also for internal migrants coming from non-bilingual regions of the same country, as well as for native-born individuals with limited competences in the local language (e.g. in the Spanish case, not all individuals born in bilingual regions are necessarily proficient in the corresponding local language). The case of the bilingual Spanish region of Catalonia, which implemented several language policies to foster the use of its local language (Catalan) after the end of the Franco regime, represents an ideal setting to investigate these issues. Indeed, some works examined the impacts of the language policy implemented in the 1980' by the Catalan Government to stimulate bilingualism and biliteracy among inhabitants of Catalonia regardless of language background and regional origin. The so-called Language Normalization Act (LNA) of 1983 was motivated by the high share of internal migrants from other Spanish regions with no knowledge of Catalan, by the lack of written skills of native Catalan speakers, and by the aspiration to recover and preserve the vitality of the local language. Indeed, the main feature of this reform consists in the introduction of Catalan as a medium of instruction at school together with Spanish (i.e. bilingual education), which has been investigated in recent works. This language policy and the increase in skills it induced (especially) among the non-Catalan speaking population have been found to impact different social and economic outcomes. For example, Cappellari and Di Paolo (2018) show that the exposure to bilingualism at school induced by the reform increased the returns to each additional year of schooling, especially among individuals with non-Catalan origins. Moreover, Caminal and Di Paolo (2019) highlight that the policy-induced increase in oral skills in Catalan among native Spanish speakers affected partnership formation, by raising the propensity to form a linguistically mixed couple (thus, reducing social segregation). Finally, a recent work by Caminal et al. (2021) reveals that the language-in-education reform, by improving skills in the local language among native Spanish speakers, generated intergenerational spillovers by affecting the language spoken with the children.

The findings from this recent research indicate that language policies that promote the formation of local language skills matter in multilingual societies. However, the evidence regarding the impact of local language policies that is available so far only refers to interventions targeted to school-age populations. That is, nothing has been said regarding the effects of the provision of language training programmes targeted to the adult population in multilingual territories. To fill this gap in the literature, in this paper we investigate the effects of a language policy implemented in Catalonia at the end of the 1980' with the aim of providing publicly subsidized (local) language training programmes. These Catalan language courses were targeted not only to adult migrants coming from other Spanish regions or from foreign countries but also to native-born individuals with limited knowledge of the local languages who were schooled before their introduction as subjects and/or a medium of instruction at school.

This work also contributes to the emerging, although still scarce, literature on the impacts of language training provision programmes, to understand whether these kinds of policies are effective in fostering the linguistic, cultural, and socioeconomic assimilation of newcomers. The few recent works on the effects of participation in language courses have focused on the effects on adult migrants and refugees (Åslund and Engdahl, 2018; Giesecke and Schuss, 2019; Kivi et al., 2020; Lochmann et al., 2019; Lang, 2021; Foged et al., 2022; Pont-Grau et al., 2023; Heller and Slungaard Mumma, 2023), which generally report positive results. We are, however, the first to provide evidence regarding large-scale language training programmes in a bilingual setting. As explained in more details below, given the absence of longitudinal individual data covering periods before and after the implementation of the language policy, we evaluate the effects of geographical exposure to language training centres exploiting information about the location of language centres and individuals' place of residence. Nevertheless, we consider that our analysis is quite relevant for policymaking, since we provide evidence that is useful for determining the spatial distribution of language training centres as well as the extent to which different subgroups of individuals react to an increase in the local supply of language courses.

Specifically, we investigate the effects of geographical accessibility of language learning centres, considering first its impact on the formation of skills in Catalan and, second, whether it affects labour market outcomes in a bilingual urban labour market. We exploit historical administrative information about the geographical location of local language schools since their creation in 1989, focussing on the Metropolitan Area of Barcelona. This is a bilingual urban labour market, where a large majority of language training centres were concentrated during the initial phase of implementation of this language policy. Geolocation information about the location of Catalan language centres is merged with data from the

Microcensus[1] of 2001, which contains individual-level information regarding several sociodemographic characteristics, local language skills, and labour market outcomes, plus the census tract identifier of the current place of residence (which is the finest geographical unit of analysis for available data).

The empirical analysis is based on a reduced-form approach that aims at estimating the causal effects of geographical exposure to language schools since their creation as well as the expansion of its geographical coverage over time. Specifically, our measure of exposure to language learning opportunities consists in the number of language centres located within a given radius from the centroid of the census tract of residence between during the period 1989–2001. In this way, we estimate the effects of local availability of language schools by comparing skills in Catalan (and labour market outcomes) of individuals residing in neighbourhoods with different degrees of potential accessibility of language learning centres. Therefore, our approach provides Intention-to-Treat (ITT) estimates, which represent the only policy-relevant parameter we are able to identify given the absence of information on participation in local language courses.[2] Indeed, we provide credible evidence that corroborates the causal interpretation of the corresponding OLS coefficient as the causal effect of geographical exposure to local language centres.

We focus our analysis on both individuals born in Catalonia and internal migrants coming from other Spanish regions, who were the most relevant users of such public services during the first decade of implementation of the language policy. Our results indicate that proximity to language training opportunities has a modest, but highly robust, effect on the formation of local language skills, fostering the probability of being able to speak and write in the local language. The results are stable with respect to the inclusion of census-tract level controls as well as district fixed effects. Falsification exercises based either on younger cohorts who were exposed to Catalan at school (and were therefore not a "target" of adults' language schools) or on the random assignment to fake census tracts provide results that speak in favour of the causal interpretation of our main estimations. The effects are not driven by language centres located in specific types of hosting institutions (public schools, community centres, and other municipality buildings) and the spatial effect shows a clear distance-decay pattern. The impacts on language skills are strongly heterogeneous, since the results suggest that only young individuals and, especially, those who were born in Catalonia benefited from geographical accessibility to language training centres. Moreover, the heterogeneous analysis by local characteristics provides some mild evidence suggesting that the impact of exposure appears to be higher for individuals residing in areas with a higher fraction of internal migrants (or a lower fraction of individuals who are proficient in Catalan). Finally, the evidence regarding whether spatial exposure to language schools translated into better labour market outcomes is far from conclusive. In fact, although our results indicate that skills in Catalan are significantly associated with better labour market outcomes (especially employment and occupation), we did not detect any significant effect of exposure to language courses on the labour market outcomes considered in a reduced form framework. This is possibly because the effect of language skills is too modest, in quantitative terms, to induce any direct improvement in labour market performance in the bilingual urban labour market.

The rest of the paper proceeds as follows. In Section 2, we provide the relevant institutional background, describing the context and implementation of the language policy under investigation. Section 3 includes an explanation of the different data sources we use in this work and descriptive statistics. Section 4 describes the empirical strategy followed to investigate the effect of exposure to language training centres. Section 5 provides the results of our analysis and several robustness checks, and Section 6 concludes the paper.

## 2. Institutional background

In this paper, we analyse the effect of geographical exposure to (local) language courses on language skills formation and labour market outcomes by exploiting the public provision of Catalan courses in the Spanish region of Catalonia, starting at the end of the 1980'. 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. Although this trend reversed 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 that occurred in Spain during the 1930s represented a major negative shock for the public use and vitality of Catalan. 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. Moreover, the new Constitution allowed regional governments to recover and stimulate the public and private use of their own languages. Several policies were implemented in Catalonia, as well as in other bilingual Spanish regions such as the Basque Country, to favour bilingualism of the population and to recover the use and vitality of local languages.

[1] As explained below, we also aggregate information from other statistics (censuses of 1991 and 2001 and the Population Statistic of 1996) to construct local-level control variables.

[2] Two other works on migrants also exploit information on the location of language schools to construct instruments for participation, both considering the German case. Specifically, Lang (2021) uses the number of participants in training programmes for each "job centre" over the number of foreign born in the corresponding area, while Giesecke and Schuss (2019) use the density of available slots in language courses at the county level.

However, the sociodemographic landscape of Catalonia after the Franco regime represented the main challenge in pursuing these aims. 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. internal migrants and their offspring) were native Spanish speakers, with limited or no knowledge of Catalan. This was particularly the case 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, but writing skills were scarce or even absent for a large fraction of native Catalan speakers.[3] It was against this background of linguistic segmentation that the local government used language policies as the main instrument of "language normalization". The main target was to guarantee bilingualism and biliteracy regardless of language background or regional origins.

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. Some years later, the Language Normalization Act (LNA) implemented by the Catalan Government (*Generalitat*) in 1983 introduced a sharp change by establishing Catalan as a medium of instruction in primary and secondary schools, alongside Spanish, making the education system effectively bilingual. The reform's objective was 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. Indeed, this language-in-education policy significantly shaped the language skills of the affected cohorts, especially among native Spanish speakers, and had important implications for labour market and social outcomes (for more details, see Cappellari and Di Paolo, 2018; Caminal and Di Paolo, 2019; and Caminal et al., 2021).

Thus, a huge effort was devoted to spreading the knowledge of Catalan among new generations. However, a significant portion of the adult population was not fully proficient in Catalan. This group included mainly monolingual Spanish speakers but also Catalan speakers who were not able to write in their native language. In the meanwhile, the regional labour market was characterized by a rising demand for local language skills. This was a result of the increasing use of the local language as a medium of communication in private business, on the one hand, and the establishment of Catalan as the main language in regional and local public institutions after the implementation of the LNA, on the other.[4]

Consequently, the availability of Catalan language courses for adults increased substantially during the 1980s, thanks to small public or even private initiatives in the city of Barcelona and its metropolitan area. However, this modest supply of language courses was clearly insufficient to guarantee equal chances to gain access to language training and to shape local language skills of the adult population. With the aim of stimulating the formation of local language skills among adults and mitigating the insufficient endowment of the workforce with this relevant asset, in 1988 the Catalan Government implemented a novel language policy aimed at providing language courses at different levels targeted to the adult population. The policy was implemented through the creation of the Consortium for the Linguistic Normalization (*Consorci per a la Normalització Lingüística*, CPLN from now on), which centralized the organization of courses, under the supervision of the General Board of Language Policy (*Direcció General de Política Lingüística*). City councils, provincial deputations, and the Catalan government itself, which guaranteed the funding of the CPLN, participated in the consortium. Until 1994, by law, 65% was funded by the *Generalitat* and 35% by other local entities; from 1995 the rule was relaxed and resulted in approximately 67% funding by the Catalan government and 33% by the local entities. To stimulate participation, language courses were generally offered free of charge except for advanced levels, although the price of these courses was highly publicly subsidized.[5]

In this paper, we focus on the first decade of existence of the CPLN, covering the academic courses from 1989/90 to 2000/01 (starting in September). This choice is driven not only by the fact that the empirical analysis exploits data from the 2001 Microcensus, containing individual-level data on skills in Catalan, but also by other features that are useful for our identification strategy. A first interesting aspect of the language policy under investigation concerns the creation of a network of language courses. The initial deployment of CPLN's Catalan courses was markedly concentrated in the Metropolitan Area of Barcelona. Among the 19 municipalities that participated in the CPLN since 1989, 11 belong to the Metropolitan Area. Moreover, another 12 municipalities joined the consortium in 1990, of which seven also form part of this large agglomeration of cities around Barcelona. Therefore, the geographical delimitation of our empirical analysis coincides with the border of the Metropolitan Area of Barcelona, which also simplifies some technical issues related to the

[3] For example, in 1986, only 31% of the Catalan population was able to write in Catalan, and the share was substantially lower for individuals born in other Spanish regions (7% versus 44% among individuals born in Catalonia). Oral skills were also far from being widespread (64%). Five years later, in 1991, the percentage of individuals who were able to write (speak) in Catalan was 40% (69%), and in 2001 up to 52% (77%) of the population had written (oral) skills in the local language.

[4] The same reform also allowed for the creation of a public Catalan TV channel (TV3) and incentivized the presence of newspapers written in Catalan. The demand for Catalan skills in the local labour market was also enhanced by the subsequent Language Policy Law (LPL) of 1998, which affected the relevance of Catalan in the labour market. First, a proficiency of level C in Catalan was set as the prerequisite to enter public sector jobs. It is worth mentioning that this proficiency certificate was automatically awarded at the end of compulsory education to individuals who had entered school since the 1978/79 school year, i.e. those born since 1972. Second, it increased the incentives to foster the use of Catalan in private business, especially among those firms with direct commercial contacts with the Catalan public sector and/or service firms with a strong level of contact with the public (e.g. the restaurant and hotel industry). That is, the LPL introduced the institutional basis for the creation of a bilingual labour market.

[5] Here we refer to the general courses, open to everybody. In addition to the general courses, the CPLN also offered sectorial courses, mainly to civil servants of regional and local institutions, who needed to attain the required level of Catalan language.

use of census tracts as the territorial unit of analysis (see Section 4). Second, classes were held in public buildings belonging to public schools, community centres, and other municipality infrastructures, on the basis of agreements between the CPLN and the hosting institutions. Therefore, the initial locations of language courses and their spatial distribution in general were generally driven by the availability of suitable infrastructures rather than by potential demand-side factors. This aspect speaks in favour of the exogeneity of the geographical availability of language courses with respect to individuals' residential locations. However, in the empirical analysis we still provide several checks regarding this possible issue as well as regarding the fact that hosting institutions could also provide other services, besides language courses, that might have affected the outcomes we investigate.

Third, until the beginning of the twenty-first century, the large majority of language students were born in Catalonia or in other Spanish regions.[6] This enables restricting the empirical analysis to individuals born in Catalonia or in other Spanish regions, thus limiting the presence of unobserved heterogeneity. Indeed, 10,173 out of 17,092 individuals (59%) who enrolled on the course in 1990/91 were born in the region. The share of foreigners in 1992/93 (the first year with available information) was just 10%. It is important to notice that the distribution across language levels varied substantially by origin: while non-Catalan students generally took beginner courses, Catalan students were mostly enrolled on intermediate and advanced courses. Indeed, among the latter there were, on the one hand, some monolingual Spanish speakers born in Catalonia (i.e. second-generation internal migrants) who had limited or no knowledge of Catalan because they were schooled before the LNA. On the other hand, a relevant share of individuals from Catalan-speaking families, who were already orally proficient, also wanted to improve their written skills in their native language. The proportion of students born in Catalonia did not change much during the decade. Actually, in the last school year we consider in this work (2000/01), students born in Catalonia represented 56% of those enrolled, while those born in other regions of Spain represented 19.78% and the rest were from other countries (7.68% from the European Union and 7.58% from outside it). However, the number of foreign students taking Catalan courses increased sharply afterwards. Indeed, in the academic year 2003/04, around 40% of enrolled students were migrants from other countries (mostly Latin American). The proportion of foreign migrants rose to 76% in 2005/06 and remained stable until today. Coinciding with the large inflow of foreign migrants, the CPLN also started a general rationalization and (territorial) reorganization of language courses, increasing the supply in areas characterized by a high share of migrants as well as in other medium and small cities of Catalonia.

## 3. Data and descriptive statistics

The empirical analysis presented in this paper focuses on the Metropolitan Area of Barcelona, a wide agglomeration of municipalities located around Barcelona, the capital of the Spanish region of Catalonia. It represents the second most populated urban area of Spain and one of the densest agglomerations in Europe and is composed of 35 different municipalities. The Metropolitan Area of Barcelona is interesting for our analysis for several reasons besides its size. First, most of the internal migrants (and subsequently foreign migrants) who moved to Catalonia are located within the borders of the Metropolitan Area, mostly in specific neighbourhoods of the capital and in some surrounding municipalities. Second, it covers virtually all the municipalities that participated in the CPLN since its creation in 1989 throughout the period we consider in this paper (academic years 1989/90 to 2000/01).[7] Third, focussing on the metropolitan area simplifies the process of tracking changes in the geographical definition of census tracts that occurred over time, which, as explained later, is fundamental for our work.

We combine several databases to answer our research questions. On the one hand, we exploit individual-level data proceeding from the Spanish Microcensus of 2001, a 5% random sample of the full sample for the year. This dataset is especially useful for our purposes for several reasons. First, it contains information about several sociodemographic characteristics such as gender, year of birth, place of birth, year of arrival in Catalonia for those who were born elsewhere, the municipality and residential dwelling of all individuals, completed education, marital status, and the number of children and adults in the household as well as self-reported knowledge of Catalan. The last variable is crucial in our analysis and is coded as a cumulative combination of different language competences. Specifically, respondents report whether they can (1) understand, (2) read, (3) speak, (4) read and speak, or (5) read, speak, and write in Catalan.[8] The Microcensus also contains information on employment status, weekly hours of work, and occupation and industry (both at two digits), which we use as labour market outcomes. Second, it refers to a period that enables a focus on the initial phase of deployment of local language training programmes provided by the CPLN.[9] Third, it makes it possible to exploit

[6] Detailed information is contained in the yearly reports of the CPLN, which are available at https://www.cpnl.cat/transparencia/memories/.

[7] It is worth noting that language courses are scheduled on the basis of the academic year, which means that in practice the starting year was September 1989. Moreover, we exclude from the analysis two small municipalities (Montgat and Tiana) located at the periphery of the metropolitan area, since they belong to a different county of the region and did not participate in the creation of the CPLN during the period considered in this paper.

[8] Unfortunately, as already mentioned, there is no information regarding participation in local language courses.

[9] Moreover, drawing on data from 2001 (and previous years) also enables us to circumvent the issues related to the subsequent large waves of arrival of foreign migrants (Gonzalez and Ortega, 2011), not only from Spanish-speaking Latin American countries. We are therefore able to focus on Spaniards, born either inside or outside Catalonia, who account for the large majority of the population (as well as users of Catalan courses), who are also all proficient in Spanish.

information about the census tract of the place of residence, which corresponds to the geographical dimension considered in this paper (and the most spatially disaggregate unit considering the available data).

On the other hand, we also use aggregate data from two different sources, which are combined with our microdata using the census tract identifiers. One consists in aggregate variables defined at the census tract level, which we use as additional controls in our empirical analysis, to rule out the effect of local confounders that might correlate with the treatment variable of interest. Specifically, the aggregate information we exploit in this paper has been constructed from the Spanish censuses of 1991 and 2001 and the so-called "Population Statistics" of 1996, a census-based database that is available only for the region of Catalonia. We retrieved: (1) the fraction of individuals aged 16 years or over who had completed some post-compulsory education, (2) the fraction of individuals born in other Spanish regions, and (3) the fraction of employed individuals in the active population in the census tract. We also obtained information about the fraction of adults who were able to: (4) at least speak or (5) speak and write in Catalan, but we use these variables only for the heterogeneity analysis (they are excessively correlated with the share of individuals born in the rest of Spain). We generally prefer using the 1991 values of these three variables, since they are less likely to be directly related to the presence of Catalan courses or to any other aggregate endogenous change that might have taken place during the period of existence of the language consortium. However, we also show the results obtained using controls from 1996 and 2001 for robustness. It is worth commenting that, in order to impute aggregate variables from 1991 and 1996 to the Microcensus, we tracked changes in census tracts with respect to the 2001 definition. In the large majority of cases, the census tract's boundaries coincide over these three years (at least within the Metropolitan Area of Barcelona), but on other occasions we found either divisions or combinations of census tracts over the years (243 mergers and 74 divisions in 1991, 94 mergers and 28 divisions in 1996). In these cases, we assumed a uniform distribution of the population within each census tract, which seems a reasonable and practical solution given the absence of information about the exact location of residential dwellings inside each area. Finally, we also consider the area (in square metres) of each census tract, the distance to the nearest school, and the distance from each census tract's centroid to Catalunya Square in Barcelona (which represents the "city centre").[10]

The second and most important aggregate data we exploit in this paper consists in historical information about language learning centres, provided by the Department of Language Policy of the Catalan Government. We obtained information about the exact geographical location (address) of each language school for each academic year since 1989/90, which made it possible to gauge the geographical coverage of language centres over the territory and its expansion over time. As we also knew the name of the centres in which the language courses took place, we were also able to classify them according to the public institution that was hosting language courses in a given year (i.e. schools, community centres, or other municipal public buildings). After cleaning the data and tracking changes in street names that occurred over time, we combined this information with the shape file at the census tract level (2001 definition) using GIS techniques. Fig. 1 displays the geographical location of language centres in the Metropolitan Area of Barcelona, considering the information from the first year of existence of the CPLN (we show the same map for 1994/95 and 2000/01 in Figures A1 and A2 of the online Appendix).

The original individual-level dataset from selected municipalities of the Metropolitan Area of Barcelona contains 102,141 observations of individuals aged between 16 and 64 years. For the purposes of our estimations, we dropped observations for individuals who were still in education (16,837 observations) and those born after 1971 (24,356). The latter restriction is motivated by the fact that these individuals automatically achieved the certification of full proficiency in Catalan (level C1) at the end of compulsory education, as explained in Section 2. After imposing these restrictions, we also eliminated from our estimation sample foreign migrants (8096) in order to avoid introducing additional (and unnecessary) unobserved heterogeneity in our model. Moreover, in order to avoid ambiguity in the definition of exposure to language courses, we excluded individuals who moved to their current residential dwelling after 1989 and who were residing in other provinces of Catalonia (1228). We obtained a final estimation sample of 51,624 observations that satisfy the above criteria. Table 1 reports descriptive statistics of individual and aggregate covariates for the selected sample.

As can be appreciated, around 63% of the sample declared that they were able to at least speak and read in Catalan and 37% that they were fully proficient (speaking, reading, and writing). However, another substantial part of the individuals in our data had limited skills in Catalan, especially considering the general tendency to overreport language proficiency. Moreover, the share of individuals born outside Catalonia is significantly high (44%), which is in line with the sociodemographic situation of this Spanish region described in Section 3 (see Table A1 in the online Appendix for more details about the province of birth). Indeed, the fraction of fully proficient individuals is substantially lower among them (only 14%, versus 54% of individuals born in Catalonia who are able to read, speak, and write in Catalan). We also report descriptive statistics for labour market outcomes, which are obtained considering only the available observations.[11]

[10] We decided to control for census tract area instead of population density because the census tracts, whose boundaries are defined for electoral purposes, generally contain the same size of population (between 500 and 2500 individuals with voting rights). Most importantly, the shape file from 1991 is not available and the conversion of the 1991 to 2001 census tracts was done by the Catalan Statistical Institute's technicians, using the original cartography. Therefore, as long as the information about the 1991 census tracts' surfaces is unavailable, we are unable to compute population density for this year. Moreover, we also consider the location of "incumbent" schools, constructed before 1990, in order to take into account the potential effect of school availability, which might have affected the individuals in our estimation sample and would not have been picked up by the fraction of highly educated individuals at the census tract level.

[11] That is, the original variable capturing the employment status excludes individuals involved in housework, and other labour market outcomes refer only to regularly employed individuals.

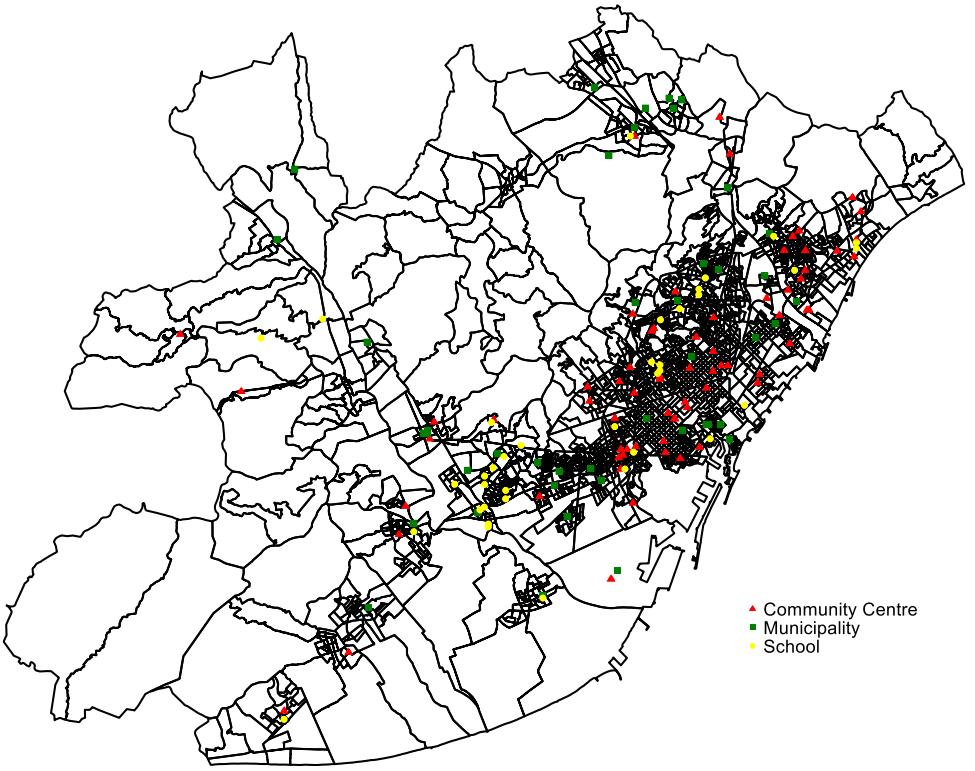


**Fig. 1.** Location of language centres by hosting institution, 1989/90.

Specifically, the Microcensus contains information about the employment status of the respondent and, for those who are employed, the hours of work, and occupation and industry (both at two digits). Here we specifically consider, besides employment and working hours, an indicator for having a highly skilled job and another one for working in the public sector, which have been derived from the information about the employment sector.

## 4. Empirical methodology

This section describes the approach that we follow to analyse the effect of the spatial deployment of language learning centres of the CPLN across the Metropolitan Area of Barcelona on individuals' proficiency in Catalan and labour market outcomes. We specifically want to focus on the initial phase of the CPLN, when the location of language learning centres across the territory was generally driven by the availability of public infrastructures rather than by the concentration of potential students (e.g. internal and foreign migrants), as has happened in more recent years. For this reason, we rely on microdata from the 2001 Microcensus (and not on more recent data such as the 2011 Microcensus), plus aggregate census-tract-level variables from different years (2001, 1996, and 1991), which we use as local controls. We generally rely on reduced-form equations as follows:

$$Y_{ic} = \alpha + \beta Exp^{r}_{ic} + \gamma' X_i + \delta' Z_c + \mu_{yri} + \varepsilon_{ic} \quad (1)$$

Here, $Y_{ic}$ represents the outcome under investigation of individual $i$ residing in census tract $c$. We initially focus on local language skills as the main outcome, considering the original ordinal scale from 1 (no knowledge) to 6 (speak, read and write), as well as a dummy that takes the value 1 if the individual $i$ is fully proficient in Catalan. The vectors $X_i$ and $Z_c$ are, respectively, individual[12] and local characteristics. The term $\mu_{yri}$ represents fixed effects for the years of residence in the current dwelling from the creation of the CPLN in 1989 until the census year, 2001. $Exp^{r}_{ic}$ is our treatment variable of interest, capturing potential exposure to language learning centres. The definition of this last variable is crucial. Our goal is to construct a measure of geographical exposure to language learning opportunities that can be reasonably taken as an Intention-to-Treat (ITT) but only using the available information about (1) the census tract of residence and (2) the

[12] The vector of individual controls includes gender, a third-order age polynomial, province of birth fixed effects, a second-order polynomial of age at migration to Catalonia (equal to zero for those born in Catalonia), and years of schooling. We also tried to include possible "bad controls" to capture household composition (number of children and adults in the household and marital status).

**Table 1**
Descriptive statistics (individual and aggregate variables).

| Variable | Obs. | Mean | Std. Dev. |
|---|---|---|---|
| ***Skills in Catalan*** | | | |
| Not understand Catalan | 51,624 | 0.042 | 0.201 |
| Understand | 51,624 | 0.217 | 0.412 |
| Read | 51,624 | 0.139 | 0.346 |
| Speak | 51,624 | 0.026 | 0.159 |
| Read and speak | 51,624 | 0.236 | 0.425 |
| Read, speak and write | 51,624 | 0.341 | 0.474 |
| ***Labour market outcomes*** | | | |
| Employed (housework = missing) | 34,931 | 0.898 | 0.303 |
| Employed extended (housework = 0) | 45,483 | 0.689 | 0.463 |
| Hours of work (per week) | 31,359 | 39.38 | 9.240 |
| Partime | 31,359 | 0.083 | 0.276 |
| Overtime | 31,359 | 0.175 | 0.380 |
| Public sector job (based on industry) | 31,359 | 0.188 | 0.390 |
| High skilled job | 31,338 | 0.376 | 0.484 |
| ***Individual characteristics*** | | | |
| Male | 51,624 | 0.483 | 0.500 |
| Age | 51,624 | 47.65 | 10.166 |
| Born in Catalonia | 51,624 | 0.542 | 0.498 |
| Years of education (imputed from levels) | 51,624 | 8.926 | 4.23 |
| Age of arrival to Catalonia (0 for natives) | 51,624 | 7.490 | 11.038 |
| Adults in the household | 51,624 | 2.843 | 1.146 |
| Children in the household | 51,624 | 0.451 | 0.762 |
| Single | 51,624 | 0.173 | 0.378 |
| Married | 51,624 | 0.722 | 0.448 |
| Widowed | 51,624 | 0.037 | 0.190 |
| Divorced | 51,624 | 0.039 | 0.193 |
| Separated | 51,624 | 0.029 | 0.168 |
| ***Local characteristics*** | | | |
| Area ($m^2$) | 51,624 | 296327 | 1419234 |
| Distance to the nearest public school (km) | 51,624 | 0.202 | 0.259 |
| Distance to Catalunya Square (km) | 51,624 | 6.413 | 4.278 |
| Share of employed 1991 | 51,624 | 0.856 | 0.046 |
| Share of highly educated ind. 1991 | 51,624 | 0.238 | 0.150 |
| Share of individuals born in Spain 1991 | 51,624 | 0.426 | 0.145 |
| Share of employed 1996 | 51,624 | 0.780 | 0.062 |
| Share of highly educated ind. 1996 | 51,624 | 0.308 | 0.163 |
| Share of individuals born in Spain 1996 | 51,624 | 0.386 | 0.120 |
| Share of employed 2001 | 51,624 | 0.886 | 0.035 |
| Share of highly educated ind. 2001 | 51,624 | 0.348 | 0.160 |
| Share of individuals born in Spain 2001 | 51,624 | 0.346 | 0.111 |
| ***Exposure measures*** | | | |
| Language centres within 0.5 km | 51,624 | 0.978 | 0.866 |
| Language centres within 1 km | 51,624 | 3.347 | 2.097 |
| Language centres within 1.5 km | 51,624 | 6.770 | 3.958 |

Note: years of schooling = 0 for illiterate individuals, 3 for incomplete primary education, 6 for completed primary education, 8 for compulsory secondary education, 10 for post-compulsory general secondary education, 12 for vocational training, 14 for advances vocational training, 15 for short-term university degree, 17 for university degree and 20 for PhD. Public sector job is equal to one if the individual works in the following sectors: public administration, defence or social security; education; health, veterinary and social services.

year of arrival in the current dwelling, plus (3) the location of language centres. Our focus on ITT effects is also driven by the fact that we do not have information on participation in language courses, which forces us to rely on reduced-form estimates that directly relate accessibility of language courses and the outcomes of interest.

Based on the available information, we define $Exp^r_{ic}$ in a way that captures the availability of language learning centres in the neighbourhood of residence over the time span during which each individual could have actually been exposed to them. We construct a measure of accessibility based on the number of locations surrounding the place of residence, based on the hypothesis that the higher presence of language learning centres in the neighbourhood would positively affect the propensity to enrol in a language course. As mentioned in Section 3, our spatial unit of references are the census tracts of the Metropolitan Area of Barcelona, which is the most disaggregated geographical information about individual's place of residence that is available in the 2001 Microcensus. Therefore, using GIS techniques, on the one hand we retrieved the census tracts' centroids and, on the other hand, we geolocated the exact address of all language centres for the relevant academic years (1989/90 to 2000/01), as shown in Fig. 1. With the aim of capturing local accessibility of language centres, we computed the sum of language centres located within different buffers around each census tract'

centroids. Specifically, we consider three different options to compute the number of available language centres, that is, the total number of centres located in buffers of radii (i) 0.5 km, (ii) 1 km and (iii) 1.5 km from the centroid of each census tracts. As explained below, this enables testing not only for the robustness of the results to different definitions of the accessibility measure, but also for the presence of spatial decay effects.[13] Moreover, to consider the fact that each individual could have been exposed during different years since 1989, we sum all the language centres located within each radius over the academic years 1989/90 to 2000/01, to capture potential exposure throughout the period of existence of the CPLN.[14] Finally, since not all the individuals in our sample resided in the same place between 1989 and 2000, because some might have moved to their current dwelling during this period, we exploit the information about the year of migration to the place of residence to sum up the number of "nearby" language centres only over the relevant years. This actually introduces some, though limited, degree of individual-level variation in our measure of exposure $Exp^r_{ic}$, which also justifies the inclusion of fixed effects for the years of residence since 1989 in the equation to be estimated.[15] Because the main variable of interest varies at the census tract level, we cluster the standard errors accordingly (using two-way clusters by census tract and years of residence provides similar results).

It seems worth mentioning that we focus on the number of language training centres surrounding the place of residence rather than on the minimum distance: first, our data contains information about the census tract of residence, but not the exact location of the dwelling, which means that we are only able to retrieve the centroid of each census tract. Because census tracts' area is significantly larger in less populated areas, working with the minimum distance from the centroid would introduce substantial measurement error for individuals residing in specific neighbourhoods of the metropolitan area. The use of the number of language training centres in surrounding areas is likely to minimize the issues due to the unavailability of information about the precise place of residence of each individual. This is because some individuals could live close to the census tract's borders and relatively far away from its centroid, which would induce some measurement error in the computation of our accessibility measure. However, the use of different radii to compute the exposure variable and the inclusion of local controls (as detailed below) are likely to mitigate the issue, which in any case would bias the OLS estimates towards zero, meaning that the estimated coefficient could represent a lower bound of the true effect of exposure.

### 4.1. Threats to identification and corresponding checks

The extent to which the estimate of the $\beta$ parameter from Eq. (1) can be plausibly interpreted as an ITT effect depends on whether the exposure variable we adopt in this paper can be considered "as good as randomly assigned". However, at least two general issues could undermine the exogeneity of $Exp^r_{ic}$, representing potential threats to the identification of the causal effect of interest. First, the variable we use to capture local exposure to language centres could pick up the effect of local factors that might be related to the spatial distribution of language centres as well as to the outcomes of interest. Second, residential decisions of individuals and families might not be completely exogenous with respect to the location of language centres or to other (directly or indirectly related) local characteristics. Given the absence of randomization or any exogenous shock in the availability of language learning centres, we carry out several checks to provide evidence that corroborates the causal interpretation of our findings. As mentioned in Section 3, an interesting feature of the initial implementation of the CPLN is that the placement of local language courses by public authorities that participated in the Consortium was mostly driven by the availability of public infrastructures. This somehow limits the degree of unobserved heterogeneity but does not exclude the possible existence of local confounders, which we try to rule out in different ways.

Therefore, we first proceed with the inclusion of a vector of local controls $Z_c$, containing the area of each census tract and the distance to the centre of Barcelona (Plaça Catalunya), to rule out the effect of residing in more isolated areas in the periphery of the metropolitan area, and the distance to the nearest public (or charter) school. This last control is especially important, because schools host a large fraction of language courses and, although we only consider adult individuals not in education, school availability in the neighbourhood of residence could correlate with their outcomes through a variety of channels.[16] We also control for other socioeconomic and demographic features of the census tracts that might be correlated with the location of language centres, specifically the local employment share (of the active labour force), the share of individuals over 15 with post-compulsory education, and the share of adult individuals born in other Spanish regions. Controlling for these aggregate variables is relevant, since they are likely to capture a substantial part of local unobservables from Eq. (1) but, to some extent, they could have been affected by the location of language centres. The

[13] We also checked the results obtained with smaller radii such as 0.25 and 0.1 km, but the estimates were quite imprecise and meaningless, due to the very limited amount of variation in the corresponding exposure variables (i.e. there are too few census tracts with at least one language centre within such small radii).

[14] We rescale the sum by 12 to facilitate the interpretation of the corresponding coefficient. We also tried separate use of the number of nearby centres in each year, which provided generally similar but less precise estimates (results are available upon request).

[15] The other side of the coin is that, as explained in Section 4, we are forced to exclude the relatively few individuals who (a) moved to the current dwelling between 1989 and 2001 and (b) came from other locations in Catalonia, since they could have been exposed to Catalan elsewhere. Therefore, we only retain individuals who migrated between 1989 and 2001 and who came from other Spanish regions.

[16] Moreover, the presence of schools could endogenously attract certain types of families who decide where to live based on their children's education opportunities (among other things), leading to endogenous residential sorting. We also carry out additional robustness checks to discard this potential issue.

ideal setting would use pre-determined values (i.e. before the creation of the CPLN), but such information is not available at the census tract level for periods before 1989. Therefore, we generally exploit the first available information, namely from the 1991 Census, but we also try to control for the value of these three local characteristics observed in 1996 and 2001 (separately and all together), to check for the stability of the coefficient of interest after partialling out the effect of neighbourhood composition. In addition, we saturate our model with district[17] fixed effects ($\pi_d$), which makes it possible to partial out time-invariant unobserved characteristics that vary at a slightly more aggregated level than our variable of interest (and thus exploit within-level variation in exposure), that is:

$$Y_{ic} = \alpha + \beta Exp^{r}_{ic} + \gamma' X_i + \delta' Z_c + \mu_{yri} + \pi_d + \varepsilon_{ic} \quad (2)$$

Second, using our preferred specification, which includes district fixed effects [i.e. Eq. (2) including local controls from 1991], we also carry out two different falsification exercises. On the one hand, we re-estimate the same equation using an additional sample of "young" individuals, born between 1972 and 1985 (who satisfy the other conditions imposed for the main estimation sample). This is because they should not be directly affected by the geographical availability of language courses since, as explained above, they were exposed to Catalan at school and, most importantly, they automatically received the proficiency certification required to enter (local and regional) public sector jobs. Therefore, they should not have any direct incentive to enrol on Catalan courses, except for the very unusual cases of intrinsically motivated individuals who want to achieve the higher qualification (level D) that is required to be a language teacher. Moreover, we also perform a permutation test in which we randomly assign census tracts (and the corresponding exposure measure) to individuals and re-estimate Eq. (2) using the placebo exposure measures. We consider two different variants of the same experiment: (1) using a uniform distribution of observation for each census tract and (2) maintaining the original proportion of observations within each census tract, and we run 10000 simulations, to obtain the corresponding distribution(s) of the fake $\beta$ coefficients.

Third, we estimate Eq. (2) considering alternative definitions of $Exp^{r}_{ic}$, based on different radii from the census tracts' centroids to calculate the number of surrounding language learning centres. Specifically, we start using a radius of 0.5 km ($Exp^{0.5}_{ic}$), which ends up being the preferred specification of the exposure variable, and we progressively increase the buffer radius to 1 km ($Exp^{1}_{ic}$) and 1.5 km ($Exp^{1.5}_{ic}$), which also allows checking for distance-decay effects of exposure to language learning opportunities. Moreover, we present the results from a variant of Eq. (2) that includes the number of language centres (summed over time) located between 0 and 0.5 km ($Exp^{0.5}_{ic}$), between 0.51 and 1 km $Exp^{0.5-1}_{ic}$, and between 1.1 and 1.5 km ($Exp^{1-1.5}_{ic}$), which corresponds to the following equation:

$$Y_{ic} = \alpha + \beta_1 Exp^{0.5}_{ic} + \beta_2 Exp^{0.5-1}_{ic} + \beta_3 Exp^{1-1.5}_{ic} + \gamma' X_i + \delta' Z_c + \mu_{yri} + \pi_d + \varepsilon_{ic} \quad (3)$$

Fourth, we also deal with other potential confounding effects, which could stem from other services offered at the same institutions hosting language courses. As explained in Section 2, Catalan courses of the CPLN took place in different types of public infrastructures, which can be generally classified as public schools, community centres, and other municipal infrastructures. The issue comes from the fact that the last two types of hosting institutions were also likely to offer other courses, workshops, and training aimed at fostering different types of skills and employability of residents (especially unemployed and disadvantaged individuals). Moreover, other social activities that take place in community centres might also have a direct effect on the propensity to learn and know Catalan, since they could foster community ties and social interactions in the neighbourhood (Domínguez and Montolio, 2021). Finally, school buildings were not used for language courses during the morning but only during the afternoon, because they were busy with regular educational activities during the morning. Therefore, there could be some difference in the composition of language students by type of hosting institution (for example, by employment status). This means that accessibility of language centres could be capturing not only the impacts of shaping language learning opportunities but also the direct and indirect effects induced by other activities offered by community centres and by the municipalities that participate in the CPLN. For this reason, using the preferred specification from the previous check, we split the exposure variable according to the type of public infrastructure where language courses were held. Specifically, we separately estimate the effect of exposure to language courses held in public schools ($Exp^{sch}_{ic}$), community centres ($Exp^{cc}_{ic}$), and other municipality buildings ($Exp^{mun}_{ic}$), including these variables in the model one by one and all together, as in Eq. (4):

$$Y_{ic} = \alpha + \beta_1 Exp^{sch}_{ic} + \beta_2 Exp^{cc}_{ic} + \beta_3 Exp^{mun}_{ic} + \gamma' X_i + \delta' Z_c + \mu_{yri} + \pi_d + \varepsilon_{ic} \quad (4)$$

Given that public schools do not provide additional services targeted to adult individuals besides Catalan courses (and considering that we retain only adults not involved in education[18]), a finding that the total effect of exposure is mostly driven by the number of language centres located in the two other types of institutions would be against the causal interpretation of our main estimates.

Fifth, we specifically deal with the issue of endogenous residential sorting. As we mentioned before, the location of language centres in the initial phase of the CPLN was generally driven by the availability of public infrastructures (and

[17] Notice that this corresponds to a municipality fixed effect when it has a single district.

[18] Notice that, as we also control for the distance to the nearest school (including adult's school), this mitigates the possibility of capturing the effect of some sorts of adult education programmes that take place in the schools hosting language courses.

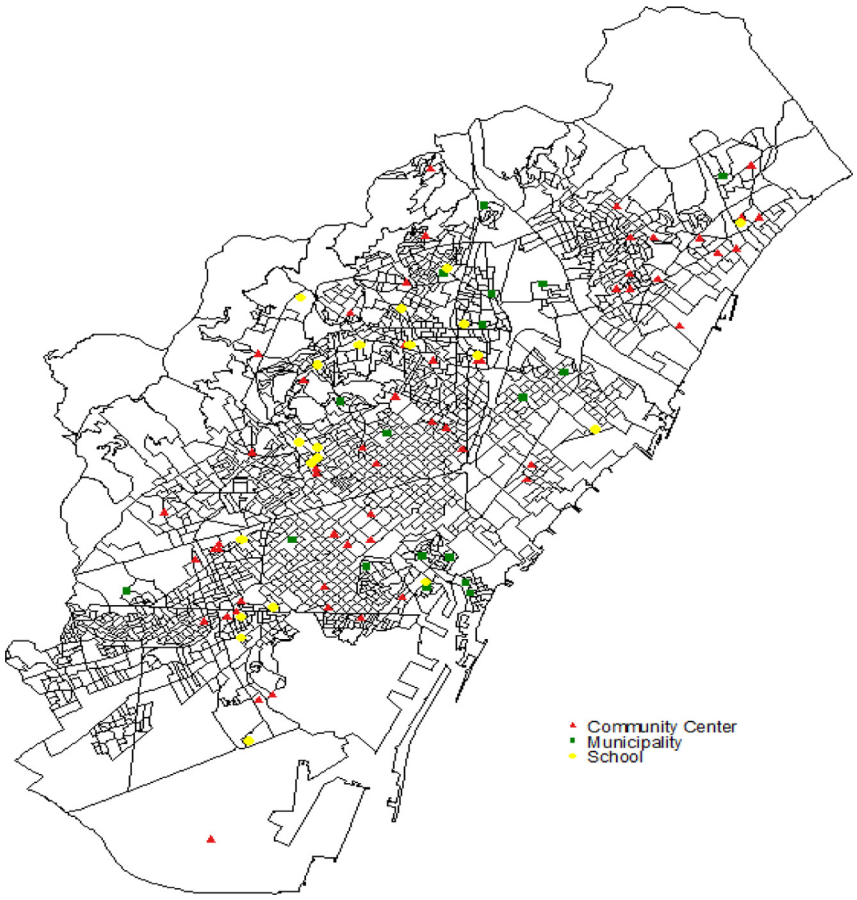


**Fig. 2.** Location of language centres by hosting institution, core of the metropolitan area (Barcelona and large and dense surrounding municipalities), 1989/1990.

the agreement between the corresponding institutions and the CPLN). However, this does not prevent the possibility that residential decisions of individuals and families are in part based on observed and unobserved characteristics of the neighbourhood. Although we already control for the distance to the nearest school and for several other local characteristics, besides the fact that the likelihood that families decide where to live on the basis of the availability of Catalan learning centres is actually very limited, we still consider that endogenous residential sorting could be somehow biasing the estimate of the parameter of interest. Therefore, we proceed by re-estimating Eq. (2), our preferred specification (using a radius of 0.5 km) after excluding individuals who moved to the current dwelling after the creation of the CPLN in 1989 (as well as those who moved after 1985 for completeness). Moreover, due to the ambiguity in the definition of exposure for individuals who arrived in the census year (2001) and considering that the interviews were done in November, we try to drop individuals who arrived at their place of residence during the same year.

Finally, our last robustness check consists in re-estimating the main equation only considering observations from "dense areas", which are the city of Barcelona and large surrounding municipalities of the metropolitan area (see Fig. 2). This check is important not only because of the unevenness of census tracts in less dense areas of the Metropolitan Area of Barcelona in terms of surface and population size (which in turn would affect the chances of social interaction), but also because there could be spatial spillovers from municipalities outside the metropolitan area. That is, it is possible that some individuals residing in the periphery of the metropolitan area could also be affected by the presence of language centres in other municipalities located outside of its limits.

After presenting the evidence from all these sensitivity checks, which favour the interpretation of our exposure variable as "conditionally exogenous", we provide additional results regarding the presence of heterogeneous effects of exposure. Specifically, we consider possible heterogeneous effects with respect to several individual characteristics such as gender, age, education, place of birth, and age at migration, as well as according to different local characteristics (mostly the share of internal migrants and the share of individuals who are proficient in Catalan).

Finally, we also analyse the relevance of skills in Catalan and the potential effect of exposure to language training courses on labour market outcomes. Specifically, exploiting the information in the Microcensus, we first show the

conditional correlation between the variable capturing skills in Catalan and the probability of being employed (considering also labour market participation of individuals involved in housework tasks), weekly hours of work, the probability of working part-time or overtime, as well as the probability of working in the public sector (inferred from industry) and the chances of having a highly skilled job (inferred from occupation codes) among the employed. Second, we estimate the reduced-form equation that directly relates the exposure measure to the above-mentioned labour market outcomes.

## 5. Results

Our first set of results are obtained by estimating Eq. (1) by OLS, whose estimates are reported in columns (1) to (6) of Table 1. We start by including only a parsimonious set of individual controls, namely gender, age (third-order polynomial), years of education, and age at migration (second-order polynomial, equal to 0 for natives), plus fixed effects for the province of birth and years of stay in the current dwelling (since 1989). The coefficient of interest ($\beta$), reflecting the impact of potential exposure to local language schools, indicates that each additional centre (per year) in a radius of 0.5 m from the centroid of the census tract of residence increases skills in Catalan by 0.046 points on the 1 to 6 ordinal scale of the dependent variable (mean 4.2, s.d. 1.71). This corresponds to approximatively 2.7% of the standard deviation of the language skills variable. In column (2) we include a first set of census tract characteristics in the vector $Z_c$ (area, distance to the most central square of Barcelona, and distance to the closest public school), which produces a certain reduction in the estimate of interest. This is possibly due to the fact that in the previous specification we were picking up some local confounders associated with the location of different census tracts within the Metropolitan Area of Barcelona. The coefficient of potential exposure obtained from this specification is 0.026, around 1.5% of the standard deviation of the outcome. In columns (3) to (6), we augment the vector of local controls by adding time-varying characteristics of the census tract, namely the share of employed individuals, the fraction of adult individuals born in other Spanish regions, and the percentage of individuals aged above 15 years with post-compulsory education. As mentioned above, we use data from 1991, 1996, and 2001, separately and jointly [column (6)].

The inclusion of these local controls slightly increases the estimate corresponding to exposure to local language courses, especially when using more recent census tract level variables, but the main conclusion remains the same (i.e. a modest but positive and significant effect of geographical exposure to language centres). However, the same local characteristics could themselves have been affected, directly or indirectly, by the deployment of Catalan courses after the creation of the CPLN and could thus represent bad controls. The ideal setting would have used "pre-determined variables", observed before 1989, but such information is unavailable. However, we are much more confident about the exogeneity of local controls measures in 1991, just two years after the policy implementation, and therefore we retain this set of variables to control for neighbourhood composition. Subsequently, in column (7) we show the results obtained including district fixed effects, which represents our preferred specification from both a conceptual and a statistical point of view.[19] The coefficient of exposure obtained from this specification is 0.030 (s.e. 0.010), which is equivalent to 1.7% of the variation in self-reported language skills.

In addition to the main evidence, Table 2 also shows other interesting effects that are worth commenting on. At the individual level, males are less likely to be proficient than females and schooling is positively associated with language proficiency, as expected. This relationship does not capture exposure to Catalan at school, since the sample only includes individuals who received compulsory schooling in Spanish only. Moreover, still at the individual level, the age of arrival in Catalonia has a U-shaped and significant effect: the older the individual was, the lower the propensity to learn Catalan, although the marginal effect is decreasing. As for census tract level variables, the share of employed individuals is positively correlated with language skills, while the fraction of individuals born in other Spanish regions has the expected negative sign, reflecting the effect of living in Spanish-speaking enclaves of the Metropolitan Area of Barcelona. The coefficient of the variable capturing the share of highly educated individuals is unexpectedly negative, but it turns out to be positive without controlling for the composition of the neighbourhood in terms of province of birth.

In Table 3, we complement the previous analysis by displaying the average marginal effects obtained from an Ordered Probit, using the preferred specification [as in column (7) of Table 2], which facilitates the quantitative interpretation of the results. The geographical accessibility to the CPLN courses has an effect that is quantitatively modest, but strongly robust, on the probability of being able to speak and write in Catalan. An increment of one school per year within a radius of 0.5 km from the place of residence increases the probability of being able to speak and write in Catalan (i.e. being fully proficient) by 0.8 percentage point (with the unconditional probability being equal to 0.34). Similar evidence is obtained by estimating a Linear Probability Model, in which the dependent variable is the dummy for being able to speak and write in Catalan (see Table A2 of the online Appendix), although the point estimate is slightly higher (1.1 percentage point). The evidence is also robust to the use of a Probit model for the same outcome, whose average marginal effects for the variable of interest are identical to those obtained from the Ordered Probit model (see Table A3 of the online Appendix).

In order to check that our results are not driven by other unobserved factors, we present the results obtained from two different falsification exercises. First, we re-estimate our model following the same specifications displayed in Table 2 but

[19] In column (8), we also show the results obtained after including the number of children and adults in the household and marital status as additional controls, which provides virtually the same results. However, we decided to retain the results displayed in column (7) since these additional variables could represent bad control, because they might be directly or indirectly related to the presence of language schools and skills in Catalan.

**Table 2**
OLS, Dependent variable = self-reported knowledge of Catalan.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Constant | 11.066*** | 11.455*** | 9.148*** | 10.497*** | 9.611*** | 8.610*** | 10.174*** | 9.795*** |
| | (0.710) | (0.709) | (0.748) | (0.728) | (0.805) | (0.786) | (0.758) | (0.800) |
| Exposure 0.5 km | 0.046*** | 0.026** | 0.028*** | 0.030*** | 0.036*** | 0.027*** | 0.030*** | 0.030*** |
| | (0.010) | (0.010) | (0.009) | (0.010) | (0.010) | (0.009) | (0.010) | (0.010) |
| Male | −0.085*** | −0.079*** | −0.063*** | −0.062*** | −0.063*** | −0.062*** | −0.063*** | −0.056*** |
| | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) |
| Age | −0.505*** | −0.513*** | −0.512*** | −0.525*** | −0.518*** | −0.514*** | −0.514*** | −0.481*** |
| | (0.047) | (0.047) | (0.046) | (0.046) | (0.046) | (0.046) | (0.046) | (0.049) |
| Age2 | 0.011*** | 0.011*** | 0.010*** | 0.011*** | 0.011*** | 0.010*** | 0.011*** | 0.010*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Age3 | −0.000*** | −0.000*** | −0.000*** | −0.000*** | −0.000*** | −0.000*** | −0.000*** | −0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Years education | 0.134*** | 0.129*** | 0.114*** | 0.113*** | 0.113*** | 0.113*** | 0.114*** | 0.114*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Age of arrival to Catalonia | −0.063*** | −0.062*** | −0.060*** | −0.059*** | −0.059*** | −0.060*** | −0.059*** | −0.060*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Age of arrival to Catalonia 2 | 0.001*** | 0.001*** | 0.001*** | 0.001*** | 0.001*** | 0.001*** | 0.001*** | 0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Area | | 0.000*** | 0.000 | 0.000* | 0.000*** | 0.000 | 0.000 | 0.000 |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Distance to the nearest public school | | 0.084 | 0.052 | 0.078 | 0.012 | 0.055 | 0.024 | 0.019 |
| | | (0.068) | (0.051) | (0.048) | (0.054) | (0.047) | (0.049) | (0.049) |
| Distance to Catalunya Square | | −0.027*** | −0.007*** | −0.018*** | −0.012*** | −0.015*** | −0.041*** | −0.042*** |
| | | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.011) | (0.011) |
| Share of employed 1991 | | | 4.081*** | | | 2.170*** | 2.707*** | 2.651*** |
| | | | (0.313) | | | (0.308) | (0.350) | (0.347) |
| Share of highly educated ind. 1991 | | | −0.965*** | | | −0.950*** | −0.254* | −0.238* |
| | | | (0.108) | | | (0.211) | (0.140) | (0.140) |
| Share of individuals born in Spain 1991 | | | −1.733*** | | | −0.698*** | −1.095*** | −1.108*** |
| | | | (0.098) | | | (0.208) | (0.120) | (0.119) |
| Share of employed 1996 | | | | 3.305*** | | 1.654*** | | |
| | | | | (0.266) | | (0.273) | | |
| Share of highly educated ind. 1996 | | | | −1.046*** | | −0.598** | | |
| | | | | (0.110) | | (0.282) | | |
| Share of individuals born in Spain 1996 | | | | −2.013*** | | −1.070*** | | |
| | | | | (0.117) | | (0.414) | | |
| Share of employed 2001 | | | | | 3.317*** | 1.370*** | | |
| | | | | | (0.433) | (0.379) | | |
| Share of highly educated ind. 2001 | | | | | −0.217** | 0.241 | | |
| | | | | | (0.104) | (0.234) | | |
| Share of individuals born in Spain 2001 | | | | | −1.918*** | −0.339 | | |
| | | | | | (0.130) | (0.376) | | |
| Adults in the household | | | | | | | | −0.009 |
| | | | | | | | | (0.007) |
| Children in the household | | | | | | | | −0.049*** |
| | | | | | | | | (0.011) |
| Single | | | | | | | | ref. cat. |
| Married | | | | | | | | 0.113*** |
| | | | | | | | | (0.020) |
| Widowed | | | | | | | | 0.090** |
| | | | | | | | | (0.035) |
| Divorced | | | | | | | | 0.171*** |
| | | | | | | | | (0.034) |
| Separated | | | | | | | | 0.140*** |
| | | | | | | | | (0.038) |

(*continued on next page*)

**Table 2** (*continued*).

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Province of birth fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Years of residence fixed effects (1989–2001) | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Municipality-district fixed effects | No | No | No | No | No | No | Yes | Yes |
| Adjusted R-squared | 0.428 | 0.431 | 0.448 | 0.449 | 0.446 | 0.451 | 0.454 | 0.454 |
| Number of observations | 51624 | 51624 | 51624 | 51624 | 51624 | 51624 | 51624 | 51624 |

Standard errors clustered at the census tract level; * significant at 10%, ** significant at 5%, *** significant at 1%.

**Table 3**
Average marginal effects from ordered probit estimates.

| | $\Delta$Pr(Cat = 1) | $\Delta$Pr(Cat = 2) | $\Delta$Pr(Cat = 3) | $\Delta$Pr(Cat = 4) | $\Delta$Pr(Cat = 5) | $\Delta$Pr(Cat = 6) |
|---|---|---|---|---|---|---|
| Exposure 0.5 km | −0.002*** | −0.005*** | −0.001*** | −0.000*** | 0.000*** | 0.008*** |
| | (0.001) | (0.001) | (0.000) | (0.000) | (0.000) | (0.002) |

Dependent variable: self-reported knowledge of Catalan (1 = not understand Catalan, 2 = understand, 3 = read, 4 = speak, 5 = read and speak read, 6 = speak and write). Control variables as in column (7) of Table 2. Standard errors clustered at the census tract level; * significant at 10%, ** significant at 5%, *** significant at 1%.

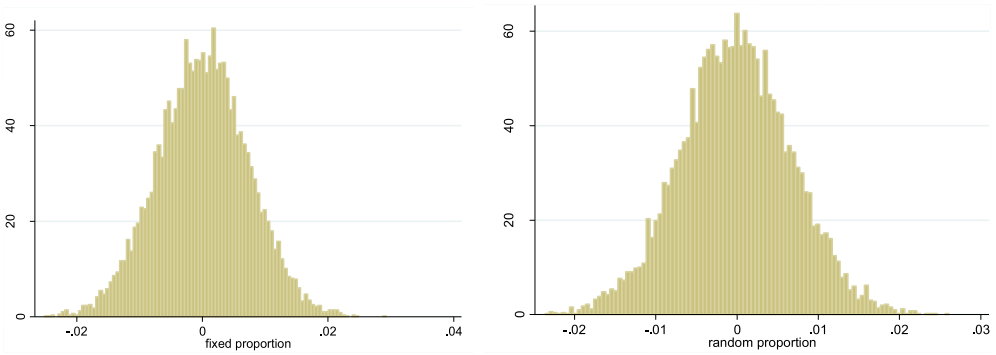


**Fig. 3.** Fake beta coefficients from the permutation test. Note: estimated beta coefficients obtained after assigning at random the census tract to each individual, 10,000 replications. Left graph: permutation carried out maintaining the same proportion of individuals in each census tract as observed in the main sample (fixed proportion). Right graph: permutation obtained without keeping the proportion of observations in each census tract observed in the main sample (random proportion).

using an additional sample of "young" individuals born after 1971 (and thus initially excluded from our estimation sample; see Table A4 in the online Appendix for descriptive statistics). As mentioned above, these individuals were exposed to Catalan as a medium of instruction at school during compulsory education and automatically received the proficiency certificate.

Therefore, they are not expected to be users of the language courses provided by the CPLN and, indeed, finding an effect of geographical exposure among them would indicate that our main estimates are likely to reflect some spurious correlation rather than a real ITT effect. Reassuringly, the results reported in Table 4 indicate that the point estimate of the placebo coefficient is generally very close to zero and not statistically significant in any specification, which speaks in favour of the validity of our approach.

Second, we randomly assigned each individual in the main sample to a census tract of the Metropolitan Area of Barcelona, with and without maintaining the proportion of observations in each census tract. We replicated this random assignment process 10,000 times and, for each permutation, we re-estimated our preferred specification [Eq. (2)] and plotted the resulting distribution of fake beta coefficients. As can be seen in Fig. 3, in both cases the placebo beta coefficients are distributed around zero and are very unlikely to be equal to the value estimated using the real exposure variable (0.030). This evidence again suggests that our estimates are not capturing unobserved local factors that correlate with language skills, but are rather capturing the real potential effect of being exposed to language learning opportunities.

Furthermore, to have a better understanding of the effect of geographical exposure to language courses, we used different radii in the definition of the variable in our preferred regression, which also makes it possible to gauge the presence of distance-decay effects. The results are reported in Table 5. The first three columns show the $\beta$ coefficient with radii of 0.5 km (the baseline), 1 km, and 1.5 km, respectively, again estimated using our preferred specification. We can see that as we enlarge the radius, the point estimate gets smaller and loses significance. To disentangle the effect of the geographic proximity, we included three variables that include information on the number of centres (by year) in each ring: closer to 0.5 km, between 0.5 km (not included) and 1 km, and between 1 km (not included) and 1.5 km. The results are shown in the last column of Table 5. The results point to a distance decay in the effect of the exposure, in both magnitude and significance. Actually, the number of centres in the outer ring seems to have no effect and the null

**Table 4**
Falsification analysis using young individuals (i.e. born after 1971).

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Constant | 8.479*** | 8.645*** | 5.523*** | 6.585*** | 5.667*** | 4.773*** | 6.823*** | 6.721*** |
| | (1.564) | (1.564) | (1.556) | (1.559) | (1.590) | (1.589) | (1.546) | (1.543) |
| Exposure 0.5 km | 0.012 | −0.008 | −0.003 | −0.001 | 0.003 | −0.004 | −0.003 | −0.003 |
| | (0.011) | (0.011) | (0.011) | (0.011) | (0.011) | (0.010) | (0.012) | (0.012) |
| Male | −0.080*** | −0.081*** | −0.086*** | −0.086*** | −0.086*** | −0.087*** | −0.086*** | −0.096*** |
| | (0.013) | (0.013) | (0.013) | (0.013) | (0.013) | (0.013) | (0.013) | (0.013) |
| Age | −0.396* | −0.406* | −0.397* | −0.398* | −0.407* | −0.394* | −0.406* | −0.350* |
| | (0.215) | (0.215) | (0.210) | (0.211) | (0.212) | (0.210) | (0.210) | (0.209) |
| Age2 | 0.013 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.011 |
| | (0.010) | (0.010) | (0.009) | (0.009) | (0.010) | (0.009) | (0.009) | (0.009) |
| Age3 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 | −0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Years education | 0.111*** | 0.108*** | 0.099*** | 0.099*** | 0.099*** | 0.098*** | 0.098*** | 0.095*** |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) |
| Age of arrival to Catalonia | −0.039** | −0.038** | −0.038** | −0.039** | −0.039** | −0.039** | −0.038** | −0.036** |
| | (0.018) | (0.018) | (0.018) | (0.018) | (0.018) | (0.018) | (0.018) | (0.018) |
| Age of arrival to Catalonia 2 | −0.002*** | −0.002*** | −0.002*** | −0.002*** | −0.002*** | −0.002*** | −0.002*** | −0.003*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Area | | 0.000 | −0.000 | −0.000 | 0.000 | −0.000 | −0.000 | −0.000 |
| | | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Distance to the nearest public school | | −0.024 | −0.015 | 0.013 | −0.045 | −0.012 | −0.007 | −0.008 |
| | | (0.063) | (0.053) | (0.054) | (0.058) | (0.052) | (0.059) | (0.058) |
| Distance to Catalunya Square | | −0.014*** | −0.005* | −0.012*** | −0.008*** | −0.009*** | −0.039*** | −0.038*** |
| | | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.012) | (0.012) |
| Share of employed 1991 | | | 3.907*** | | | 2.297*** | 2.542*** | 2.427*** |
| | | | (0.397) | | | (0.403) | (0.425) | (0.412) |
| Share of highly educated ind. 1991 | | | −0.580*** | | | −0.913*** | −0.125 | −0.106 |
| | | | (0.111) | | | (0.327) | (0.143) | (0.143) |
| Share of individuals born in Spain 1991 | | | −0.486*** | | | −0.479 | −0.338*** | −0.340*** |
| | | | (0.098) | | | (0.292) | (0.127) | (0.125) |
| Share of employed 1996 | | | | 2.838*** | | 0.956*** | | |
| | | | | (0.329) | | (0.354) | | |
| Share of highly educated ind. 1996 | | | | −0.432*** | | −0.181 | | |
| | | | | (0.124) | | (0.411) | | |
| Share of individuals born in Spain 1996 | | | | −0.311** | | −0.490 | | |
| | | | | (0.123) | | (0.533) | | |
| Share of employed 2001 | | | | | 3.301*** | 1.546*** | | |
| | | | | | (0.515) | (0.443) | | |
| Share of highly educated ind. 2001 | | | | | 0.165 | 0.383 | | |
| | | | | | (0.107) | (0.290) | | |
| Share of individuals born in Spain 2001 | | | | | −0.190 | 0.563 | | |
| | | | | | (0.132) | (0.447) | | |
| Adults in the household | | | | | | | | −0.015* |
| | | | | | | | | (0.009) |
| Children in the household | | | | | | | | −0.109*** |
| | | | | | | | | (0.024) |
| Single | | | | | | | | ref. cat. |
| Married | | | | | | | | −0.075** |
| | | | | | | | | (0.038) |
| Widowed | | | | | | | | −0.935 |
| | | | | | | | | (0.581) |
| Divorced | | | | | | | | −0.148 |
| | | | | | | | | (0.128) |
| Separated | | | | | | | | −0.195 |
| | | | | | | | | (0.203) |

(*continued on next page*)

**Table 4** (*continued*).

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Province of birth fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Years of residence fixed effects (1989–2001) | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Municipality-district fixed effects | No | No | No | No | No | No | Yes | Yes |
| Adjusted R-squared | 0.181 | 0.183 | 0.201 | 0.199 | 0.197 | 0.204 | 0.209 | 0.212 |
| Number of observations | 24334 | 24334 | 24334 | 24334 | 24334 | 24334 | 24332 | 24332 |

OLS, Dependent Variable = self-reported knowledge of Catalan. Standard errors clustered at the census tract level; * significant at 10%, ** significant at 5%, *** significant at 1%.

**Table 5**
Exposure at different radii.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Exposure 0.5 km | 0.030*** | | | 0.036*** |
| | (0.010) | | | (0.011) |
| Exposure 1 km | | 0.021*** | | |
| | | (0.006) | | |
| Exposure 1.5 km | | | 0.008** | |
| | | | (0.004) | |
| Exposure 0.51–1 km | | | | 0.017*** |
| | | | | (0.006) |
| Exposure 1.1–1.5 km | | | | −0.000 |
| | | | | (0.005) |
| Adjusted R-squared | 0.454 | 0.454 | 0.453 | 0.454 |
| Number of observations | 51624 | 51624 | 51624 | 51624 |

OLS, dependent variable: self-reported knowledge of Catalan. Control variables as in column (7) of Table 2. Standard errors clustered at the census tract level; * significant at 10%, ** significant at 5%, *** significant at 1%.

hypothesis of equality of the exposure coefficients in column (4) is soundly rejected. This is clear evidence that proximity matters and indeed the number of centres located within 0.5 km from the centroid of the census tract is what matters more.[20]

Additionally, to understand whether the estimated $\beta$ coefficient is picking up direct or indirect effects of other activities that took place in the same centre in which the language courses were offered, as well as differences in student composition, we split the overall exposure variable by type of building (see Table 6). The results reported in Table 6 indicate that the point estimate of exposure to language courses undertaken in public school buildings is higher (and more statistically significant) than those of exposure to language centres located in community centres or other municipal infrastructures. This evidence goes against the idea that the baseline estimate is capturing the effect of other services provided in the same locations, but one could argue that it could be driven by differences in language student composition by infrastructure type (i.e. those who attend language courses located in public schools are those who would learn the language more easily anyway). However, the test for the equality of the coefficients of exposure to the three types of infrastructures does not allow the null hypothesis to be rejected ($p$-value = 0.68), which overall leads us to consider that our main estimates are capturing a "genuine" effect of geographical exposure to language schools (and thus we retain overall exposure in what follows).

We also would like to discard the (residual) possibility that our estimates are still biased by the presence of endogenous residential sorting, even after controlling for a large set of census tract characteristics and district fixed effects. Hence, we replicated the analysis by excluding the individuals who might have chosen their place of residence considering the location of language centres. Specifically, Table 7 contains the results obtained after excluding individuals who had been living in the same place since before the creation of the CPLN [i.e. since 1989 in column (2), since 1985 in column (3), and since 1980 in column (4)]. This exercise provides estimates slightly higher than the baseline (but not statistically different), increasing with the length of stay in the current dwelling of residence, which goes against the idea that endogenous residential choices are affecting the validity of our main results. Moreover, given the ambiguity of the definition of exposure for those who moved to the current place of residence in 2000, just before the 2001 census, we also try to exclude them [column (5)] and again the evidence remains stable. Finally, as a last sensitivity check, we replicate the estimations by excluding observations of individuals residing in the periphery of the metropolitan area, which is useful for two possible issues. On the one hand, census tracts located away from the city centres are larger in terms of area, that is, the population density[21] is much lower in these places and this could be related to omitted local unobservables. On

[20] This evidence is also suggestive that measurement error in the computation of the measure of accessibility based on the number of language centres surrounding census tract's centroids is unlikely to explain the relatively small effect (in terms of the point estimate) of exposure. This is because the likelihood of incorrectly assigning surrounding language centres to each census tract should decrease with the length of the radii.

[21] Note that we do not have exact information about census tract area in 1991; we only observe the area in 2001. In alternative specifications, we tried to control for population density in 2001, which provides similar results (available upon request).

**Table 6**
Exposure by type of building.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Exposure 0.5 km | 0.030*** | | | | |
| | (0.010) | | | | |
| Exposure 0.5 km - schools | | 0.043** | | | 0.042** |
| | | (0.017) | | | (0.017) |
| Exposure 0.5 km - com. centres | | | 0.021 | | 0.022 |
| | | | (0.015) | | (0.015) |
| Exposure 0.5 km - munic | | | | 0.026 | 0.026 |
| | | | | (0.022) | (0.021) |
| Adjusted R-squared | 0.454 | 0.453 | 0.453 | 0.453 | 0.453 |
| Number of observations | 51624 | 51624 | 51624 | 51624 | 51624 |

OLS, dependent variable: self-reported knowledge of Catalan. Control variables as in column (7) of Table 2. Standard errors clustered at the census tract level; * significant at 10%, ** significant at 5%, *** significant at 1%.

**Table 7**
Restrictions on year of arrival in the current dwelling.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Exposure 0.5 km | 0.030*** | 0.031** | 0.038*** | 0.043*** | 0.033*** |
| | (0.010) | (0.013) | (0.014) | (0.016) | (0.010) |
| Adjusted R-squared | 0.454 | 0.460 | 0.466 | 0.477 | 0.456 |
| Number of observations | 51624 | 35572 | 30431 | 23002 | 50187 |

OLS, dependent variable: self-reported knowledge of Catalan. Control variables as in column (7) of Table 2. Standard errors clustered at the census tract level; * significant at 10%, ** significant at 5%, *** significant at 1%. Column (1): baseline estimates (column (7) of Table 2)). Column (2): only individuals living in the same place since 1989 or before. Column (3): only individuals living in the same place since 1985 or before. Column (4): only individuals living in the same place since 1980 or before. Column (5): only individuals living in the same place since 2000 or before.

the other hand, individuals residing in the periphery of the metropolitan area could have been exposed to other language centres located outside its limits (i.e. spatial spillovers from adjacent areas). The results are included in Table A5 of the online Appendix and are very similar to our main estimates.

### *5.1. Heterogeneous effects and local labour market outcomes*

The next two tables show the heterogeneous effects of exposure. The former concerns individual characteristics and the latter considers local characteristics. In column (2) of Table 9, we consider differential effects of spatial accessibility of language centres according to gender. The point estimates are positive and significant for both genders, although they are somewhat higher for females. However, the test of equality of the coefficients does not allow its null hypothesis to be rejected. In column (3), we allow for heterogeneous coefficients according to age intervals. Interestingly, the effect decreases with age, being much stronger for individuals in the age range of 29 to 37 years. This is possibly due to the fact that these individuals (born between 1964 and 1971) were only partially exposed to Catalan at school and did not obtain the proficiency certificate that enables people to work in the (regional and local) public sector in Catalonia. Because they are likely to compete for the same jobs with individuals who are slightly younger, they had an incentive to enrol on courses offered by the CPLN to obtain the language certificate. There is an additional incentive for younger adults to enrol in the courses compared to older individuals. Since acquiring better language skills can be seen as an investment in human capital, the expected benefits (return-to-investment) of becoming proficient in Catalan are likely to be higher if this investment is carried out earlier in life (and can be rewarded for a longer spam of the working life). Moreover, also the cost of acquiring new language skills is lower for younger individuals.

We also analysed heterogeneous effects depending on completed education [see column (4)]. Although the coefficient of exposure is clearly much higher for individuals with a low education level and not statistically significant for those with post-compulsory schooling, the null hypothesis of equality of coefficients is not rejected. Columns (5) and (6) provide information about the heterogeneous effects depending on the individual's origin. First, it is possible to appreciate that geographical exposure to language training centres only benefited, in terms of proficiency in the local language, individuals born in Catalonia. The coefficient estimated for individuals born in the rest of Spain is virtually zero and not significant.

In Column (6), we consider both the province of birth and age on arrival in Catalonia for those who were born abroad. Again, exposure to language schools does not affect language proficiency among individuals born outside Catalonia even if they arrived during childhood. This evidence is indeed consistent with the results reported in Table 3, which indicate that accessibility of language centres mostly affects the probability of being able to speak and write in Catalan. Indeed, language courses provided by the CPLN during its first decade of existence were very effective in shaping written proficiency among adult individuals born in Catalonia. Many of them were indeed native Catalan speakers who were orally fluent in Catalan thanks to intergenerational transmission of the language (Caminal et al., 2021) but did not have written skills because they

received education in Spanish only. Therefore, they took advantage of the language policy targeted to adults to acquire written skills in their native language. Moreover, it is very likely that the lack of effect of geographical exposure to language courses on language skills of individuals born in other Spanish regions could be related with information issues. That is, internal migrants might have had insufficient knowledge about the existence of publicly-provided language courses, which could explain why the availability of language centres in their neighbourhood of residence does not affect their skills in Catalan. Indeed, it is also possible that individuals proceeding from other Spanish region had a certain lack of knowledge about the potential socioeconomic benefits associated with the knowledge of Catalan in the bilingual region of Catalonia, therefore reducing their incentives to invest in the acquisition of skills in Catalan.

Finally, we explore the existence of heterogeneous effects according to local characteristics. We analyse whether the effect of exposure to language centres varies according to (a) the fraction of individuals born in the rest of Spain in the census tract, (b) the fraction of adult individuals who are able to speak in Catalan, and (c) the fraction of individuals who are able to speak and write in Catalan.

For the three contextual variables, we consider whether the share observed in the census tract is higher than the overall proportion or not. The results are reported in Table 9 and suggest that the impact of exposure on language skills is somewhat higher for individuals living in neighbourhoods with a higher share of individuals born outside Catalonia and (symmetrically) where there is a lower proportion of individuals who are proficient in Catalan. This seems to indicate that the public provision of local language training could mitigate the impact of neighbourhood composition, that is, the negative effect of living in a Spanish-speaking enclave on proficiency in Catalan. However, for any of the three variables considered here, the test for the equality of the estimated coefficients does not allow the corresponding null hypothesis to be rejected. Nevertheless, we consider that this issue deserves special attention and it will represent the focus of our future research.

As the last evidence, we looked into possible labour market effects induced by geographical exposure to local language schools. As mentioned before, we consider as outcomes the probability of being employed (including among those who are not employed, mostly women, who are involved in household tasks) and, among employed individuals, weekly hours of work and indicators of having a part-time job (below 30 h per week), working in the public sector, and having a highly skilled job. The first evidence we provide consists in the conditional correlation between skills in Catalan and these labour market outcomes (see Table 6A in the online Appendix). The results indicate that skills in Catalan are conditionally correlated with labour market outcomes and both the size and sign of the estimated coefficients are as expected. Specifically, the probability of working is positively associated with the level of proficiency in Catalan. The effect is larger when we keep in the sample individuals involved in housekeeping tasks, suggesting that language skills are also related to labour market participation. Only the highest levels of proficiency are positively related to the probability of working in the public sector and having a high-skilled job. The only exception is for hours of work (and the derived dummy for part-time jobs), for which any of the coefficient of skills in Catalan are statistically different from zero.

Nevertheless, when considering the reduced form estimates of the direct effect of exposure to language training centres, all the estimated coefficients are very low in size and not significantly different from zero (as can be appreciated in Table 10). This result is possibly due to the fact that, although skills in Catalan are rewarded in the labour market and actually correlate positively with several labour market outcomes considered here, the impact of exposure to Catalan learning centres on language skills is possibly too modest in size to induce any positive effect on labour market outcomes.[22]

## 6. Conclusions

In this paper, we investigated the effect of the geographical accessibility of local language learning centres on the formation of language skills and labour market outcomes in a bilingual urban labour market, the Metropolitan Area of Barcelona. We exploited the creation of the so-called Consortium for Language Normalization (CPLN), a public institution of the Spanish region of Catalonia that provided publicly subsidized courses of Catalan (the local language of Catalonia) targeted to adult individuals who were schooled in Spanish only. The empirical analysis considers the first decade of existence of the CPLN (1989/90 to 2000/01) and is based on a measure of spatial availability of language courses that captures potential exposure in an Intention-to-Treat (ITT) framework.

Our results show a quantitatively modest but robust effect of exposure to language centres on language proficiency, especially regarding the probability of being able to speak and write in Catalan. All the evidence from falsification exercises and sensitivity checks points towards the validity of the causal interpretation of our findings. The impact of the local supply of language courses is more pronounced for younger and less educated individuals and, most strikingly, is relevant only for those who were born in Catalonia: they are indeed those who are more sensitive to an increase of spatial accessibility (i.e. having more language centres near to the place of residence). Actually, this result is consistent with the fact that the availability of language learning centres only affects the probability of speaking and writing in Catalan. In fact, the latter domain of language proficiency is precisely what people born in Catalonia (many of them being native Catalan

[22] We also analysed possible heterogeneous effects on labour market outcomes, considering the same variables that we used in Tables 8 and 9. However, in this case too we were unable to detect any significant effect of exposure to language schools on labour market outcomes (results are available upon request).

**Table 8**
Heterogeneous effects, individual characteristics.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Exposure 0.5 km | 0.030*** | | | | | |
| | (0.010) | | | | | |
| Exposure 0.5 km - female | | 0.033*** | | | | |
| | | (0.011) | | | | |
| Exposure 0.5 km - male | | 0.027** | | | | |
| | | (0.012) | | | | |
| Exposure 0.5 km - age 30–37 | | | 0.064*** | | | |
| | | | (0.016) | | | |
| Exposure 0.5 km - age 38–46 | | | 0.031* | | | |
| | | | (0.017) | | | |
| Exposure 0.5 km - age 47–55 | | | 0.014 | | | |
| | | | (0.016) | | | |
| Exposure 0.5 km - age 56–64 | | | 0.015 | | | |
| | | | (0.015) | | | |
| Exposure 0.5 km - low education | | | | 0.034*** | | |
| | | | | (0.012) | | |
| Exposure 0.5 km - high education | | | | 0.021 | | |
| | | | | (0.013) | | |
| Exposure 0.5 km - born in Catalonia | | | | | 0.053*** | |
| | | | | | (0.011) | |
| Exposure 0.5 km - born in the rest of Spain | | | | | −0.006 | |
| | | | | | (0.014) | |
| Exposure 0.5 km - born in Catalonia | | | | | | 0.047*** |
| | | | | | | (0.011) |
| Exposure 0.5 km - arrived with age 1–13 | | | | | | 0.001 |
| | | | | | | (0.026) |
| Exposure 0.5 km - arrived with age 14–29 | | | | | | 0.002 |
| | | | | | | (0.018) |
| Exposure 0.5 km - arrived with age 30 or more | | | | | | −0.046 |
| | | | | | | (0.037) |
| p-value for coefficients' equality test | – | 0.605 | 0.030 | 0.418 | 0.000 | 0.013 |
| Adjusted R-squared | 0.454 | 0.453 | 0.453 | 0.424 | 0.454 | 0.451 |
| Number of observations | 51624 | 51624 | 51624 | 51624 | 51624 | 51624 |

OLS, dependent variable: self-reported knowledge of Catalan. Standard errors clustered at the census tract level; * significant at 10%, ** significant at 5%, *** significant at 1%. Control variables as in column (7) of Table 2, except: column (3), control for age groups dummies rather than cubic age. Column (4), control for high–low education (post-compulsory vs lower levels) rather than years of schooling. Column (6), control for age of arrival dummies rather than quadratic age at arrival.

speakers) wanted to achieve, because they were not exposed to language at school and had limited written skills. However, although competences in Catalan are positively related to labour market performance (especially employment probability and occupation), the effects induced by the geographical accessibility are too modest to lead to any improvements in any of the labour market outcomes considered in this work.

The evidence reported in this paper indicates that the public provision of local language training and its spatial accessibility positively affect the acquisition and improvement of language competences in urban labour markets. Therefore, a first policy implication that can be derived from our findings is that geographic proximity helps to increase the effectiveness of language training programmes targeted to adults, which is also in line with the results regarding the spatial decay effect we have detected. This means that decentralizing the location of language courses and trying to reach target individuals is a sensible route to follow. However, the results regarding the heterogeneous effects, especially the fact that only individuals born in Catalonia are affected by the presence of language schools near to their place of residence, should also be carefully considered by policymakers. In fact, this indicates that additional effort is needed to stimulate the acquisition of language skills among individuals who are in a less advantaged position, who are (in this case) individuals born in other parts of Spain. Indeed, especially during the period of analysis, internal migrants and their descendants were likely to be spatially segregated in specific areas of the city (Garcia-Lopez et al., 2020). Accordingly, the geography of the supply of language centres should be more directly based on the residential location of the target population, in order to increase their availability in areas with a higher concentration of inhabitants with more need to improve their skills in the local language. This will possibly contribute to stimulating participation in language courses among individuals who have much more to gain from training in local languages. Moreover, it is possible that information about the availability of local language courses, as well as about the potential benefits related to local language acquisition, could also play some role in explaining the absence of any reaction to local supply of language courses among individuals

**Table 9**
Heterogeneous effects, local characteristics.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Exposure 0.5 km | 0.030*** | | | |
| | (0.010) | | | |
| Exposure 0.5 km - low % born in the rest of Spain | | 0.032*** | | |
| | | (0.012) | | |
| Exposure 0.5 km - high % born in the rest of Spain | | 0.041** | | |
| | | (0.018) | | |
| Exposure 0.5 km - low % at least speak in Catalan | | | 0.040** | |
| | | | (0.018) | |
| Exposure 0.5 km - high % at least speak in Catalan | | | 0.031*** | |
| | | | (0.012) | |
| Exposure 0.5 km - low % speak and write in Catalan | | | | 0.038*** |
| | | | | (0.018) |
| Exposure 0.5 km - high % speak and write in Catalan | | | | 0.027** |
| | | | | (0.011) |
| p-value for coefficients' equality test | – | 0.681 | 0.636 | 0.578 |
| Adjusted R-squared | 0.454 | 0.453 | 0.454 | 0.455 |
| Number of observations | 51624 | 51625 | 51626 | 51627 |

OLS, dependent variable: self-reported knowledge of Catalan. Standard errors clustered at the census tract level; * significant at 10%, ** significant at 5%, *** significant at 1%. Control variables as in column (7) of Table 2, except: column (3), control for an indicator of high/low fraction of individuals born in the rest of Spain rather than the share. Column (4), control for an indicator of high/low fraction of individuals who are able to speak or speak and write in Catalan rather than the share of individuals born in the rest of Spain. Column (5), control for an indicator of high/low fraction of individuals who are able to speak write in Catalan rather than the share of individuals born in the rest of Spain.

**Table 10**
Labour market outcomes.

| Outcome: | Employed | Empl. ext. | Hours of work | Partime | Public sector | High skilled |
|---|---|---|---|---|---|---|
| Exposure 0.5 km | 0.000 | 0.003 | 0.11 | −0.002 | 0.000 | −0.001 |
| | (0.002) | (0.003) | (0.075) | (0.002) | (0.003) | (0.003) |
| Adjusted R-squared | 0.025 | 0.237 | 0.054 | 0.044 | 0.156 | 0.339 |
| Number of observations | 34931 | 45483 | 31528 | 31358 | 31359 | 31337 |

OLS, control variables as in column (7) of Table 2. Standard errors clustered at the census tract level; * significant at 10%, ** significant at 5%, *** significant at 1%.

born outside Catalonia. Although this possible channel cannot be investigated with available data, policymakers should more carefully consider implementing active information policies to reach their target groups and increase their language skills endowment. Finally, the absence of effects on labour market outcomes should be better analysed in the future, but one tentative policy recommendation in this line is to review the content of language courses and their organization not only to foster their impact on language skills acquisition but also to orientate language courses towards the competences that are relevant in the bilingual labour market.

## Acknowledgements

We acknowledge the financial support of the project PID2021-122575NB-I00 (Spanish Ministry of Science and Innovation). This paper benefited from comments received in seminars (IREA, URV and Secretaria de Política Lingüística, Generalitat de Catalunya), workshops and conferences (Catalan Economic Society Conference, LEER Conference on Education Economics, IAAEU Workshop on Labour Economics), as well as for the comments received by two anonymous referees. We are also grateful to the support provided by technicians of CPNL, IDESCAT and INE and for providing us the data used in this paper.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.eap.2023.08.001.

## References

Åslund, O., Engdahl, M., 2018. The value of earning for learning: Performance bonuses in immigrant language training. Econ. Educ. Rev. 62, 192–204.
Bleakley, H., Chin, A., 2004. Language skills and earnings: Evidence from childhood immigrants. Rev. Econ. Stat. 86 (2), 481–496.
Bleakley, H., Chin, A., 2010. Age at arrival, English proficiency, and social assimilation among US immigrants. Am Econ. J.: Appl. Econ. 2 (1), 165–192.

Caminal, R., Cappellari, L., Di Paolo, A., 2021. Language-in-education, language skills and the intergenerational transmission of language in a bilingual society. Labour Econ. 70, 101975.
Caminal, R., Di Paolo, A., 2019. Your language or mine? The non-communicative benefits of language skills. Econ. Inq. 57 (1), 726–750.
Cappellari, L., Di Paolo, A., 2018. Bilingual schooling and earnings: Evidence from a language-in-education reform. Econ. Educ. Rev. 64, 90–101.
Chiswick, B.R., Miller, P.W., 2007. The Economics of Language: International Analyses. Routledge.
Domínguez, M., Montolio, D., 2021. Bolstering community ties as a means of reducing crime. J. Econ. Behav. Organ. 191, 916–945.
Foged, M., Hasager, L., Peri, G., Arendt, J.N., Bolvig, I., 2022. Language training and refugees' integration. Rev. Econ. Stat. 1–41.
Garcia-Lopez, M.À., Nicolini, R., Roig, J.L., 2020. Segregation and urban spatial structure in Barcelona. Pap. Reg. Sci. 99 (3), 749–772.
Giesecke, M., Schuss, E., 2019. Heterogeneity in Marginal Returns to Language Training of Immigrants. IAB-Discussion Paper No. 19/2019.
Gonzalez, L., Ortega, F., 2011. How do very open economies adjust to large immigration flows? Evidence from Spanish regions. Labour Econ. 18 (1), 57–70.
Heller, B., Slungaard Mumma, K., 2023. Immigrant integration in the United States: The role of adult english language training. Am. Econ. J.: Econ. Policy (forthcoming).
Kivi, L.H., Sõmer, M., Kallaste, E., 2020. Language training for unemployed non-natives: who benefits the most? Baltic J. Econ. 20 (1), 34–58.
Lang, J., 2021. Employment effects of language training for unemployed immigrants. J. Popul. Econ. 35, 719–754.
Lochmann, A., Rapoport, H., Speciale, B., 2019. The effect of language training on immigrants' economic integration: Empirical evidence from France. Eur. Econ. Rev. 113, 265–296.
Pont-Grau, A., Lei, Y.H., Lim, J.Z., Xia, X., 2023. The effect of language training on immigrants' integration: Does the duration of training matter? J. Econ. Behav. Organ. 212, 160–198.