



**HAL**  
open science

# Ethnic diversity and educational success: Evidence from France

Sylvain Chareyron, Amélie Chung, Patrick Domingues

► **To cite this version:**

Sylvain Chareyron, Amélie Chung, Patrick Domingues. Ethnic diversity and educational success: Evidence from France. *Research in Economics*, 2021, 75 (2), pp.133-143. 10.1016/j.rie.2021.04.002 . hal-04268529

**HAL Id: hal-04268529**

**<https://hal.science/hal-04268529v1>**

Submitted on 22 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Ethnic Diversity and Educational Success: evidence from France

Sylvain Chareyron <sup>1</sup>, Amélie Chung <sup>2</sup> and Patrick Domingues <sup>3</sup>

## Abstract:

This paper explores the relationship between ethnic diversity and adolescent educational outcomes at a very fine level. The French Labor Force Survey is used to obtain information on neighborhood composition and we use repetition of the school year at age 15 to 16 as an indicator of school underperformance. We use the method of Bayer, Ross, and Topa (2008) to take account of endogeneity. Although high ethnic diversity of the residential area and educational success seem to be negatively related at first sight, we show that the causal relationship between the two disappears when all confounding factors are properly controlled for. This result is robust for alternative definitions of origin.

**Keywords:** neighborhood effects, ethnic diversity, education

**JEL Code:** I21, J15, R23

## Declarations:

Declaration of Interest: None

## Section 1: Introduction

---

<sup>1</sup> University Paris Est Créteil (UPEC), ERUDITE (EA 437), UPEC, UGE, TEPP (FR 3435): sylvain.chareyron@u-pec.fr, 61 Av. du Général de Gaulle 94010 Créteil

<sup>2</sup> University of New Caledonia, LARJE : [chung.amelie@gmail.com](mailto:chung.amelie@gmail.com), 145 Av. James Cook 98851 Nouméa

<sup>3</sup> University Paris Est Créteil (UPEC), ERUDITE (EA 437): patrick.domingues@u-pec.fr, 61 Av. du Général de Gaulle 94010 Créteil

In some developed countries, people from the ethnic majority tend to be reluctant to live in ethnically diverse neighborhoods and may even choose to move out of neighborhoods that they consider have become overly diverse (Card, Mas, and Rothstein, 2008). This reaction tends to reinforce the dynamics of spatial segregation and its associated adverse effects (Boozer *et al.*, 1992; Grogger, 1996). It may be explained by an aversion to ethnic diversity (Alesina and La Ferrara, 2005) but also by the fear of people from the main ethnic group that their children will suffer from poor educational conditions. This thinking may come from first sight observations, as can be illustrated in the French educational system where a *prima facie* analysis would tend to conclude that more ethnically diverse neighborhoods are associated with a higher school failure rate. Figure 1 plots the correlation between ethnic diversity and repetition of an academic year in France. The figure shows a positive correlation between diversity and educational delay. However, this apparent negative relationship between high ethnic diversity and educational success may be biased by many confounding factors, since ethnically diverse neighborhoods tend to be associated with other factors that reduce educational success. In this paper, we are trying to delve deeper into this apparent relationship and disentangle the connections between ethnic composition of the residential area and school performance.

**[Insert Figure 1 here]**

A substantial literature has built up in evaluating the economic and social implications of ethnic diversity in both developing and developed countries. For instance, the canonical paper by Easterly and Levine (1997) pointed out that ethnic diversity is an explanatory factor in weak economic growth performance. More recently, Awaworyi Churchill, Nuhu, and Lopez (2018), based on an analysis of macroeconomic data, found that greater ethnic diversity

is associated with increased gender gaps. It also occurs alongside lower participation in community life (Alesina and La Ferrara, 2005), and is found to have a negative effect on social solidarity and the level of inter-racial trust (Putnam, 2007). Diversity is also associated with negative business performances (O'Reilly, Williams, and Barsade, 1998) as well as lower spending on productive public goods (Alesina, Baqir, and Easterly, 1999; Algan, Hémet, and Laitin, 2016). In addition, some studies have shown that expenditure on public goods and services are lower in more fragmented areas, particularly expenditure on public education (Poterba, 1997; Goldin and Katz, 1999; Vigdor, 2004).

Scholars addressing the effects of ethnic composition on educational indicators have more mixed findings. In some cases, a high proportion of non-natives has a partially negative effect on test scores (Hoxby, 2000; Hanushek, Kain, and Rivkin, 2009). Others find that the proportion of ethnic minorities does not have any significant effect on performance (Card and Rothstein, 2007; Angrist and Lang, 2004). While still others, such as Maestri (2017), find that ethnic diversity has a positive effect on the test score of minority students. This lack of consensus is also reflected in studies on the effect of racial segregation and isolation in the US context. In 1968, the Coleman report offered evidence that ethnic diversity undermined academic achievement, showing that the performance of black students declined in schools with a high proportion of black students, and Hanushek (1972) found that a higher concentration of blacks affects both blacks and whites. More recently, Kain and O'Brien (2000) find that black pupils gain more value than white pupils when they move to higher quality schools in more mixed neighborhoods, while Rivkin (2000) finds no evidence of a positive effect on academic attainment for blacks who are more exposed to whites. Hanushek

*et al.* (2009) underline that the contradictory findings in the literature dealing with ethnic composition stem from a difficulty with identifying the causal effect of peer characteristics.<sup>4</sup>

Several mechanisms have been proposed to explain why ethnic diversity can affect academic performance in a positive or negative way (Dronkers and van der Velden, 2013). For instance, Lazear (1999) suggests that ethnic diversity boosts language acquisition, especially in a school context where minorities have more incentives to espouse the majority culture. Ottaviano and Peri (2006) pursue the idea advanced by Lazear (1998), according to which ethnic diversity stimulates students' curiosity. The idea that diversity is enriching to a society comes from the fact that different ethnic groups bring different skills which can mutually enhance each other and contribute to productivity. Negative effects may arise from the fact that a more heterogeneous student population may inhibit a teacher from specializing her or his teaching, thus reducing the effectiveness of teaching.

These seminal papers have focused on ethnic diversity at the school or the classroom levels; but we find few studies in the economic literature on the links between the ethnic heterogeneity of a residential area and school achievement. Furthermore, although 9.2% of the French population were born abroad and 11% are individuals born in France with at least one immigrant parent,<sup>5</sup> little work has been done on the economic impacts of ethnic diversity in this country.<sup>6</sup> Consequently, lack of proper identification of the real causes and effects is

---

<sup>4</sup> Manski (1993) and Moffitt (2001) provide explanations of the problems associated with concepts and data treatment.

<sup>5</sup> From the INSEE statistics on Nationality and Immigration (2015)

<sup>6</sup>One exception is the recent paper by Hémet and Malgouyres (2018) which explores how diversity affects employment probability and productivity at a local level. They show that the probability of employment is negatively correlated with the ethnic diversity of the *aire* and positively correlated with the diversity of the employment area.

problematic in the French context because it can lead to a belief in and a reliance on apparent relationships that may be biased by many confounding factors.

Therefore, we are using French data to evaluate the potential effect of the ethnic composition of an individual's residential area on his or her school performance. The aim is to determine whether ethnic heterogeneity is a contributory factor to educational underachievement. To this end, we follow Goux and Maurin (2005) who chose the repetition of the school year between the ages of 15 and 16 as an indicator of school underperformance. We measure ethnic heterogeneity at a very fine level so that we can obtain a credible measure of the exogenous effect of ethnic diversity on educational success. Following the example of Hémet and Malgouyres (2018), we define ethnic origin using three methods of categorization: (i) nationality; (ii) country of birth; (iii) origin of parents. In the next section we focus on the various definitions and categorizations of ethnicity used in the literature. Finally, we apply the reasoning of Bayer, Ross, and Topa (2008) to the endogeneity of ethnic diversity at the neighborhood level. Although a superficial analysis of the results shows that ethnic diversity and educational success are negatively related, this causal connection disappears when all confounding factors are properly controlled for, regardless of the definition of ethnic origin used.

The paper is organized as follows. Section 2 focuses on the plurality of approaches to ethnic diversity and its endogeneity in the economic literature. It surveys and assesses the literature for the positive and negative effects of ethnic diversity on economic policies and outcomes. Data treatment and variable construction are described in Section 3 and Section 4 outlines the empirical strategy. Some concluding elements are presented in Section 7.

## **Section 2: Ethnic diversity**

Defining ethnicity is not an easy task. The definitions and tools used to measure it depend on the economic approach and the regional context studied. Alesina and La Ferrara (2005) highlight the variety of criteria used for ethnic identification in the literature in terms of analyses and theoretical or empirical contributions. For instance, the empirical literature on inter-country comparisons generally uses a measure called ethno-linguistic fractionalization (identification of ethnicity based on skin-color and/or language), while the development literature makes more use of language or tribal affiliation. Studies in urban economics, and more specifically those conducted in American cities, use ‘race’ as defined by the five categories used in the census: (i) White, (ii) Black, (iii) American Indian, (iv) Asian, Pacific Islander, (v) Other (in which Hispanic is included).

Furthermore, there are generally two main ways to consider ethnicity. The first, described as objective, is based on criteria such as nationality, citizenship, country of birth or parents’ country of origin. The second is takes a more subjective approach and considers such factors as the feeling of belonging to a culture or society.<sup>7</sup> Objective identification of ethnicity has some advantages over subjective identification. First, it can be difficult to delimit an ethnic group or culture, precisely and without bias (Horowitz, 2001). Indeed, several studies show that it is difficult for an individual to classify others into given ethnic categories. Second, identifying oneself in an ethnic category necessarily depends on the socio-economic and political context at a given time, as people tend to define themselves in response to a specific context (Laitin, 2000; Bloch and Rao, 2001).

---

<sup>7</sup> Constant, Gataullina, and Zimmermann (2009) have developed the ethnosizer test to measure the intensity of ethnic identity.

The collection and recording of information indicating the ethnic origins or religious affiliations of individuals remains not allowed in France. However, data on individuals by nationality are available. Therefore, following Hémet and Malgouyres (2018), we adopted an objective approach and selected three parameters to categorize ethnic origin: (i) nationality; (ii) country of birth and (iii) origin of parents. These three parameters of ethnic identification are not without consequences since, particularly in education, the pupils best equipped to benefit from school culture are those who already possess the codes within their cultural and family environment. Thus, first-generation students have quite a different experience from second-generation students in the host country (Dronkers and van der Velden, 2013). The 2012 OECD report on indicators of immigrant integration reveals significant differences in average reading literacy scores (PISA) by ethnic origin. The score difference in France between children whose parents were born in the country and children who either immigrated to the country or were born to foreign-born parents is among the highest in OECD countries (OECD and European Union, 2015).

### **Section 3: Data and Variables**

To investigate the effect of ethnic diversity on academic success, we use data from the French Labor Force Survey (LFS) conducted each year by the *Institut National de la Statistique et des Etudes Economiques* (INSEE) between 2005 and 2017. This survey is designed to collect information on individuals aged 15 years and over living in groups of approximately 20 adjacent dwellings<sup>8</sup>, defined as a small neighborhood unit called an *aire*.<sup>9</sup>

---

<sup>8</sup> Only main and ordinary dwellings are covered, secondary residences or holiday homes are excluded from the sample.

<sup>9</sup> Compared to the “Output Area” used by Gibbons, Silva, and Weinhardt (2013), applying neighbourhood units that are smaller than the standard previously used (125 households on average), our neighbourhood units are



The specificity of this survey comes from its sample design. The sample is rotational, one sixth of the *aires* selected being replaced by new *aires* each quarter. A dwelling is surveyed over six consecutive quarters before being removed from the sample, so that the survey takes the form of a quasi-panel.

Data recorded by the survey are multipurpose. This survey collects information on gender, date and place of birth, nationality, family composition, labor market situation, educational level, etc. For our purposes, we focus on 15-year-olds for whom we have two survey responses one year apart and who have at least one other 15-year-old in their *aire*. The sample therefore consists of 7,987 individuals. For each individual, we construct indicators of academic success based on academic delay. Although there has been a large decline in the rate of grade repetition in France since 2000, the need to repeat a grade remains a good indicator of underperformance at school. The 2015 report by the Organization for Economic Co-operation and Development's report (OECD), the Program for International Student Assessment (PISA), indicates that the French school system is among the most unequal in terms of student performance, and that public policies must invest in resources to combat school failure.<sup>10</sup> Although the proportion of repeating students fell by 16 points between 2009 and 2015, it is still double the OECD average, with 22% of students repeating at least once before the age of 15.<sup>11</sup> Hence, we consider that a 15-year-old individual has been held back in the grade between 15 and 16 years if he or she stays in the same grade between the two interviews. We consider that he or she has been held back in a grade before 15 if he or she is in a grade lower than the ninth grade or is in vocational orientation at 15.

---

small enough (20 dwellings on average) to provide a credible scale for the assessment of social interaction effects.

<sup>10</sup> See the 2016 OECD report: PISA 2015, The Challenges of the French Education System and Best International Practices (*Les défis du système éducatif français et les bonnes pratiques internationales*).

<sup>11</sup> From: PISA 2015 Results (Volume I): Excellence and Equality in Education – Preliminary Version. 2016

To this dataset, we add information on the individuals' environments. The main environmental information we add is a fractionalization index that indicates the level of ethnic diversity in the area. The fractionalization index is calculated using the same method as Alesina *et al.* (2003):

$$DIV_j = \sum_{i=1}^N s_{ij}(1 - s_{ij}) = 1 - \sum_{i=1}^N s_{ij}^2 \quad (1)$$

where  $s_{ij}$  is the share of individuals from group  $i$  in geographic area  $j$ . The index is 0 in a homogenous area and converges to 1 with high population heterogeneity. In line with Hémet and Malgouyres (2018), we consider six groups of origins: France, Southern Europe, Other European countries, Maghreb, Other African countries and Rest of the World.

The group to which one belongs depends on which definition of origin is considered. Origin is defined using three different parameters: nationality, country of birth and origin of the parents. For instance, in the definition by nationality, an individual is classified in the Maghreb group if he or she holds the nationality of a Maghrebian country. Under the definition by country of birth, an individual is classified in the Maghreb group if he or she was born in a Maghrebian country. Under the definition by origin of parents, an individual will be classified in the Maghreb group if he or she holds the nationality of a Maghrebian country or if at least one of his or her parents has the nationality of a Maghrebian country or was born in a Maghrebian country. As argued by Hémet and Malgouyres (2018), these different measures of diversity capture different dimensions of diversity. When diversity is defined by origin of parents, it can better reflect ethnic diversity because it is more likely to capture visible differences such as color of the skin. Diversity defined by nationality can be

considered to be an indicator of cultural diversity as two people of the same nationality are more likely to speak the same language.

Table 1 and Table 2 present the variety of origins of individuals living in France, based on the three alternative definitions of origin. The proportion of individuals defined as French is lower when the definition of diversity applied is based on origin of parents than when it is based on the individual's nationality alone. In other words, the French population appears quite homogeneous when nationality is the only factor considered. This is because people of foreign origin tend to acquire French nationality after they have lived in the territory for some years.

**[Insert Table 1 here]**

We observe in Table 2 that diversity is higher when the origins of the parents are considered rather than nationality: the probability that two individuals randomly selected from an *aire* belong to two different ethnic groups is 10% on average when nationality is the defining attribute but is 34% on average when origin of the parents is the defining attribute.

**[Insert Table 2 here]**

Table 3 presents the characteristics of the adolescents in the sample as well as the characteristics of their environment. Finally, to take account of peer effects that may influence the level of education (Izaguirre and Di Capua, 2020), we calculate, for each 15-year-old individual, the proportion of other 15-year-olds in the *aire* who repeated a grade before the

age of 15. It should be noted that for each adolescent, the proportion is calculated using only information on individuals who do not belong to his or her family.

**[Insert Table 3 here]**

#### **Section 4: Empirical Strategy**

We use the following equation to estimate the effect of diversity on educational performance:

$$REP_{ijt} = \alpha + \beta DIV_{jt} + \omega P_{jt} + \gamma X_{it} + \Phi_g + \mu_t + \varepsilon_{ijt} \quad (2)$$

where  $REP_{ijt}$  is a dummy variable indicating repetition of the grade between 15 and 16 years of age of adolescent  $i$  living in *aire*  $j$  at time  $t$ .  $DIV_{jt}$  is the level of diversity in *aire*  $j$  at time  $t$ . We alternate the use of the three definitions of origin diversity to produce separate estimates.  $P_{jt}$  is the proportion of other adolescents in the *aire* who have repeated the grade at 15 years old in order to capture peer effects.  $X_{it}$  is a set of individual control variables: socio-economic category of mother and father, gender and origin of the individual. In each estimation, the definition of origin applied to the individual is the same as the one used to measure diversity of the *aire*. The values for the origin variable are: France, Southern Europe, Other European countries, Maghreb, Other African countries and the Rest of the World.  $\Phi_g$  are the geographic fixed effects and  $\mu_t$  are time fixed effects (quarter and year dummies).  $\varepsilon_{ijt}$  is an error term.

The estimations of equation (2) give unbiased estimates of the effect of the level of diversity on educational performance if the level of diversity in the *aire* is exogenous relative to the probability of grade repetition. However, this assumption is unlikely to be true as immigrants probably choose, for economic reasons, to settle more in places with a lower than average level of education. This would lead to an incorrect attribution of an effect to diversity, one that in fact comes from unobserved characteristics of the environment. To address this issue, we use a methodology similar to that of Hémet and Malgouyres (2018). Based on the work of Bayer, Ross, and Topa (2008), they argue that households are unable to choose a specific neighborhood within the area where they decide to live for reasons such as limited housing stock and the difficulty of identifying and comparing the characteristics of each neighborhood in an area. This assumption is realistic in the case of the LFS because the neighborhoods/*aires* considered are very small (about 20 dwellings) and do not correspond to any administrative area. This means that once the characteristics at the relevant geographical scale for household location choice have been taken into account, the distribution of households within the *aire* of this zone is exogenous. Furthermore, the *aire* is not an official or administrative division: it is a unit used only in the LFS, thus individuals are not aware of what an *aire* is and they cannot expressly decide to live in a particular *aire* (Hémet and Malgouyres 2018). In the dataset, we have three geographical units that are larger than the unit of the *aire*: the *région*, the *département* and the *secteur*.<sup>12</sup> We believe that the *secteur* is the most appropriate relevant geographical unit to control for, but we are able to test the different possible scales by adding fixed effects to the estimation.

---

<sup>12</sup> The *secteur* is the smallest identifiable area after the *aire*. It contains between 120 and 240 households and thus between 6 and 11 *aires*. In France a *département* is a geographical and administrative unit that can be thought as a county. There are 101 of them with an average population of about 660,000 inhabitants. A *region* is an administrative unit above the *département*. There are 18 of them with an average population of about 3.7 million.

Figure 2 presents, for the different definitions of origins, the distribution of the level of diversity in the *aires* and the deviation with the average level of diversity in the *région*, *département* and *secteur*. First, it can be observed that the distributions are similar for the definitions of diversity based on nationality and country of birth, with a substantial concentration of very homogenous *aires*. However, when diversity is based on origin of parents, the distribution is much less concentrated. We also observe that the variability of the distributions decreases with an increase in scale of the geographical areas from which we center the measure. A big step appears when with de-mean at the *secteur* level, which is to be expected as this is at a much smaller scale than the *région* and *département*. However, there is still a certain amount of variation when the deviation from the *secteur* mean is examined.

**[Insert Figure 2 here]**

## **Section 5: Results**

Table 4 presents the results of the linear probability models for the different definitions of ethnic diversity: nationality, country of birth and origin of the parents.<sup>13</sup> We begin in column (1) with a naïve regression of the probability of grade repetition with the ethnic diversity in the *aire*. This first approach shows a large and significant increase of grade repetition as the level of ethnic diversity increases, regardless of which of the three alternative definitions of origin is used. Then, in column (2), as we add individual control variables we find that our estimates remain large and significant although their magnitude tends to decrease. In column (3), we control for the proportion of other adolescents in the *aire* who

---

<sup>13</sup> Probit estimates give similar results and are available upon request.

have repeated the grade at 15 years old to capture peer effects and we continue to find very similar results.

As indicated in the previous section, the geographical units studied are so small that we can consider the selection of individual from one *aire* to another as random taking into account the selection at the relevant geographic scale. Therefore, to go further, we account for the potential endogeneity of the level of diversity at the *aire* level, including geographic fixed effects at different scales. In column (4), we include time fixed effects and geographic fixed effects at the *région* level while in column (5) we use the *département* level.<sup>14</sup> In column (5), when all control variables and the *département* fixed effects are included, the magnitude of the estimates are between 30 and 97 percent lower than in the naïve estimate (depending on the definition of origin used) and the relationship is not statistically significant when diversity is defined by parents' origin. There subsists, however, a significant effect with the two other definitions of origin.

It is likely that, inside a *département*, endogenous sorting of individuals could still be an issue. This is why we include in column (6) fixed effects at the *secteur* level, a much finer geographical unit, generally below the municipal level. When controls for *secteur* fixed effects are included, it becomes clear that all the estimated coefficients statistically arrive at zero. Considering that the *secteur* is presumed to be the most relevant scale for controlling spatial selection in our set of available variables, it can be seen that ethnic diversity has no effect on education at the very local level.

**[Insert Table 4 here]**

---

<sup>14</sup> The estimated coefficients of diversity are presented in Table 4 but more extensive results of the regression to evaluate the relationship between grade repetition and diversity by nationality can be found in Table A1. It can be seen in Table A1 that we explain only a very limited part of the variance of grade repetition. However, the  $R^2$  are in range of the ones obtained by Goux and Maurin (2007).

## **Section 6: Robustness Checks**

We conduct three robustness checks. First, we estimate the same specification as in column (6) of Table 4, but this time by clustering the standard errors at the *département* level rather than at the *aire* level. The *aire* level permits to flexibly account for the structure of the error term but is exposed to relatively large standard errors due to the small number of observations per *aire*. The results are presented in Table A2. Standard errors, however, change only very slightly when they are clustered at the *département* level and the significance of the results does not change. Second, we remove the French overseas territories from the sample because they may have particular characteristics that could bias the results. Again, the results presented in Table A3, are quite insensitive to this change. Finally, we consider a different outcome to measure educational success: repetition of a grade before 16 years old. Grade repetition between 15 and 16 is a more instantaneous measure of educational success and is more likely to reflect present living conditions than repetition of a previous grade. However, this second outcome allows us to capture a cumulated effect and avoids being conditional on prior repetition. In any case, we find similar effects in both outcomes. Table A4 in appendix also shows that the negative effect of diversity on educational success disappears when control variables are included.

## **Section 7: Conclusion**

The literature shows that in some developed countries, ethnically diverse neighborhoods tend to be avoided by the ethnic majority (Card, Mas, and Rothstein, 2008). In order to explain this phenomenon, scholars generally advance a theory of aversion to ethnic



diversity. Alesina and La Ferrara (2005) explain that inter-group behaviors can be described by considering that “individuals may attribute positive utility to wellbeing of members of their own group and negative utility to that of members of other groups”. However, another reason for this behavior could be that people from the main ethnic group fear that their children will suffer from poor educational conditions in ethnically diverse neighborhoods.

In this paper, using French data, we analyze the influence of ethnic diversity on educational performance at a very fine geographical level. We believe that this small scale allows us to credibly estimate the exogenous effect of ethnic diversity on educational success. We find that although a naïve estimate could conclude that ethnic diversity has a negative effect on educational success, the level of ethnic diversity in the *aire* has no effect on the probability of repeating a grade between the ages of 15 and 16 when endogenous sorting is properly controlled for.

This result shows that even at a very fine geographical level, endogenous selection may lead to a false conclusion when proper controls are not included. This issue is more pregnant at higher aggregate scales where it is more difficult to consider selection as random. This underlines the difficulty of studying the effect of ethnic diversity on economic or social outcomes and highlights the need to avoid drawing conclusions from superficial observation.

## References:

- Alesina, A., R. Baqir, and W. Easterly. 1999. “Public Goods and Ethnic Divisions.” *Quarterly Journal of Economics* 114 (4): 1243–84.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. “Fractionalization.” *Journal of Economic Growth* 8 (2): 155–94.
- Alesina, Alberto, and Eliana La Ferrara. 2005. “Ethnic Diversity and Economic Performance.” *Journal of Economic Literature* 43 (3): 762–800.
- Algan, Yann, Camille Hémet, and David D. Laitin. 2016. “The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation.” *Journal of Political Economy* 124 (3): 696–733.
- Angrist, Joshua D., and Kevin Lang. 2004. “Does School Integration Generate Peer Effects? Evidence from Boston’s Metco Program.” *American Economic Review* 94 (5): 1613–34.

- Awaworyi Churchill, Sefa, Ahmed Salim Nuhu, and Karely Lopez. 2018. "Persistence of Gender Inequality: The Role of Ethnic Divisions." *Applied Economics*, August, 1–16.
- Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy* 116 (6):
- Bloch, Francis, and Francis Rao. 2001. "Statistical Discrimination and Social Assimilation" 10 (2): 1–5.
- Boozer, Michael A., Alan B. Krueger, Shari Wolkon, John C. Haltiwanger, and Glenn Loury. 1992. "Race and School Quality Since Brown v. Board of Education." *Brookings Papers on Economic Activity. Microeconomics* 1992: 269–338.
- Card, David. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics* 19 (1): 22–64.
- Card, David, Alexandre Mas, and Jesse Rothstein. 2008. "Tipping and the Dynamics of Segregation." *Quarterly Journal of Economics* 123 (1): 177–218.
- Card, David, and Jesse Rothstein. 2007. "Racial Segregation and the Black–White Test Score Gap." *Journal of Public Economics* 91 (11): 2158–84.
- Caselli, Francesco, and Wilbur John Coleman II. 2013. "On The Theory Of Ethnic Conflict." *Journal of the European Economic Association* 11 (s1): 161–92.
- Coleman, James S. 1968. "Equality Of Educational Opportunity." *Equity & Excellence in Education* 6 (5): 19–28.
- Constant, Amelie F., Liliya Gataullina, and Klaus F. Zimmermann. 2009. "Ethnosizing Immigrants." *Journal of Economic Behavior & Organization* 69 (3): 274–87.
- Dronkers, Jaap, and Rolf van der Velden. 2013. "Positive but Also Negative Effects of Ethnic Diversity in Schools on Educational Performance? An Empirical Test Using PISA Data." In *Integration and Inequality in Educational Institutions*, edited by Michael Windzio, 71–98. Dordrecht: Springer Netherlands.
- Goldin, Claudia, and Lawrence F. Katz. 1999. "Human Capital and Social Capital: The Rise of Secondary Schooling in America, 1910-1940." *The Journal of Interdisciplinary History* 29 (4): 683–723.
- Goux, Dominique, and Éric Maurin. 2005. "Composition sociale du voisinage et échec scolaire: Une évaluation sur données françaises." *Revue économique* 56 (2): 349.
- Goux, Dominique, and Eric Maurin. 2007. "Close Neighbours Matter: Neighbourhood Effects on Early Performance at School." *Economic Journal* 117 (523): 1193–1215.
- Grogger, Jeff. 1996. "Does School Quality Explain the Recent Black/White Wage Trend?" *Journal of Labor Economics* 14 (2): 231–53.
- Hanushek, Eric. 1972. *Education and Race: An Analysis of the Educational Production Process*.
- Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2009. "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement." *Journal of Labor Economics* 27 (3): 349–83.
- Hémet, Camille, and Clément Malgouyres. 2018. "Diversity and Employment Prospects: Neighbors Matter!" *Journal of Human Resources* 53 (3): 825–58.
- Horowitz, Donald L. 2001. *The Deadly Ethnic Riot*. Berkeley: University of California Press.
- Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." w7867. Cambridge, MA: National Bureau of Economic Research.
- Izaguirre, Alejandro, and Laura Di Capua. 2020. 'Exploring Peer Effects in Education in Latin America and the Caribbean'. *Research in Economics* 74 (1): 73–86.
- Kain, John F., and Daniel M. O'Brien. 2000. "Black Suburbanization in Texas Metropolitan Areas and Its Impact on Student Achievement." Working Paper. Green Center for the Study of Science and Society, University of Texas at Dallas.
- Laitin, David D. 2000. "What Is a Language Community?" *American Journal of Political Science* 44 (1): 142–55.
- Lazear, Edward. 1998. "Diversity and Immigration." Cambridge, MA: National Bureau of Economic Research.
- Lazear, Edward P. 1999. "Culture and Language." *Journal of Political Economy* 107 (S6): S95–126.

- Maestri, Virginia. 2017. "Can Ethnic Diversity Have a Positive Effect on School Achievement?" *Education Economics* 25 (3): 290–303.
- Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60 (3): 531–42.
- Moffitt, Robert A. 2001. "Policy Interventions, Low-Level Equilibria And Social Interactions." In *Social Dynamics*, Durlauf SN, Young HP, 45–82. MIT Press.
- OECD, and European Union. 2015. *Indicators of Immigrant Integration 2015: Settling In*. OECD.
- O'Reilly, Charles A., Katherine Y. Williams, and Sigal Barsade. 1998. "Group Demography and Innovation: Does Diversity Help?" In *Research on Managing Groups and Teams, Vol. 1. Composition*, D. H Gruenfeld, 183–207. US: Elsevier Science/JAI Press.
- Ottaviano, Gianmarco I.P., and Giovanni Peri. 2006. "The Economic Value of Cultural Diversity: Evidence from US Cities." *Journal of Economic Geography* 6 (1): 9–44.
- Poterba, James M. 1997. "Demographic Structure and the Political Economy of Public Education." *Journal of Policy Analysis and Management* 16 (1): 48–66.
- Putnam, Robert D. 2007. "E Pluribus Unum: Diversity and Community in the Twenty-First Century The 2006 Johan Skytte Prize Lecture." *Scandinavian Political Studies* 30 (2): 137–74.
- Rivkin, Steven G. 2000. "School Desegregation, Academic Attainment, and Earnings." *Journal of Human Resources*, 35 (2): 333–46.
- Vigdor, Jacob L. 2004. "Community Composition and Collective Action: Analyzing Initial Mail Response to the 2000 Census." *The Review of Economics and Statistics* 86 (1): 303–12.

Table 1: Distribution of the origins of individuals, 2005-2017 (in %)

	Nationality	Country of Birth	Parents' origins
France	94.18	87.86	76.77
Southern Europe	1.36	2.16	6.27
Rest of Europe	1.01	1.72	4.01
Maghreb	1.63	4.62	8.52
Rest of Africa	0.65	1.46	2.09
Rest of the World	1.19	2.17	3.20
Observations	4,873,442	4,044,318	4,873,465

Note: Statistics are computed on the entire LFS population.

Table 2: Diversity in individuals' living environment

	Observations	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
<b>Aire diversity</b>							
Nationality	7,987	0.105	0.143	0.000	0.000	0.1	1
Country of Birth	6,658	0.175	0.183	0.000	0.033	0.275	0.782
Parents' Origins	7,987	0.337	0.205	0.000	0.162	0.505	0.813

Note: Statistics are computed on the sample of 15-year-olds.

Table 3: Individuals' characteristics and characteristics of the living environment

	Mean	Standard Deviation
<b>Individual characteristics</b>		
Boy	0.517	0.500
Held back a grade at 15	0.189	0.392
Grade repetition between 15 and 16	0.148	0.355
French nationality	0.974	0.160
<b>Characteristics of the other 15-year-olds living in the <i>aire</i></b>		
% repeated a grade	0.183	0.311
Observations		7,987

Note: Educational level is coded from de 0 (unqualified) to 7 (bachelor's degree and above)

Table 4: Grade repetition and diversity: summary of the results

	(1)	(2)	(3)	(4)	(5)	(6)
<b>1. Diversity by Nationality</b>						
	0.143***	0.132***	0.131***	0.101**	0.086**	-0.092
	(0.035)	(0.038)	(0.038)	(0.042)	(0.043)	(0.121)
<b>2. Diversity by Country of Birth</b>						
	0.126***	0.118***	0.118***	0.103**	0.088**	-0.010
	(0.028)	(0.032)	(0.032)	(0.037)	(0.039)	(0.134)
<b>3. Diversity by Parents' Origins</b>						
	0.080***	0.053**	0.053**	0.028	0.010	-0.007
	(0.022)	(0.023)	(0.023)	(0.029)	(0.030)	(0.082)
Individual	NO	YES	YES	YES	YES	YES
Year and quarter F.E.	NO	NO	NO	YES	YES	YES
Prop. held back a grade	NO	NO	YES	YES	YES	YES
<i>Région</i> F.E.	NO	NO	NO	YES	NO	NO
<i>Département</i>	NO	NO	NO	NO	YES	NO
<i>Secteur</i> F.E.	NO	NO	NO	NO	NO	YES

Notes: The dependent variable indicates grade repetition of an individual between 15 and 16 years of age. The value is 1 if the individual repeated a grade, and 0 otherwise. It is regressed for diversity by nationality in the first set of regressions (1), for diversity by country of birth and for parents' origins in the second (2) and third (3) sets of regressions respectively. Fractionalization indices are based on the 6 categories of the origin variables. Each column corresponds to a different specification. In Column (1), the dependant variable is regressed for the *aire*, without any other control. Column (2) controls for individual characteristics: origin group (6 categories), gender, socio-economic category of both parents. Column (3) = (2) + Proportion of individuals who repeated a grade at 15 years in the *aire* (excluding the individual). Column (4) = (3) + year and quarter fixed effects + *région* fixed effects. Column (5) = (4) but with *département* fixed effects. Column (6) = (5) but with *secteur* fixed effects. Extensive results of the regression of grade repetition with diversity by nationality can be found in Table A1. Standard errors clustered at the *aire* level are reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

## Appendix

Table A1: Grade Repetition and Diversity by Nationality

	(1)	(2)	(3)	(4)	(5)	(6)
<b>Diversity by Nationality</b>						
	0.143*** (0.035)	0.132*** (0.038)	0.131*** (0.038)	0.101** (0.042)	0.086** (0.043)	-0.092 (0.121)
Male		0.009 (0.008)	0.010 (0.008)	0.009 (0.008)	0.010 (0.008)	-0.006 (0.014)
<b>Nationality (Ref: French)</b>						
Southern European		-0.119*** (0.044)	-0.118*** (0.043)	-0.117*** (0.044)	-0.126*** (0.046)	-0.162 (0.118)
Other European		-0.008 (0.065)	-0.007 (0.065)	-0.005 (0.066)	-0.008 (0.065)	-0.020 (0.161)
North African		-0.091** (0.047)	-0.091* (0.047)	-0.088* (0.046)	-0.073 (0.047)	-0.133 (0.090)
Other African		0.129* (0.077)	0.129* (0.077)	0.131* (0.080)	0.122 (0.083)	0.145 (0.166)
Other nationality		-0.142*** (0.037)	-0.141*** (0.037)	-0.128*** (0.036)	-0.134*** (0.036)	-0.107 (0.087)
Observations	7,987	7,987	7,987	7,987	7,987	7,987
Adjusted R <sup>2</sup>	0.003	0.011	0.011	0.040	0.051	0.205

Notes: The dependent variable indicates grade repetition by an individual between 15 and 16 years of age. The value is 1 if the individual repeated a grade, and 0 otherwise. It is regressed for diversity by nationality based on the 6 categories of the nationality variable. Each column corresponds to a different specification. In Column (1), the dependant variable is regressed on *aire*, without any control variables. Column (2) controls for individual characteristics: origin group (6 categories), gender, socio-economic category of both parents. Column (3) = (2) + Proportion of individuals who repeated a grade at 15 years in the *aire* (excluding the individual). Column (4) = (3) + year and quarter fixed effects + *région* fixed effects. Column (5) = (4) but with *département* fixed effects. Column (6) = (5) but with *secteur* fixed effects. Standard errors clustered at the *aire* level are reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001



Table A2: Standard errors computed at the *département* level

	Dependent variable: grade repetition between 15 and 16 years		
	Nationality	Country of Birth	Parents' Origins
Diversity	-0.092 (0.130)	-0.010 (0.185)	-0.007 (0.103)

Notes: The dependent variable indicates grade repetition by an individual between 15 and 16 years of age. The value is 1 if the individual repeated a grade, and 0 otherwise. It is regressed for diversity index based on the 6 category nationality variable. Each regression controls for the full set of individual characteristics, year, quarter and *région* fixed effects, proportion of individuals from the *aire* that repeated the grade at 15 and characteristics of the *département*. Standard errors clustered at the *département* level are reported in parentheses. \* p<0.10, \*\*p<0.05, \*\*\* p<0.001

Table A3: Sample excluding individuals living in French overseas territories

	Dependent variable: grade repetition between 15 and 16 years		
	Nationality	Country of Birth	Parents' Origins
Diversity	-0.063 (0.140)	-0.010 (0.189)	0.027 (0.085)

Notes: The dependent variable indicates grade repetition by an individual between 15 and 16 years of age. The value is 1 if the individual repeats a grade, and 0 otherwise. It is regressed for a diversity index based on the 6 categories of the nationality variable. Each regression controls for the full set of individual characteristics, year, quarter and *région* fixed effects, proportion of individuals from the *aire* that repeated a grade at 15 and characteristics of the *département*. Standard errors clustered at the *aire* level are reported in parentheses. \* p<0.10, \*\*p<0.05, \*\*\* p<0.001

Table A4: Grade repetition before 16 and diversity: summary of the results

	(1)	(2)	(3)	(4)	(5)	(6)
<b>Diversity by Nationality</b>						
	0.167***	-0.033	-0.035	0.020	0.024	-0.094
	(0.042)	(0.044)	(0.044)	(0.049)	(0.050)	(0.183)
<b>Diversity by Country of Birth</b>						
	0.083**	-0.026	-0.026	0.019	0.021	-0.158
	(0.036)	(0.040)	(0.040)	(0.047)	(0.049)	(0.205)
<b>Diversity by Parents' Origins</b>						
	-0.006	0.010	0.011	0.043	0.050	-0.016
	(0.029)	(0.030)	(0.030)	(0.037)	(0.039)	(0.116)
Individual	NO	YES	YES	YES	YES	YES
Year and quarter F.E.	NO	NO	NO	YES	YES	YES
Prop. held back back a grade	NO	NO	YES	YES	YES	YES
<i>Région</i> F.E.	NO	NO	NO	YES	NO	NO
<i>Département</i>	NO	NO	NO	NO	YES	NO
<i>Secteur</i> F.E.	NO	NO	NO	NO	NO	YES

Notes: The dependent variable indicates grade repetition by an individual before 16 years of age. The value is 1 if the individual has repeated a grade, and 0 otherwise. It is regressed for diversity by nationality in the first set of regressions (1.), for diversity by country of birth and by parents' origins in the second (2.) and third (3.) sets of regressions respectively. Fractionalization indices are based on the 6 categories of the origin variables. Each column corresponds to a different specification. In Column (1), the dependant variable is regressed on *aire*, without any other control. Column (2) controls for individual characteristics: origin group (6 categories), gender, socio-economic category of both parents. Column (3) = (2) + Proportion of individuals who repeated a grade at 15 in the *aire* (excluding the individual). Column (4) = (3) + year and quarter fixed effects + *région* fixed effects. Column (5) = (4) but with *département* fixed effects. Column (6) = (5) but with *secteur* fixed effects. Standard errors clustered at the *aire* level are reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

Figure 1: Correlation between ethnic diversity and grade repetition between 15 and 16 years of age

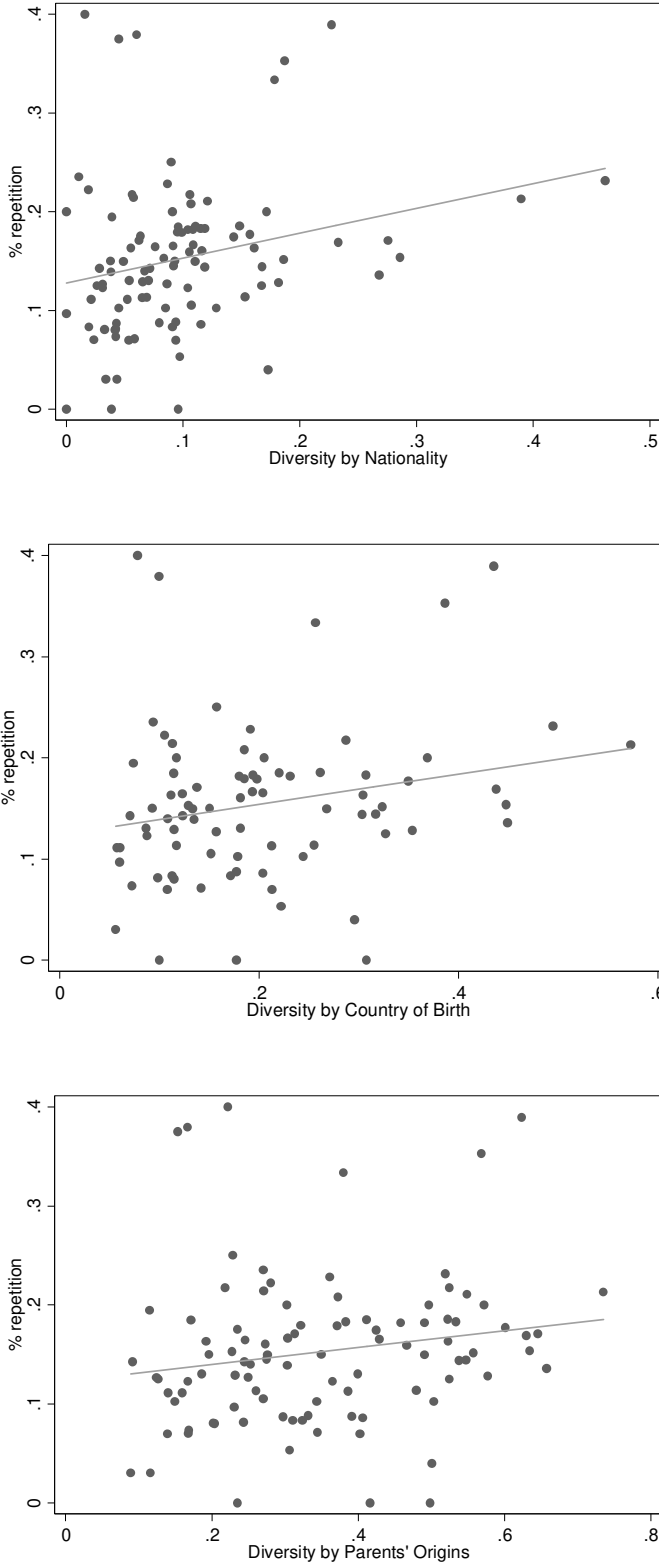
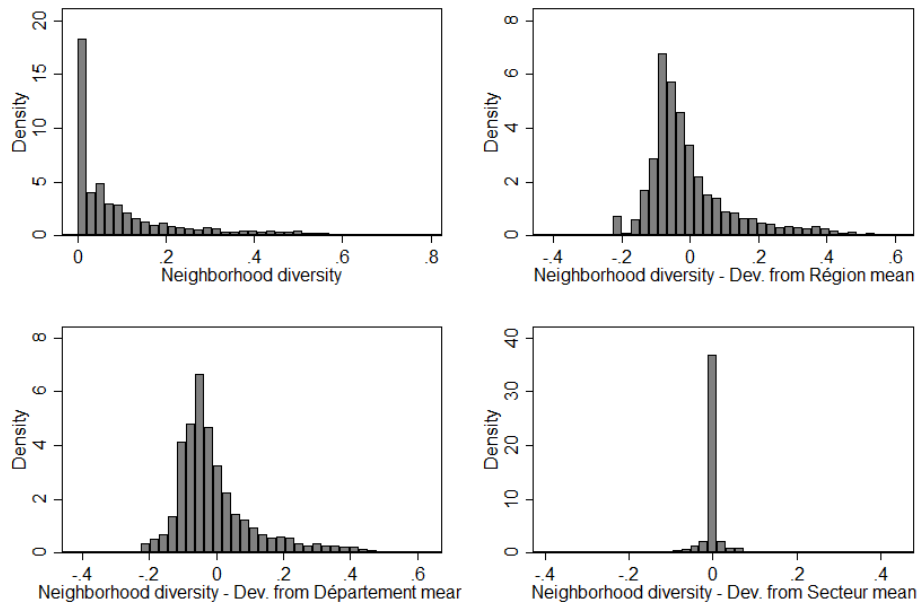
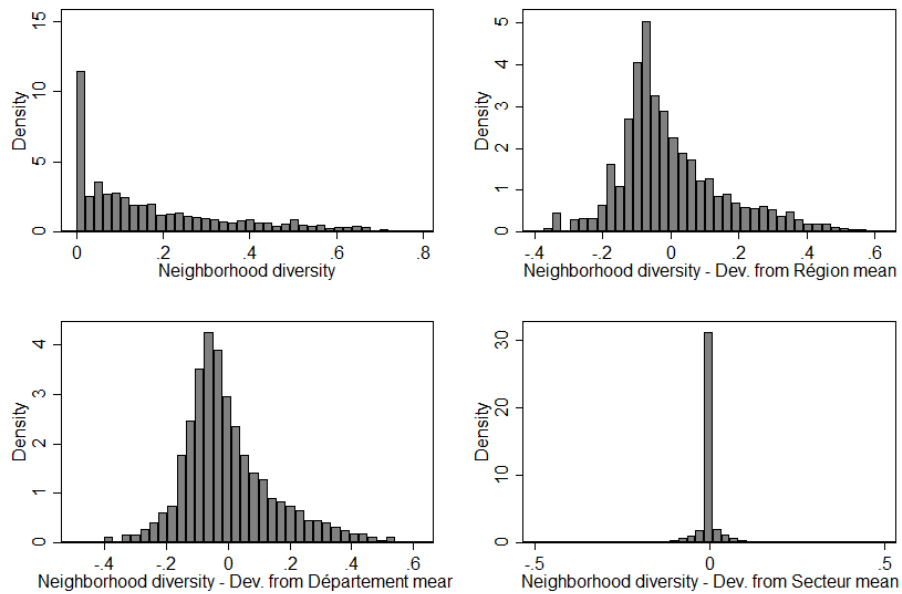


Figure 2: Aire Composition, Variation in Diversity levels.

### Diversity by Nationality



### Diversity by Country of Birth



# Diversity by Parents' Origins

