Neighbor Effects and Early Track Choices

Manon Garrouste[*] Camille Hémet[†]

July 2024

Abstract

The choice between vocational and academic education at the end of secondary school has important long-run effects, and is made at an age where peers' influence might be paramount. In this paper, we investigate the effect of *neighbors*' track choices on 9th graders choices at the end of lower secondary education, in Paris. This question is central to understand the extent to which residential segregation can reinforce social segregation across vocational and academic tracks. We rely on neighbors from the preceding cohort in order to bypass the reflection problem, and use within-catchment-area variation in distance between pairs of students to account for residential sorting. We use a pair-wise model that enables us to carefully study the role of distance between neighbors, and to perform detailed heterogeneity analysis. Our results suggest that close neighbors do influence track choices at the end of 9th grade, particularly for pupils pursuing a vocational track. This effect is driven by neighbors living in the same building, and is larger for pairs of boys and for pairs of pupils from low social background. Overall, our results suggest that neighbor effects tend to accentuate social segregation across high school tracks.

Keywords: Neighbor effects, Peer effects, Education, Early track choice, Segregation, Fixed-effects models

JEL Codes: I21, I24, C21, R23

[*]Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Économie Management, F-59000 Lille, France. Email: manon.garrouste@univ-lille.fr

[†]Université Paris 1 Panthéon-Sorbonne - Paris School of Economics, Paris, France. Email: camille.hemet@psemail.eu

Acknowledgements: We would like to thank Simon Briole, Guillaume Chappelle, Pierre-Philippe Combes, Gabrielle Fack, Laurent Gobillon, Nina Guyon, Elise Huillery, Miren Lafourcade, Eric Maurin and Mariona Segù for their useful comments and suggestions, as well as participants at various seminars and conferences. We are grateful to the statistical service of the French Ministry of Education (DEPP) for providing us with the data and in particular to Axelle Charpentier, Laurence Dauphin, Marine Guillerm, Soazig Jolivet, Alexis Lermite and Olivier Monso for their helpful insights. Funding from the French National Research Agency (ANR-18-CE22-0013-01, ANR-17-EURE-0001 and ANR-21-CE28-0004) is gratefully acknowledged.

1 Introduction

Selecting an educational track, and particularly choosing between vocational and general education, holds significant importance in shaping future educational and labor market outcomes (Silliman and Virtanen, 2022). Importantly, track choice is socially marked, with a notable over-representation of students from lower socio-economic backgrounds in vocational training, and a higher concentration of pupils pursuing vocational tracks in poorer neighborhoods (the maps in Figure A1 illustrate this point in the case of Paris). This critical decision generally occurs during adolescence, a period when peer influence is paramount (Albert et al., 2013; Brown and Larson, 2009; Deutscher, 2020). If peers influence each other regarding educational decisions, in particular peers living in the same neighborhood, this type of social interaction could contribute to reinforcing social segregation across educational tracks, thereby exacerbating existing disparities. Understanding whether individual educational decisions are influenced by neighbors' choices is therefore a key question that can have important policy implications. Although the literature provides evidence that neighbors influence the pursuit of higher education (see, e.g. recent work by Barrios-Fernández, 2022), research on the role of close neighbors in determining early track choice is still scarce.

In this paper, we aim to understand whether and to what extent interactions with other pupils living in the same neighborhood affect individual education decisions at the end of 9$^{\text{th}}$ grade. While it is established that social reproduction is an important dimension of educational track choices, especially in France (OECD, 2019), much less is known about the way in which these choices can be influenced by neighborhood peers' own education decisions. Using administrative data on the universe of pupils in Paris in 2017, geolocated at the address level, we show that close neighbors do influence track choices, this effect being driven by pupils enrolling in vocational training. Importantly, because we rely on a measure of distance between each pupil and any of their neighbors in a pair-wise model, we are able to precisely document the spatial scope of neighbor effects. We find that this effect rapidly fades with distance, and is actually driven by neighbors living in the same building. To the best of our knowledge, this is the first paper to precisely measure the spatial range of neighbors' effect, and to show how highly localized it is. This has important implications for the literature on neighbors' effects, as researchers might fail to identify such effects if the reference neighborhood is too large. Barrios-Fernández (2022) also provides convincing suggestive evidence that neighbors' effects are very local. However, while his estimates for various ranges of distance are obtained on different samples of students, we are able to estimate the precise effect of distance for each pupil in our sample.

Our study relies on a pair-wise model, where each ninth grader is paired with any of their neighbors, based on the Euclidian distance between each ninth grader and all other pupils from the preceding cohort (tenth graders) living in the surroundings. For computational reasons, we alternatively limit the set of neighbors to those living in the same catchment area, and to those living in a 1-kilometer radius around the reference pupil. We then measure the effect of distance between two neighbors on the likelihood that they pursue the same track, and interpret a significant coefficient of distance as indicating the existence of a neighbor effect. This model has three main advantages. First, it enables us to precisely measure the effect of distance between peers, considering distance in a continuous way, while the literature typically looks at neighbors within a given range. Therefore, we can precisely document the spatial decline of neighborhood peers' effects, which is an important methodological contribution to the literature on neighborhood effects. Second, by relying on pairs of pupils, we can study the heterogeneity of peer effects with respect to peer characteristics. In particular, our results confirm that neighbor effects are larger for pairs of boys and of pupils from disadvantaged backgrounds. Finally, using this pair-wise model enables us to abstract from the linear-in-means model which relies on strong assumptions on the structure of peer effects, as recently emphasized by Boucher et al. (2024), and to bypass the famous reflection problem pointed out by Manski (1993). Nonetheless, for the sake of comparison with the rest of the literature, we also resort to a more standard linear-in-means model, where we estimate the effect of neighbors' average track choice on individual choice within various ranges of distance. In any case, we avoid the reflection problem as we consider the effect of the track choices made by pupils from the previous cohort, whose decision was taken one year earlier and therefore cannot have been influenced by the choices made by the ninth graders under study.

A fundamental methodological challenge in identifying peer effects comes from the fact that peers share common unobservable characteristics and an environment, the effects of which (known as "correlated effects") are difficult to distinguish from social interaction effects (Manski, 1993). When looking at neighbor effects, the crucial issue is endogenous neighborhood selection, i.e. the fact that individuals decide where to live based on unobserved preferences. In the French context, this problem is particularly salient, as residential location largely determines which school will be attended, based on a catchment area system. It is therefore likely that socially and economically more advantaged families, who are more involved in their children's education and less constrained in their housing decisions, choose to reside in high-quality school catchment areas. In turn, this means that the catchment area of residence, and consequently the set of neighbors, is not exogenous to schooling decisions and preferences. To bypass this issue, we rely on a fixed-effect strategy, whereby the identifying variation comes from comparing local neighbors *within* catchment areas, which arguably constitute a relevant selection zone in the residential selection process in Paris. This method is similar in spirit to the one used in Bayer et al. (2008); Hémet and Malgouyres (2018); Solignac and Tô (2018), but is much more precise as we use individuals' exact geographic coordinates rather than the block or census tract where they reside.

Our results show that close neighbors do have an influence on track choices at the end of lower secondary education, after including catchment-area fixed effects accounting for residential selection across catchment areas. In particular, we find that this result is driven *i)* by pupils enrolling in a vocational track after ninth grade, and *ii)* by pupils living at the exact same address. More precisely, we find that living in the same building as a one-year-older student enrolled in vocational training increases the likelihood that a ninth grader enrolls in vocational education after ninth grade by 2 percentage points. By contrast, we do not find any significant effect of neighbors living further away. Our results thus show that neighbor effects are extremely local in nature, and that they rapidly disappear with distance. This raises the question of how to define a neighborhood, and reveals that neighborhood peer effects are likely to be diluted when considering relatively large neighborhoods. We check that these results are not driven by pupils living in social housing, which creates local clusters of low socio-economic status households. We also verify that our results are not due to the influence of older siblings (Altmejd et al., 2021), by excluding probable siblings from the sample. Finally, our results are robust to the inclusion of national test scores taken at the end of ninth grade, which account for ninth-graders' academic ability.

Crucially, our pair-wise design enables us to investigate the heterogeneity of neighbors' effects along various characteristics of the pairs. In other words, not only do we examine how the neigbhors' effect varies depending on neighbors' characteristics, but we are able to look at how this effect varies depending on the characteristics of the pair. In particular, we find that the effect of close neighbors is reinforced when both peers (the ninth grader and the one-year older neighbor) are from a low socio-economic background. Therefore, our results suggest that neighbor effects may accentuate social segregation across tracks. In addition, looking at the heterogeneity of the effect with respect to gender, we find that the effect of close neighbors is strongest when both pupils are boys.

Our paper contributes to the large and growing literature on peer effects in education. Several studies analyze the role of school peers on educational attainment (see for instance Epple and Romano, 2011; Sacerdote, 2011, for a review). Some of them use the US Add Health Survey to shed light on the role of peers on tests scores, drop out rates, or college enrollment (see Calvó-Armengol et al., 2009; Lin, 2010; Bifulco et al., 2011; Patacchini et al., 2017, among others). However, few studies look at the influence of *neighborhood* peers rather than *school* peers. In India, Helmers and Patnam (2014) show that a child's nearest neighbors' cognitive skills positively affect their own cognitive achievement. Using American data, Del Bello et al. (2015) find that only peers at school influence academic results and that peers from the same neighborhood do not. Relying on French data, Goux and Maurin (2007) show that close neighbors' characteristics influence test scores and grade repetition, but their setting does not allow them to report on education decisions. Few papers look directly at track choice, and, when they do, they document the role of school peers' characteristics (in particular educational attainment and social background) rather than peers' choices *per se* (Battiston et al., 2020; Bifulco et al., 2011; Gibbons and Telhaj, 2016; Jonsson and Mood, 2008; Landaud et al., 2020; Ly and Riegert, 2014). Instead, we want to take examine how individual decision-making is influenced by peers' choices.

This question is addressed by De Giorgi et al. (2010) and Barrios-Fernández (2022), but only in higher education. De Giorgi et al. (2010) show that Italian students are more likely to choose a given college major when many of their peers make the same choice. Barrios-Fernández (2022) finds that the closest neighbor significantly influences individual decisions to go to university in Chile. While they focus on enrollment decisions or major choice for undergraduate students or applicants, we will rather study track choices made by students at the end of lower secondary education (end of ninth grade). This is a pivotal moment in children's educational path, particularly in France, as students may decide, for the first time in their schooling, to embark on a vocational training rather than an academic track. The decision made at the end of ninth grade thus deeply influences future education and labor market outcomes. For this reason, it is crucial to understand how this decision is taken and how pupils from the same neighborhood may impact it.[1]

As such, our paper also adds to the existing research on neighborhood effects, which has demonstrated that growing up in a better neighborhood leads to better long-run eco-

[1] Several closely related papers also look at this type of decision between vocational and academic training, but they focus on the role of information or aspirations rather than neighborhood peer effects *per se* (Fricke et al., 2018; Goux et al., 2017; Guyon and Huillery, 2020).

nomic outcomes (Chetty et al., 2016; Chetty and Hendren, 2018). However, these studies do not clarify whether these outcomes are primarily influenced by better peers or improved environment. This paper specifically examines the role of peers by isolating the effects of neighbors while keeping neighborhood characteristics constant. Importantly, we further ensure that our results are not specifically driven by schoolmates either. Our findings indicate that neighbors do indeed play a role, independently of neighborhood and schoolmate effects.

The rest of this paper is organized as follows. We first provide some background by presenting the French education system and the associated administrative data in Section 2. Section 3 then describes our empirical model and estimation strategy. Our results are presented in Section 4. In Section 5, we propose a discussion on the underlying mechanisms. Section 6 concludes.

2 Institutional setting and data

2.1 Compulsory education and track choice in France

In France, education was compulsory from age 6 to age 16 up to 2019.[2] Children first attend primary school for five years, then they enter lower secondary education, spending four years in middle schools (*collèges*), before moving to upper secondary education in high schools (*lycées*) which takes two to three years. Primary and lower secondary education is the same for everyone. However, at the end of lower secondary education (in grade nine), pupils (usually aged 14 or 15) have to choose between differentiated school tracks. They can continue to higher secondary education, either through an academic track or a vocational track. They can also choose to repeat ninth grade, or to leave school if they have reached the age of 16. The majority of them enter an academic track, which prepares for a national academic examination (the "baccalauréat") that students take at the end of

[2]This was true at the time of the study. Since the 2019-2020 school year, education has been compulsory from the age of 3.

high school. They can then continue to higher education. The vocational track can take two forms: a short (two-year) vocational track preparing for a vocational certificate and direct entry into the labor market, or a longer (three-year) vocational track preparing for a vocational "baccalauréat" diploma, which gives access either to the labor market or to higher education.

The track choice procedure starts at the beginning of the second term of ninth grade (in January).[3] Each pupil and their family indicate their choice between an academic track, a short vocational track, or a long vocational track. At the end of the term, the teaching staff provides an answer in the form of a temporary proposal to each family. At the beginning of the third term, pupils make a list of final track choices, indicating a choice of high school for each track. They can rank multiple choices, but they are given priority in the public high school of their catchment area. At the end of the academic year, the middle school issues a final recommendation in terms of track choice. If it matches the family's own track choice, the pupil is allocated to that track. If not, the family meets with the school headmaster. If they disagree on the student's track choice, the family can appeal the middle school's decision. In this case, an appeal board makes the final decision. In any case, pupils can always choose to repeat ninth grade.

Importantly, track choice is socially marked, with a notable over-representation of students from lower socio-economic backgrounds in vocational training, and a higher concentration of pupils pursuing vocational tracks in poorer neighborhoods. Table 1 displays the socio-economic characteristics of all 9th graders enrolled in France in 2016, and how they are distributed across track choices (observed in 2017). Strikingly, pupils from a high socio-economic status background represent 26.5 % of all pupils enrolled in an academic track in 10th grade, but only 4.1 % and 6.5 % of pupils enrolled in short and long vocational tracks respectively. Conversely, pupils from a low socio-economic background represent 59.1 % of pupils entering short vocational training. As a result, pupils receiving means-tested scholarships are over-represented in vocational education. We can also note

[3]The academic year is organized in three terms between September and early July.

that vocational training has a higher proportion of pupils of foreign nationality.

Table 1 – Socioeconomic background of 9th graders by track choice in 10th grade

	Academic track	Vocational track	
		Short	Long
Socioeconomic status			
High SES	26.5	4.1	6.5
Medium SES	50.1	36.8	48.3
Low SES	23.4	59.1	45.2
Scholarship status			
No scholarship	83.4	67.8	69.7
Scholarship	16.6	32.2	30.3
Citizenship			
French	96.1	84.9	92.8
Not French	3.9	15.1	7.2

Source: MENJS DEPP, FAERE-Apprenants 2017. Sample: all 9^th graders enrolled in France in 2016.

The decision to enroll in vocational or academic education strongly influences mid- to long-term education and labor market outcomes. For instance, after seven years, the employment rate of young people who left the education system in 2010 after a short vocational certificate was 73% as opposed to 84% for those who obtained a vocational "baccalauréat", and 92% for those who completed a Master's degree. The median monthly wage was about 1,500 euros for young people with a vocational degree, 1,600 euros for those with an academic "baccalauréat", and 2 200 euros for those with a Master's degree (DEPP, 2020). Overall, in France, the academic track is generally perceived as the most academically prestigious, whereas the vocational track is often considered a low-status track or a default option for pupils who do not perform well enough to pursue the academic track.

2.2 Allocation to schools

Allocation to middle schools is based on a catchment area system. Each pupil has priority in the public middle school of their catchment area, which depends on their home address. Allocation to another public school may be granted by the local education authority (*académie*) only if there are remaining places once all pupils from the catchment area have been allocated. Allocation to another public school can be granted for the following reaso, in order of priority: medical reasons, scholarship, brother or sister already in the school, distance, specialized subject only available in the school. Furthermore, pupils can also choose to enroll at a private middle school, as these are not subject to the catchment area system. Because of this, the choice of middle school is largely determined through location decision (Boutchenik et al., 2020; Monarrez, 2023). When deciding where to live, some families carefully choose the corresponding catchment area school. In particular, socially and economically more advantaged families are expected to reside in high-quality schools' catchment areas, meaning that residing in a given catchment area, and therefore the set of neighbors, is not exogenous to schooling decisions and preferences.

Similarly, allocation to high schools is also based on a catchment area system, relying on a centralized allocation mechanism at the local authority level. Each student has priority in the public high school of their catchment area. Depending on the local education authority, past academic achievement and a favorable opinion from the middle school's or high school's headmaster may also be considered. For vocational high schools, the catchment area is typically larger than for general high schools and often corresponds to the entire local education authority area.

As there is some degree of freedom in school choice, it is important to note that not all students are enrolled at their catchment area school. In particular, when entering middle school, about 10% of students go to a public school other than their catchment area middle school, and about 20% go to a private school (Boutchenik et al., 2020). This situation may introduce additional selection, with school choice coming on top of endogenous location decision. As for location decisions, socially and economically more advantaged families are expected to choose high-quality schools, potentially outside of their catchment area. To account for this, we will carefully control for students' school choices in our estimations.

2.3 Data and descriptive statistics

To analyze the effect of neighbors on individual schooling decisions, we rely on geocoded administrative data on the universe of pupils enrolled in secondary education in France in 2017, provided by the statistical service of the French Ministry of Education. For each pupil, we observe the school and grade in which they are enrolled in years t and $t-1$. We focus on two samples. First, we consider pupils enrolled in the last year of lower secondary education (ninth grade) in 2017, and we observe their grade and track in year $t+1$. Second, we consider the sample of pupils enrolled in the first year of upper secondary education (tenth grade) in 2017, and we observe their track in year t. For each pupil, we know the exact location of the home address, given by geographic coordinates. We also observe each pupil's gender, country of birth, parents' occupations, and whether or not the pupil receives a public income-based scholarship.

In addition, each pupil needs to be assigned to the catchment area – and thus the public middle school – corresponding to their home address. Unfortunately, there is no nationwide GIS documenting the borders of each catchment area in France. In this study, we focus on the municipality of Paris, for which we were able to obtain the geocoded data of every public middle school's catchment area. We therefore restrict our sample to pupils living in Paris and who were enrolled in ninth grade in 2016 or 2017, as we can thus identify the catchment area of their public middle school.[4]

In total, we observe 17,708 ninth graders, and 19,614 tenth graders in 2017, living in 111 middle school catchment areas in Paris. On average, Parisian pupils are socially more advantaged than pupils living in the rest of France. About 45% come from a high socio-economic background in the sense than their reference parent is an executive or a highly-qualified worker, 34% come from a medium socio-economic background, and 21% from a low socioeconomic background (with parents who are blue-collar workers or unemployed). Students benefiting from an income-based scholarship represent about

[4]The catchment area is defined as the one that was in use at the time when the ninth graders entered middle school.

20% of the sample. Almost 90% of children are French-born. About 11% of pupils live in a deprived neighborhood in the sense that it benefits from a special urban policy scheme.[5]

Although most pupils were enrolled in a public middle school in ninth grade (43% in their catchment-area school and 28% in another public school), a large proportion of them (29%) attended a private school. After ninth grade, about 82% entered an academic track and 15% a vocational track. Among those who continued to a vocational track, the vast majority followed a long, three-year track (see Table 2 for descriptive statistics of the student sample).

Table 2 – Socioeconomic characteristics of pupils in the sample

	9th graders		10th graders	
Socioeconomic status				
High SES	7,987	45.1	8,344	42.5
Medium SES	6,087	34.4	6,768	34.5
Low SES	3,628	20.5	4,502	23.0
Gender				
Female	8,975	50.7	9,617	49.0
Male	8,733	49.3	9,997	51.0
Scholarship status				
No scholarship	14,153	79.9	16,429	83.8
Scholarship	3,555	20.1	3,185	16.2
Citizenship				
French	15,765	89.0	17,545	89.5
Not French	1,943	11.0	2,069	10.5
Middle school				
Public	12,658	71.5	14,198	72.4
Private	5,050	28.5	5,416	27.6
Track choice				
Repetition	494	2.8		
General track	14,645	82.7	15,517	79.1
Vocational track	2,569	14.5	4,097	20.9
Total	17,708	100.0	19,614	100.0

Source: MENJS DEPP, FAERE-Apprenants 2017, Paris

Note: High SES = children of executives or highly-qualified workers, Medium SES = children of workers with intermediate occupations, or white-collar workers, Low SES = children of blue-collar workers or unemployed. Repetition refers to students repeating ninth grade and is therefore not observed for the 10th grader cohort.

[5]These neighborhoods are part of the priority areas for urban policy *Quartiers prioritaires de la politique de la ville*. They are defined as economically disadvantaged neighborhoods, based on residents' median income.

3 Empirical strategy

3.1 Conceptual framework

To bypass the issue of endogenous location, we use the within-catchment-area variation in distance between neighbors. To do so, we construct all pairs of pupils living in the same catchment area. Every ninth grader is paired with all peers from the preceding cohort living in the same catchment area.

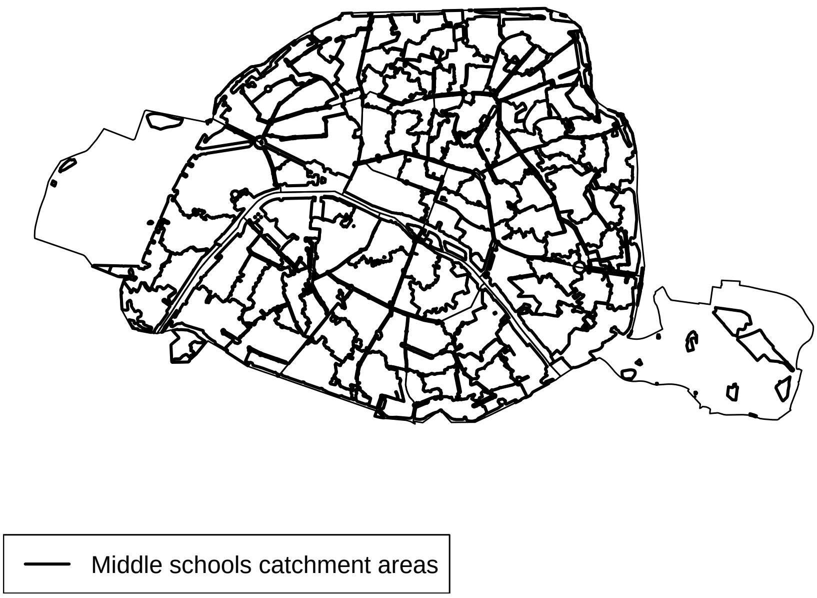
The distance to neighbors is then defined as the Euclidean distance to all 10th graders living in the same catchment area. Figure 1 represents middle schools catchment areas in the municipality of Paris with a close-up of the 14th *arrondissement*. Each catchment area corresponds to one public middle school (black squares). The municipality of Paris includes 111 catchment areas, each of them holding an average of 159 ninth graders. Blue dots represent a (fictive) sample of ninth graders in each catchment area, grey diamonds represent a (fictive) sample of tenth graders, and Figure 2 represents the two-by-two distance between every ninth grader and every tenth grader in a given catchment area. Note that this creates a perfectly dense network of neighbors within each catchment area.

Although some families carefully choose where to live based on the quality of the catchment area middle school, we can assume that, within a given catchment area, they do not choose a particular location depending on the location of other tenth graders. In the following, we will therefore use the fact that within-catchment-area variation in two-by-two distance between neighbors is exogenous to unobserved schooling preferences.[6]

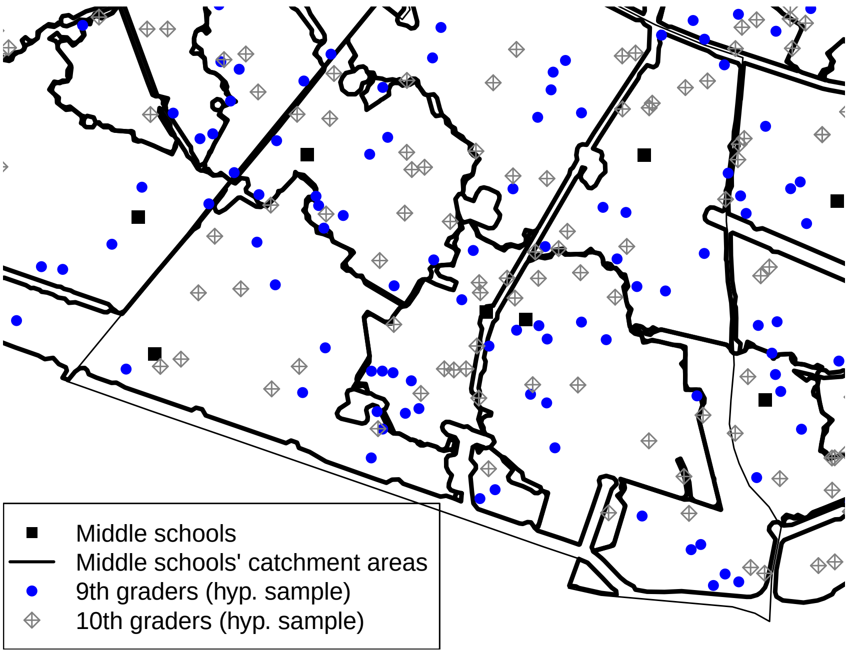
[6]This assumption will be discussed and tested in Section **??**.

Figure 1 – Middle schools catchment areas

(a) in Paris



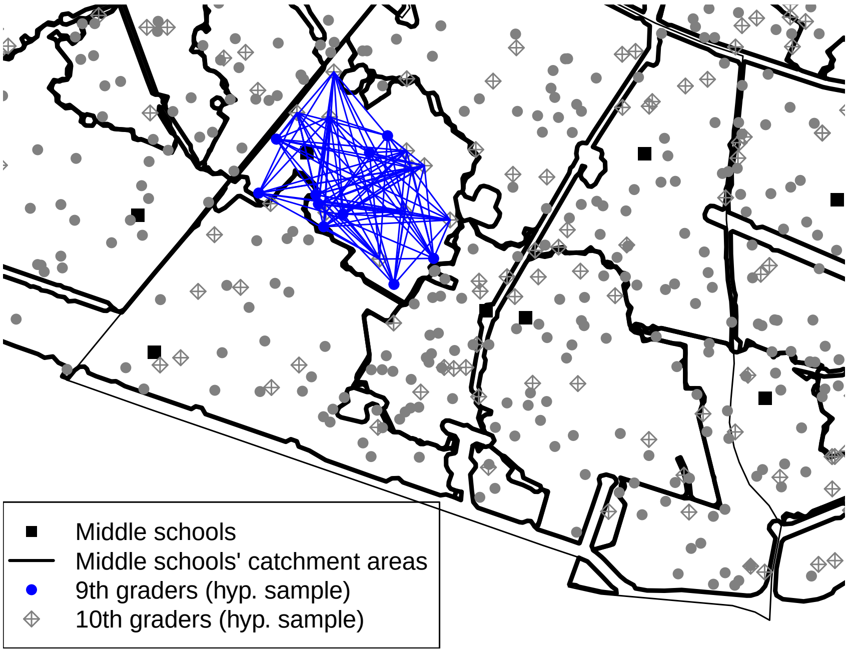
(b) in the 14th *arrondissement* of Paris



Source: Insee, IGN, Direction des Affaires Scolaires - Ville de Paris

Note: Thick lines represent the boundaries of middle schools catchment areas in the municipality of Paris. Black squares represent middle schools. Blue dots represent a (fictive) sample of ninth graders. Grey diamonds represent a (fictive) sample of tenth graders.

Figure 2 – Middle schools catchment areas



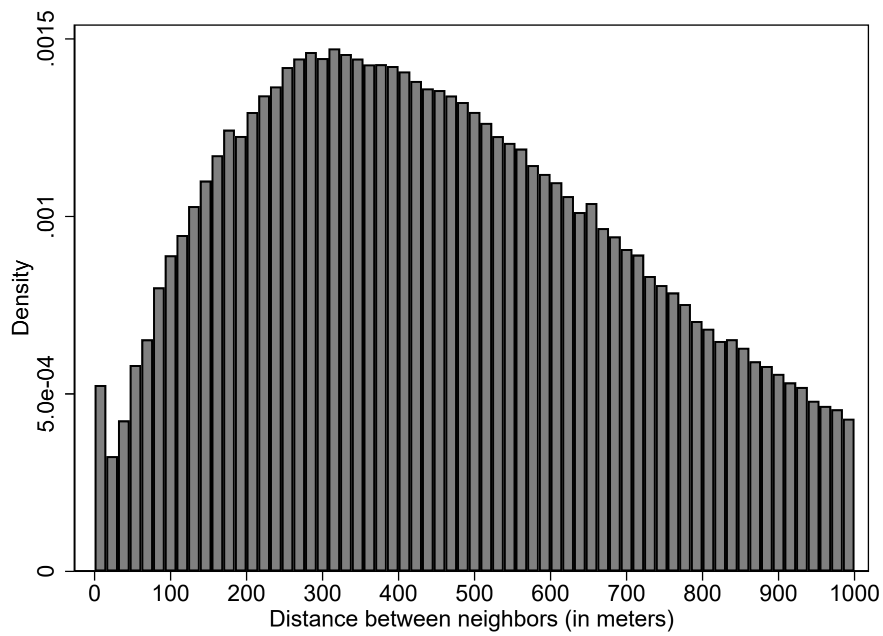
Source: Insee, IGN, Direction des Affaires Scolaires - Ville de Paris

Note: Thick lines represent the boundaries of middle schools catchment areas in the municipality of Paris. Black squares represent middle schools. Blue dots represent a (fictive) sample of ninth graders. Grey diamonds represent a (fictive) sample of tenth graders. Blue lines represent distances between ninth graders and tenth graders within one catchment area.

3.2 Sample description

We observe more than 3,200,000 pairs of pupils living in the same catchment area (see Tables 3 and 4 for descriptive statistics of the pair sample). Within a given catchment area, the average distance between a ninth grader and a tenth grader is about 500 meters (see Figure 3). Note that we censored distance to be less than one kilometer for computational reasons and because almost 90% of pairs live less than one kilometer away from each other. Importantly, the mass point at zero illustrates that in Paris many pairs live at the same address, corresponding to the exact same geographical coordinates. More than 21,000 pairs of pupils live at distance zero, which corresponds to about 8,400 ninth graders (i.e. almost 48% of the sample of ninth graders) with at least one same-address tenth-grader neighbor. For them, the number of same-address tenth-grader neighbors varies between 1 and 22, with a mean of 2.5.

Figure 3 – Distribution of distance between neighbors



Source: MENJS DEPP, FAERE-Apprenants 2017

Table 3 – Description of pairs of ninth- and– tenth-graders

	Percentage
Socioeconomic status	
i High SES	
j High SES	23.7
j Medium SES	14.0
j Low SES	8.8
i Medium SES	
j High SES	13.4
j Medium SES	12.0
j Low SES	8.3
i Low SES	
j High SES	7.5
j Medium SES	7.2
j Low SES	5.3
Gender	
i Female	
j Female	24.9
j Male	26.0
i Male	
j Female	24.0
j Male	25.1
Scholarship status	
i No scholarship	
j No scholarship	70.0
j Scholarship	10.9
i Scholarship	
j No scholarship	15.0
j Scholarship	4.1
Citizenship	
i French	
j French	79.6
j Not French	9.3
i Not French	
j French	9.8
j Not French	1.3
Observations	3272473

Source: MENJS DEPP, FAERE-Apprenants 2017, Paris

Note: The unit of observation is a pair of individuals living in the same catchment area. i refers to 9^{th} graders and j refers to 10^{th} graders.

Table 4 – Track choices of pairs of ninth- and– tenth-graders

	Percentage
Repetition-General	2.2
Repetition-Vocational	0.6
General-General	67.7
General-Vocational	15.7
Vocational-General	10.4
Vocational-Vocational	3.4
Observations	3272473

Source: MENJS DEPP, FAERE-Apprenants 2017, Paris

Note: The unit of observation is a pair of individuals living in the same catchment area. Tracks are reported using the following order: 9th grader-10th grader.

3.3 Empirical model

For any pair of ninth grader i and tenth grader j living in the same catchment area C, let us denote by Y_{ij} a dummy variable which equals one if i pursues the same track as j after ninth grade and zero otherwise. $dist_{ij}$ is the distance (in meters) between i's and j's places of residence. Let us consider the following equation:

$$Y_{ij} = \beta dist_{ij} + \zeta dist_{ij}^2 + X_i'\gamma_1 + X_j'\gamma_2 + X_{ij}'\gamma_3 + \mu_C + \varepsilon_{ij} \quad (1)$$

where X_i, X_j, and X_{ij} are a set of individual characteristics: gender and socio-economic background (parents' occupations, scholarship status, and whether or not they are French-born), as well as the pair's school choices (same school or not and type of school, i.e. catchment area public school, another public school, or a private school).

3.3.1 Mitigating residence- and school-based sorting: a fixed-effect approach

As families choose the catchment area where they want to live, $dist_{ij}$ is supposed to be endogenous in general. For instance, if socially more advantaged families choose to live in the catchment area of better-quality schools, then we expect $dist_{ij}$ to be negatively correlated with the error term (distant neighbors' unobserved preferences regarding education choices are less similar than close neighbors'). A naive estimation would then overestimate the effect of close neighbors' track choices. Therefore, our estimation strategy consists in including catchment area fixed effects μ_C in the equation. Under the assumption that families do not choose exactly where to live in a particular catchment area, the within-catchment-area variation in distance can then be used to recover a causal effect of neighbors' choices on own choice. In other words, we assume that the distribution of students within catchment areas is as good as random once we account for their choice of catchment area. To test this assumption, we will check that there is no spatial auto-correlation within catchment areas regarding students' observed characteristics (see Section **??**).

This strategy is similar to the one used in the seminal paper by Bayer et al. (2008), and more recently by Hémet and Malgouyres (2018) and Solignac and Tô (2018) for France, but uses students' exact geographic coordinates instead of individuals' census tracts. We are thus able to analyze the effect of distance to neighbors very precisely rather than using residence in the same census tract as proxy. As we observe the entire distribution of two-by-two distances, we can analyze the distance at which neighbor effects begin to fade. To test whether only very close neighbors have an influence, let us consider the following model, where distance to neighbors is divided into concentric rings of 100 meters around each student (see Figure 4). For all students i and j such that $100 \times k \leq dist_{ij} < 100 \times (k+2)$, let us consider the model:

$$Y_{ij} = \beta_k \mathbb{1}_{100 \times k \leq dist_{ij} < 100 \times (k+1)} + X_i' \gamma_1 + X_j' \gamma_2 + X_{ij}' \gamma_3 + \mu_C + \varepsilon_{ij} \tag{2}$$

$\forall k = 0, \ldots 5$.

The parameters of interest, β_k, measure the additional probability of choosing the same track as a close neighbor rather than someone living further away, controlling for catchment area selection. For instance, β_0 corresponds to the effect of living less than 100 meters away within a given catchment area rather than 100 to 200 meters away on a pair's track choice.

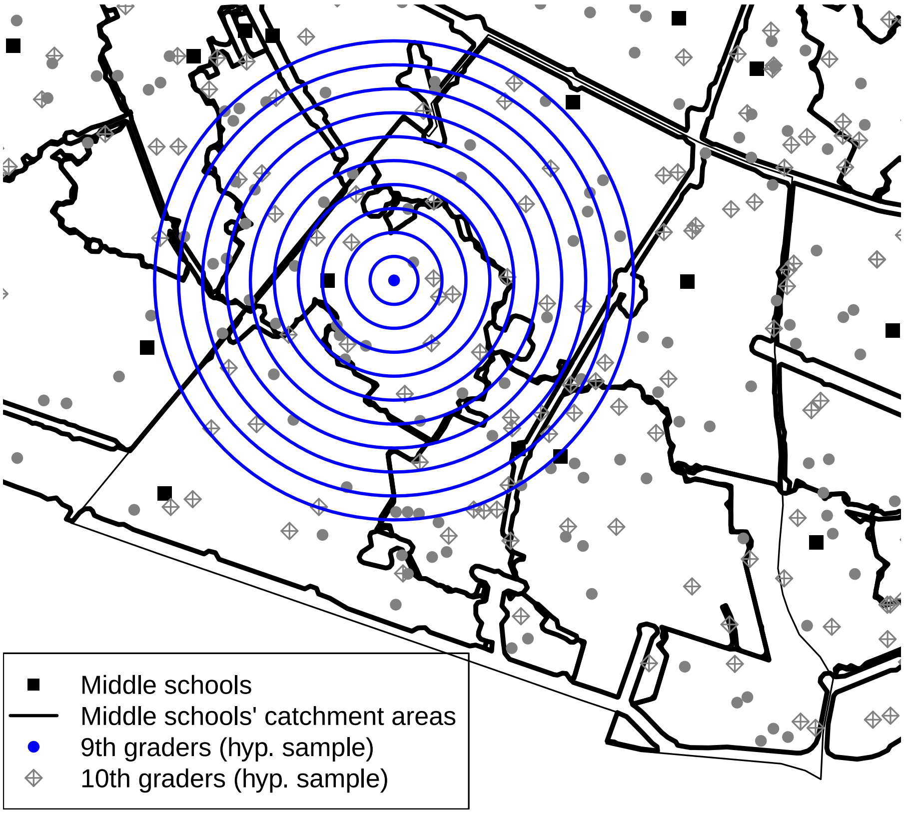
The magnitude of these parameters therefore provides a test for the importance of residence-based peer influence in determining the type of track that students choose. Of course, there might be several mechanisms explaining such neighborhood peer effects. The most direct one would be that pupils are (more) influenced by their neighbors (than by their non-neighbors) because they interact more frequently, for instance going home from school together, or spending more time together during evenings and weekends (again, whether or not they were enrolled at the same middle school is included in the model). Another, more indirect channel would be through parents: it could be that parents who are close neighbors interact more frequently with each other than parents who live further away, and therefore have more opportunities to exchange information on track choices and discuss their plans for their children, thus influencing each other and eventually affecting students' actual choices.[7] Given the data at hand, disentangling the precise mechanisms through which neighbors affect students' track choice is out of the scope of this paper. However, the heterogeneity analysis can give us some insights. For instance, if same-sex peers interact more frequently or easily, then one could expect larger neighbor effects for pairs of girls or pairs of boys than for opposite-sex pairs. Potential mechanisms are investigated in Section 5.

3.3.2 Testing the no-spatial-autocorrelation assumption in *within* estimations

In order to test the identifying assumption of equations (1) and (2), we conduct a formal statistical test that consists in verifying that neighbors are not distributed within catchment areas in a way that would create spatial auto-correlation with respect to certain observed characteristics. In particular, even within catchment areas, individuals may be clustered on the basis of their social background. For instance, this may be the case in a context of residential segregation (which is particularly high in the Paris municipality) whith wealthy and poor families living in separate neighborhoods, or if many low-socioeconomic status families live close to each other in social housing. To test for this, we run join-count

[7] Note that these two channels are not mutually exclusive, and can even be complementary.

Figure 4 – Effect of distance to neighbors (Equation 2)



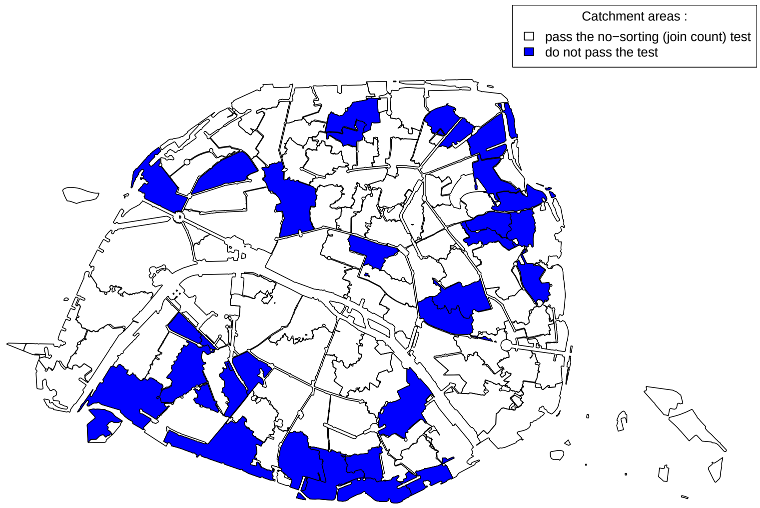
Notes: For each student i residing at the center of the rings, Equation (2) considers sequential sub-samples of all neighbors living less than 100 meters away, 100 to 200 meters away, etc. Coefficient β_0 corresponds to the effect of living at distance zero rather than less than 100 meters away. Coefficient β_1 corresponds to the effect of living less than 100 meters away rather than 100 to 200 meters away, etc.

tests, which consist in comparing the theoretical probability that two students share the same characteristic, under the assumption of absence of spatial auto-correlation, with the observed probability. In other words, this test assesses whether pairs with a common characteristic (for instance both coming from the same social background) are closer to each other than would be expected by chance, taking into account the proportion of students from each social background in the catchment area.[8]

We perform this test on the sample of ninth graders. The results show that the null

[8]Social backgrounds are defined as parents' occupation status, defined by the statistical service of the French Ministry of Education as a four-category variable. Very high socio-economic status students are children of executives, High SES students are children of parents with intermediate occupations, Medium SES students are children of white-collar workers, and Low SES students are children of blue-collar workers or the unemployed.

Figure 5 – Map of the catchment areas that do not pass the no-spatial-autocorrelation test



Source: MENJS DEPP, FAERE-Apprenants 2017, Direction des Affaires Scolaires - Ville de Paris

hypothesis of an absence of spatial auto-correlation in the pairs' social background is rejected in 30 catchment areas out of 111 (representing about 28% of the sample of student pairs) with a 5% significance level. Note that some spatial auto-correlation due to randomness is of course likely to happen (in about 5% of cases). For the sake of comparison, we also test for spatial auto-correlation in the pairs' gender composition (a dimension in which there cannot be spatial sorting), which leads to a rejection of the null hypothesis of an absence of spatial auto-correlation in 5 catchment areas out of 111. The 30 catchment areas that do not pass the sorting test are mapped in Figure 5. A quick look at this map suggests that these areas correspond to a large extent to neighborhoods with a high concentration of public housing buildings. It is therefore mechanical to observe a significant clustering of pupils from the same socio-economic background in these areas.

In what follows, we first estimate the model on the entire sample, and then systematically exclude these 30 catchment areas to make sure that the estimation results are not driven by the potential clustering of families within catchment areas based on socioeconomic characteristics.

4 Results

4.1 Main results

Model with quadratic distance term. Table 5 presents the results obtained when estimating equation (1) by ordinary least squares with catchment-area fixed effects including all Parisian catchment areas. The first column corresponds to the effect of the distance between neighbors (measured in 100 meters) on the probability of choosing the same track after ninth grade, controlling for the choice of catchment area. The remaining columns report the results of estimating the equation separately for each type of track (grade repetition, academic, or vocational).

Interestingly, the only significant effect of distance is on the probability that i and j both choose the vocational track. The results suggest that a one-hundred-meter increase in the distance between neighbors corresponds to a 0.2 percentage point increase in the probability that i and j choose the vocational track. Compared to the results obtained when estimating the same model without catchment-area fixed effects, the coefficient is slightly smaller (see Table B1 in Appendix B), suggesting that the effect of distance between neighbors would be over-estimated if we did not control for sorting across catchment areas.

Our estimation strategy relies on the assumption that the distribution of students within catchment areas is as good as random once we account for their choice of a catchment area. However, this assumption may not hold if households are spatially clustered on the basis of characteristics that are correlated to their unobserved preferences regarding education trajectories. In this case, we would risk wrongly attributing to neighbors' influence something that is actually due to correlated effects.

Residential segregation, poverty clusters, or social housing may all be reasons why this assumption would be violated, especially in the Paris municipality. To make sure that the estimation results are not driven by the residential sorting of families within catchment areas, we exclude from the analysis the 30 catchment areas for which the no-spatial-sorting test does not hold.

Table 6 shows that the results are not due to spatial sorting on the basis of the pairs' social composition, as the results are robust to the exclusion of the catchment areas that do not pass the no-spatial-auto-correlation test discussed in Section 3.3.2. The estimate of the effect of distance on the probability that i and j choose the vocational track is smaller (0.16 percentage point) but still very significant. The coefficient corresponding to the square of distance suggests that neighbor effects are likely non-linear in distance between students.[9] In the rest of the paper, all estimates are based on the sample excluding the 30 catchment areas that fail to pass the sorting test.

Order of magnitude of the results. Although statistically significant, the estimated effects of neighbors on track choices may seem economically small. Moreover, the results are not easily comparable to other studies on the effects of neighbors on education outcomes (Goux and Maurin, 2007; Del Bello et al., 2015), as these studies look at schooling performance and not schooling decisions, and use linear-in-means models at the individual level, whereas we use models at the pair level. We can nevertheless compare neighbor effects with the size of coefficients of other covariates in the model, such as students' parents' occupation. Of course, we do expect neighborhood peer effects to be much smaller than the effect of students' own social background. For the sake of comparison, the 2-percentage point increase in the probability of entering a vocational track after ninth grade when a same-building older neighbor has chosen a vocational track is about ten times smaller than the increase in the same probability when coming from a high socioeconomic background.

Model with concentric rings of distance. Considering linear or quadratic effects of distance between neighbors may be inaccurate as neighborhood peer effects are presumably non-linear and mixing very close and very distant neighbors is likely misleading.

[9]The effect of distance between students fades away or changes sign after $\frac{0.0016}{2\times-0.0001} = 8$ hundred meters.

To further understand neighborhood peer effects on track choices, let us present the results obtained when estimating Equation 2. Figure 6 plots the coefficients associated with estimating the β_k parameters. When focusing on neighbors within a two-hundred-meter range, who are thus close to one another, the results suggest that older neighbors have no effect on track choice when considering all tracks. However, very close neighbors seem to influence younger neighbors when they are enrolled in a vocational track. Figure 6 (c) illustrates that for pairs living at distance zero, i.e. in the same building or block of buildings, the probability that i pursues an academic track after ninth grade when j has chosen a vocational track is lower by 2 percentage points compared to pairs of neighbors living between 0 and 100 meters away. Symmetrically, for pairs living at distance zero, the probability that i chooses a vocational track after ninth grade when j is pursuing a vocational track is higher by 2 percentage points, compared to pairs of neighbors living 0 to 100 meters away (Figure 6 (e)). Older neighbors living further away do not influence their younger neighbors.

Table 5 – Neighborhood peer effects on track choices at the end of ninth grade - Equation (1) - All catchment areas

	i's track-j's track						
	Same track	Rep^{ion}-Gen^{al}	Rep^{ion}-Voc^{al}	Gen^{al}-Gen^{al}	Gen^{al}-Voc^{al}	Voc^{al}-Gen^{al}	Voc^{al}-Voc^{al}
Dist. btw neighbors (100 meters)	-0.0005	0.0006*	-0.0001	0.0015	-0.0002	0.0001	-0.0020***
	(0.0010)	(0.0003)	(0.0001)	(0.0012)	(0.0010)	(0.0007)	(0.0005)
Dist.^{2}	0.0000	-0.0000	0.0000	-0.0001	-0.0000	0.0000	0.0001***
	(0.0001)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Intercept	0.7422***	0.0144***	0.0031***	0.7069***	0.1258***	0.1146***	0.0353***
	(0.0062)	(0.0022)	(0.0008)	(0.0082)	(0.0053)	(0.0053)	(0.0028)
Nbr obs	3,271,014	3,271,014	3,271,014	3,271,014	3,271,014	3,271,014	3,271,014
Nbr clusters	111	111	111	111	111	111	111

Source: MENJS DEPP, FAERE-Apprenants 2017

Note: * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$). The unit of observation is a pair of individuals living in the same catchment area. Standard errors in parentheses are clustered at the catchment area level. All estimations include catchment-area fixed effects, individual characteristics, and characteristics of the pair (gender, parents' occupations, scholarship status, country of birth, and middle school choice).

Table 6 – Neighborhood peer effects on track choices at the end of ninth grade - Equation (1) - Excluding the catchment areas that do not pass the no spatial auto-correlation test

	i's track-j's track						
	Same track	Rep^{ion}-Gen^{al}	Rep^{ion}-Voc^{al}	Gen^{al}-Gen^{al}	Gen^{al}-Voc^{al}	Voc^{al}-Gen^{al}	Voc^{al}-Voc^{al}
Dist. btw neighbors (100 meters)	-0.0003	0.0004	-0.0001	0.0013	-0.0006	0.0007	-0.0016***
	(0.0011)	(0.0004)	(0.0002)	(0.0014)	(0.0012)	(0.0008)	(0.0005)
$Dist.^2$	0.0000	-0.0000	0.0000	-0.0001	0.0000	-0.0000	0.0001*
	(0.0001)	(0.0000)	(0.0000)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Intercept	0.7465***	0.0175***	0.0041***	0.7139***	0.1252***	0.1069***	0.0326***
	(0.0068)	(0.0024)	(0.0008)	(0.0090)	(0.0064)	(0.0064)	(0.0034)
Nbr obs	2,318,373	2,318,373	2,318,373	2,318,373	2,318,373	2,318,373	2,318,373
Nbr clusters	81	81	81	81	81	81	81

Source: MENJS DEPP, FAERE-Apprenants 2017

Note: * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$). The unit of observation is a pair of individuals living in the same catchment area. Standard errors in parentheses are clustered at the catchment area level. All estimations include catchment-area fixed effects, individual characteristics, and characteristics of the pair (gender, parents' occupations, scholarship status, country of birth, and middle school choice).

Take-away. To sum up, close neighbors do influence their younger neighbors' track choices at the end of ninth grade, when it comes to enrolling into vocational track. Importantly, the effect is very local: it is detected only for neighbors living at the same address (relative to other neighbors living less than 100 meters away), and immediately fades away with distance.

4.2 Robustness checks

Controlling for academic achievement. At the end of middle school, students take a standardized national exam called the *Brevet*, which consists of a written exam in four subjects (French, Mathematics, History-Geography, and Sciences), and an oral presentation. Students' final score also includes their average in all subjects as evaluated by their teachers throughout the academic year.

Passing the *Brevet* exam is not a prerequisite for admission to higher secondary education. However, almost all students take it and their score is positively correlated to their probability of pursuing an academic track. In order to check that our neighbor effect estimates are not driven by students' academic achievement, we include ninth graders' *Brevet* exam scores as a covariate in the estimation.[10] Figure 7 compares the previous estimation results (baseline estimates) with the results obtained when controlling for academic achievement (second estimate in the graph). Although the estimated coefficient corresponding to the effect of very close neighbors (living in the same building) is smaller, it is still significantly different from zero and not significantly different from the previous estimate.

Social housing. Some very close neighbors, living in the same building or block of buildings, may very well be living in large social housing estates. These students are likely to share unobserved characteristics that could explain the estimated correlation between track choices. To exclude any effects driven by social housing, we match our data with

[10] Unfortunately, the *Brevet* exam score is not available for the cohort of 10^{th} graders.

the national geocoded dataset on social housing. Every student's place of residence is matched with the national registry of social housing (*Répertoire des logements locatifs des bailleurs sociaux*) to determine which ones live in such housing. Whether i, j, or both of them live in social housing is then included as additional covariates in the estimation (i.e. in addition to the *Brevet* exam score).

Figure 7 shows that our results are robust to controlling for social housing as the estimate (third estimate in the graph), although smaller, is still significant and not significantly different from the previous one.

Siblings. One-year-apart students living in the same building may very well be siblings. As there is no identifier for family in the data, we cannot observe real siblings. However, we define probable siblings as any two students living in the same place and sharing the same parental characteristics (parents' occupations and scholarship status). As a robustness check, we then exclude all probable siblings from the estimation sample. This methode, which is more conservative than excluding actual siblings, as two students living in the same place and sharing the same parental characteristics may not be siblings, leads to the exclusion of 835 pairs.

Figure 7 confirms that our results are not driven by siblings and are robust to excluding probable siblings as the estimate remains unchanged (fourth estimate in the graph) and, although slightly less precise, still significant and not significantly different from the previous estimate.

In the rest of the paper, all estimates are obtained controlling for both academic achievement, social housing status, and excluding pairs of probable siblings from the sample.

Linear-in-means model. The literature on peer effects often uses linear-in-means models, in which an individual's outcome depends on the mean outcome of their peers. The pairwise model is our preferred specification because it is the only way to measure the effect of the exact distance between neighbors and to measure heterogeneous effects depending on the pairs' social composition. However, to check that our results are not due to this pairwise specification, as well as to make them more comparable to the existing literature, let us estimate a linear-in-means model in which a ninth grader's probability of enrolling in the academic (respectively vocational) track after middle school is a function of the proportion of their neighbors from the preceding cohort in each bin of distance pursuing an academic (respectively vocational) track.

Figure 8 confirms that only very close neighbors (at distance zero) have an impact on the probability of pursuing an academic or a vocational track. The magnitude of the estimates is 0.035, meaning that a 1% increase in the proportion of older neighbors in the vocational (respectively academic) track corresponds to a 3.5 percentage point increase in the individual probability of pursuing the vocational (respectively academic) track.

Non-linear model. Thus far and for the sake of simplicity, we have estimated independent linear probability models to analyze pairs' track choices. However, choosing the same track as one's neighbor could also be modelled as a multinomial model where:

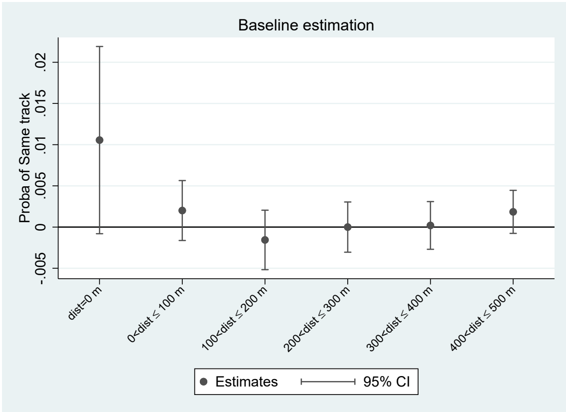
$$\begin{aligned} Y_i &= k \text{ if } U_i^k > U_i^l, \ \forall k \neq l \\ U_i^k &= \beta^k \bar{Y}_i^k + X_i' \gamma^k + \lambda^k \bar{X}_i^k + \mu_C^k + \varepsilon_i^k \end{aligned} \tag{3}$$

with $k = \{1, 2, 3\}$ indicating each possible track: Repetition, Academic, or Vocational track, and $\bar{Y}_i^k$ the share of i's older neighbors in track k.

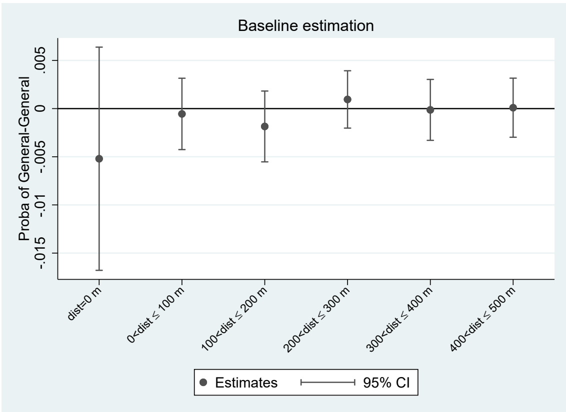
Estimating Equation (3) using a multinomial logit model gives results that are perfectly equivalent both qualitatively and quantitatively to those obtained when estimating linear probability models by ordinary least squares. Results are displayed in Appendix C in Figure C2. The corresponding odds-ratio is 1.5, meaning that a 1% increase in the proportion of older neighbors in the vocational track multiplies by 1.5 the individual probability of pursuing the vocational track rather than the academic track (the reference).

Figure 6 – Neighborhood peer effects on track choices at the end of ninth grade - Equation (2)

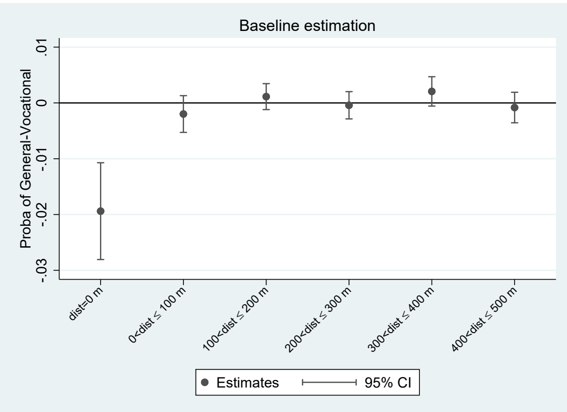
(a) Same track



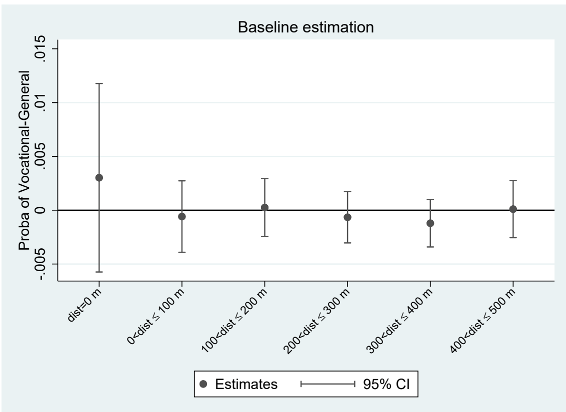
(b) i Academic - j Academic



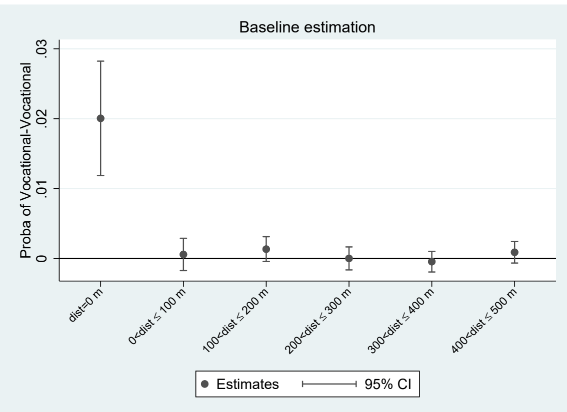
(c) i Academic - j Vocational



(d) i Vocational - j Academic



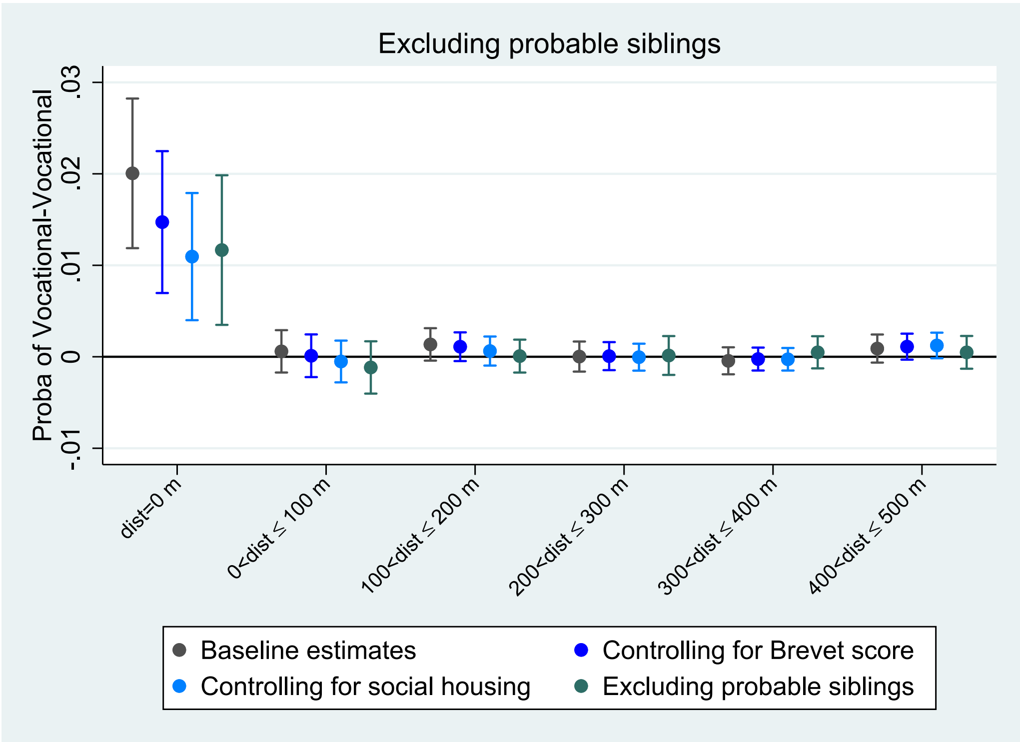
(e) i Vocational - j Vocational



Source: MENJS DEPP, FAERE-Apprenants 2017

Note: The unit of observation is a pair of individuals living in the same catchment area. The first estimate measures the effect of living at the same address (distance = 0 m) relative to living between 0 and 100 meters apart (0 < distance $\leq$ 100 m). Standard errors are clustered at the catchment area level. Catchment areas which do not pass the sorting test are excluded from the sample. All estimations include catchment-area fixed effects and characteristics of the pair (dummies for each combination of genders, parents' occupations, scholarship status, and country of birth).

Figure 7 – Robustness checks

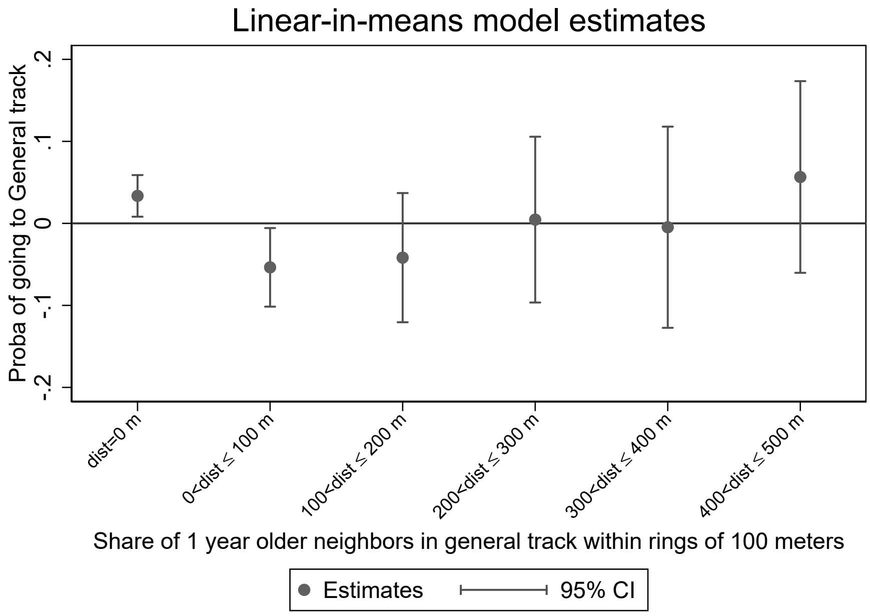


Source: MENJ, DEPP, FAERE-Apprenants 2017, RPLS 2017

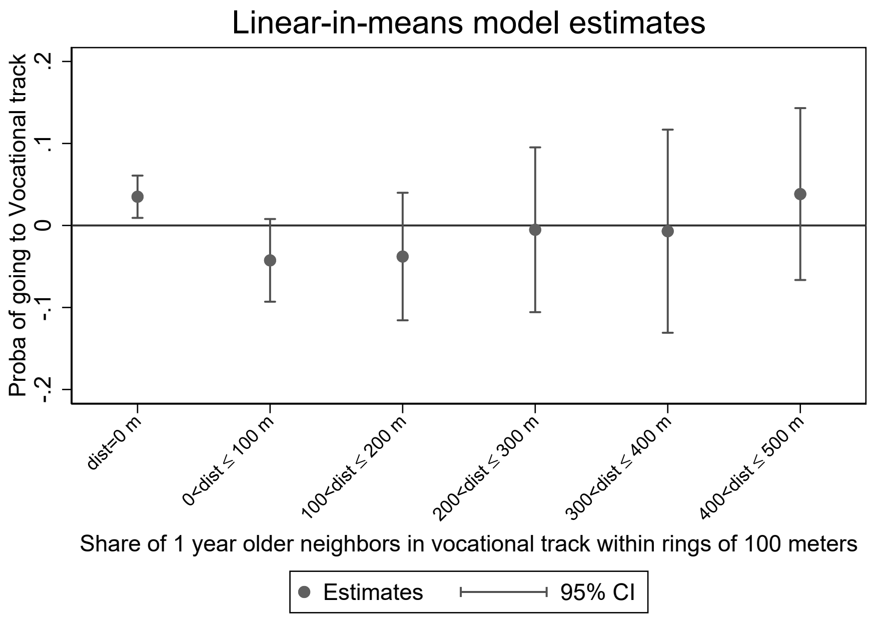
Note: The unit of observation is a pair of individuals living in the same catchment area. Each point estimate is obtained from a separate regression. The first set of estimates measure the effect of living at the same address (distance = 0 m) relative to living between 0 and 100 meters apart ($0 < \text{distance} \leq 100$ m). Standard errors are clustered at the catchment area level. Catchment areas which do not pass the sorting test are excluded from the sample. All estimations include catchment-area fixed effects and characteristics of the pair (dummies for each combination of genders, parents' occupations, scholarship status, and country of birth). Baseline estimate = same as Panel (e) of Figure 6. Controlling for Brevet score = additionally controls for i's *brevet* score. Controlling for social housing = additionally controls for social housing status of the pair. Controlling for social housing = additionally excludes probable siblings from the estimation.

Figure 8 – Linear-in-means model estimates

(a) Academic track



(b) Vocational track



Source: MENJ, DEPP, FAERE-Apprenants 2017

Note: The unit of observation is a ninth grader. Point estimates are obtained from a single regression. The first estimate measures the effect of the share of tenth graders living at the same address (distance = 0 m). The second estimate measures the effect of the share of tenth graders living between 0 and 100 meters apart (0 < distance $\leq$ 100 m), etc. Standard errors are clustered at the catchment area level. Catchment areas which do not pass the sorting test are excluded from the sample. All estimations include catchment-area fixed effects, characteristics of i and mean characteristics of neighbors (gender, parents' occupations, scholarship status, country of birth, social housing status).

5 Discussion of the mechanisms

Several mechanisms may be underlying our results. One direct explanation is that pupils are more influenced by their neighbors because they interact more frequently, such as walking home from school together or spending time together during evenings and weekends. This holds true regardless of whether they attended the same middle school. Another, more indirect, channel is through parents: parents who are close neighbors may interact more often than those who live further apart, providing more opportunities to exchange information about track choices and discuss their children's plans, thereby influencing each other and ultimately affecting the students' decisions. Although the data do not allow us to precisely observe these different mechanisms, analyzing the effect along different characteristics of the pairs provides some insights. First, under the assumption that same-sex teenagers interact more, one might expect stronger neighbor effects for same-sex pairs. Second, recent evidence from the sociological literature suggests that high-SES adults have stronger local social ties in the French context (Cayouette-Remblière and Charmes, 2024; Authier and Cayouette-Remblière, 2021). Therefore, looking at the heterogeneity of our results by parents' socio-economic background may provide some insights about the mechanisms. In particular, if neighbor effects are stronger for pupils from a lower socio-economic background, this would tend to suggest that the effects go through the children themselves rather than through the parents. Conversely, if neighbor effects are stronger for those with high SES parents, we would not be able to disentangle whether this is driven by teenagers themselves or by their parents. To test these assumptions, let us pre-multiply the model with the pairs' composition in terms of gender and parents' socio-economic status, respectively.

Table 7 presents the differentiated effects depending on the pairs' gender composition, with pairs of girls as the reference group. It shows that the effect of distance is not significantly different from zero for pairs of girls (first row), nor for opposite-sex pairs (second and third rows). Interestingly, the effect of distance is entirely driven by pairs of boys pursuing a vocational track. Every one-hundred-meter decrease in distance between neighbors increases the probability that i and j both pursue a vocational track by 0.38 percentage point more for pairs of boys compared to pairs of girls (the reference). Importantly, this heterogeneity with respect to the pairs' gender composition suggests that the effect of distance captures genuine interactions between students and is not only due to unobserved characteristics or family ties, for instance.

Table 8 reports the differentiated effects depending on the pairs' socioeconomic backgrounds.[11] For pairs composed of a low-SES younger student i and a high-SES older student j, a one-hundred-meter decrease in distance increases the probability of both students pursuing a vocational track after ninth grade by 0.35 percentage point, compared to pairs of high-SES students (the reference). Remarkably, the effect of distance is larger when only the older student j comes from a low socioeconomic background (third row), as compared to high-SES pairs. In that case, every one-hundred-meter decrease in distance increases the probability that i chooses the same track as j by 1.1 percentage point more for high-SES-low-SES student pairs than for exclusively high-SES pairs. This is due to the fact that distance has a larger impact on the probability that both students choose the academic track, while simultaneously having a smaller effect on the probability that they choose different tracks (academic and vocational, with a 0.85 percentage point smaller effect with every one-hundred-meter decrease, or vocational and academic, with a 0.38 percentage point smaller effect with every one-hundred-meter decrease). Compared to pairs of high-SES pupils, pairs of low-SES pupils have a 0.75 percentage point higher probability of both pursuing a vocational track after ninth grade with every one-hundred-meter decrease in distance. There is a marginally significant 0.81 percentage point decrease in the probability that both students pursue an academic track when distance is shorter by one hundred meters.

In the same vein, we also tested whether pairs' composition in terms of scholarship status – a proxy for parental income – influence the results, but we find no significant

[11]For sake of simplicity, let us consider two categories: low-SES pupils are defined as children of blue-collar workers or of unemployed parents; high-SES pupils are defined as children of executives or intermediate occupations.

differences in neighbor effects in this dimension.[12]

To summarize the results, we find that neighbor effects are driven by pairs of boys, and are larger for low-SES students. This tends to support the intuition that these effects go through teenagers influencing each other more than through their parents' exchanging information. This comforts the idea that we capture the existence of peer effects in the neighborhood, rather than an effect of the environment.

[12] Alternatively, we investigated the heterogeneity of the results depending on whether the pupils are born in France or not, and we do not find any significant difference either. This is not surprising as about 90% of our sample is born in France.

Table 7 – Neighbor effects on track choices at the end of ninth grade - Equation (1) - By gender

	i's track-j's track				
	Same track	Gen^{al}-Gen^{al}	Gen^{al}-Voc^{al}	Voc^{al}-General	Voc^{al}-Voc^{al}
Dist. (100 meters)	0.0020	0.0022	-0.0032*	0.0005	-0.0003
	(0.0021)	(0.0021)	(0.0019)	(0.0014)	(0.0006)
i Male $\times$ Dist.	-0.0011	-0.0025	0.0027	-0.0007	0.0014
	(0.0027)	(0.0028)	(0.0018)	(0.0024)	(0.0009)
j Male $\times$ Dist.	-0.0042	-0.0045	0.0040	0.0006	0.0003
	(0.0031)	(0.0029)	(0.0032)	(0.0014)	(0.0010)
Male $\times$ Male $\times$ Dist.	-0.0007	0.0031	-0.0036	0.0034	-0.0038**
	(0.0036)	(0.0028)	(0.0023)	(0.0023)	(0.0018)
Intercept	0.1721***	-0.0564***	-0.1524***	0.7821***	0.2285***
	(0.0244)	(0.0202)	(0.0237)	(0.0257)	(0.0196)
Nbr obs	1,033,653	1,033,653	1,033,653	1,033,653	1,033,653
Nbr clusters	81	81	81	81	81

Source: MENJS DEPP, FAERE-Apprenants 2017

Note: * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$). The unit of observation is a pair of individuals living in the same catchment area, and within a 1km-distance radius. Standard errors in parentheses are clustered at the catchment area level. Catchment areas which do not pass the sorting test are excluded from the sample. All estimations include catchment-area fixed effects and characteristics of the pair (dummies for each combination of genders, parents' occupations, scholarship status, and country of birth). For the sake of simplicity, we do not display the two outcomes involving grade repetition, but the results, which are not significant, are available upon request.

Table 8 – Neighbor effects on track choices at the end of ninth grade - Equation (1) - By socioeconomic status (Low SES)

	i's track-j's track				
	Same track	Gen^{al}-Gen^{al}	Gen^{al}-Voc^{al}	Voc^{al}-General	Voc^{al}-Voc^{al}
Dist. (100 meters)	0.0025	0.0018	-0.0033*	0.0002	0.0007*
	(0.0019)	(0.0020)	(0.0019)	(0.0009)	(0.0004)
i Low SES=1 $\times$ Dist.	-0.0019	0.0016	-0.0006	0.0033	-0.0035**
	(0.0029)	(0.0028)	(0.0019)	(0.0032)	(0.0015)
j Low SES=1 $\times$ Dist.	-0.0110***	-0.0119***	0.0085**	0.0038***	0.0009
	(0.0036)	(0.0039)	(0.0039)	(0.0011)	(0.0010)
Low SES=1 $\times$ Low SES=1 $\times$ Dist.	0.0006	0.0081*	0.0026	-0.0052	-0.0075**
	(0.0044)	(0.0041)	(0.0044)	(0.0035)	(0.0030)
Intercept	0.2368***	0.0194	-0.2294***	0.7945***	0.2173***
	(0.0241)	(0.0191)	(0.0237)	(0.0266)	(0.0191)
Nbr obs	1,033,653	1,033,653	1,033,653	1,033,653	1,033,653
Nbr clusters	81	81	81	81	81

Source: MENJS DEPP, FAERE-Apprenants 2017

Note: * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$). The unit of observation is a pair of individuals living in the same catchment area, and within a 1km-distance radius. Standard errors in parentheses are clustered at the catchment area level. Catchment areas which do not pass the sorting test are excluded from the sample. All estimations include catchment-area fixed effects and characteristics of the pair (dummies for each combination of genders, parents' occupations, scholarship status, and country of birth). For the sake of simplicity, we do not display the two outcomes involving grade repetition, but the results are available upon request.

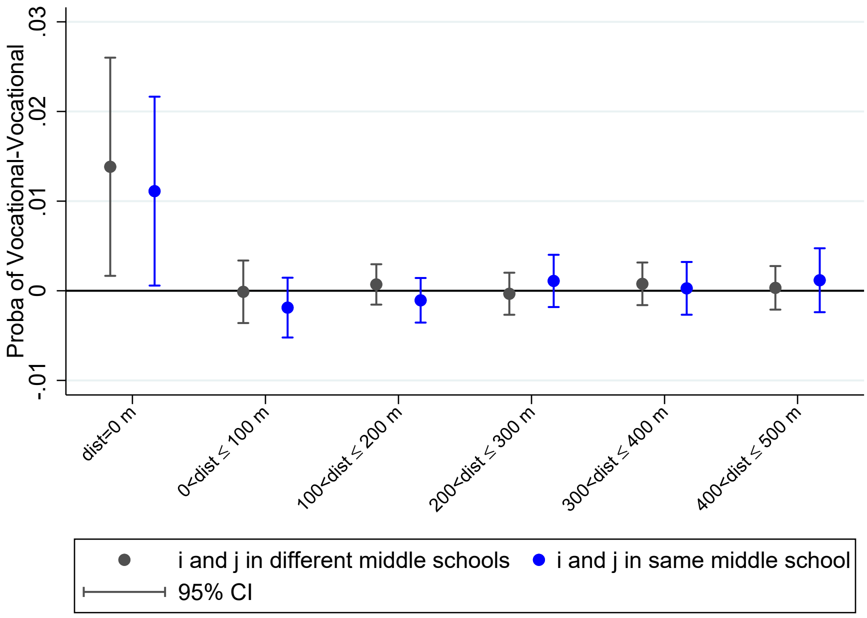
One concern could be that our results do not only reflect neighbor effects but also school peer effects, as many students in our sample do not attend their default (catchment-area) middle school. For instance, around 28 % of pupils in our sample are enrolled at a private school (Table 2). In this case, if close neighbors systematically attend the same school to a larger extent than neighbors living further away, our results could simply reflect this pattern. We therefore split our sample between pairs of pupils who attended the same middle school and those who went to different schools.

Figure 9 reports the point estimates from equation 2 on each subsample, for the probability that both i and j go to the vocational track (the results for the other outcomes are displayed in Figure E4). The results for close neighbors who attended the same school (in blue) are not statistically different from those who do not, confirming that our results indeed reflects neighbor effects and not schoolmate effects.

Obviously, we do not expect all pupils to be equally influenced by their neighbors when it turns to track choice, and in particular to the decision to go to a vocational track. Indeed, as discussed in Section 2.1, vocational track is often considered a low-status track or a default option for pupils who do not perform well enough to pursue the academic track. This implies that pupils who perform relatively well in middle school, and whose admission to a general track is almost automatic, are very unlikely to be influenced by their neighbors towards choosing a vocational track. As a matter of fact, we do observe that 99 % of pupils who enroll in an academic track in high school obtained more than 47.5 % (9.5/20) at the *Brevet* (end-of-middle-school national exam). Conversely, this test score corresponds to the median grade for pupils who enroll in a vocational track. To investigate this further, we split our sample into four subsamples based on the score obtained by the ninth grader (i) at the *Brevet* exam. Results for the probability that both i and j go to the vocational track are displayed in Figure 10.[13] We can see that neighbor effects are only significant on the two subsambles of ninth graders with low test scores, and that this effect if twice as large as the one we obtained on the whole sample. This confirms that pupils

[13]Results for all other outcomes are displayed in Figure D3

Figure 9 – Neighbor effects on track choices at the end of ninth grade - Equation (2) - By same middle school

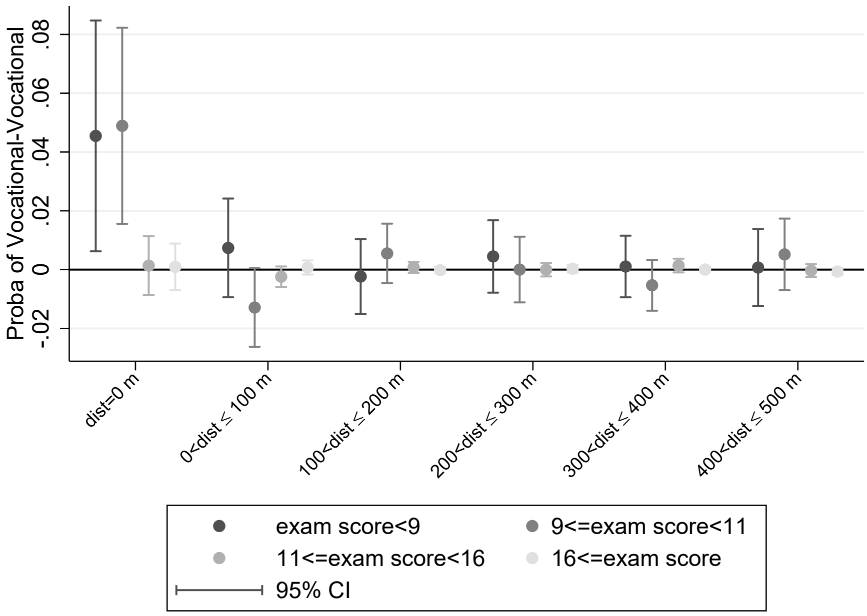


Source: MENJS DEPP, FAERE-Apprenants 2017

Note: The unit of observation is a pair of individuals living in the same catchment area. The sample is split into two subsamples based on whether the two pupils attended the same middle school or not. Each point estimate is obtained from a separate regression. The different colors correspond to the two subsamples. The first set of estimates measure the effect of living at the same address (distance = 0 m) relative to living between 0 and 100 meters apart (0 < distance ≤ 100 m). Standard errors are clustered at the catchment area level. Catchment areas which do not pass the sorting test are excluded from the sample. All estimations include catchment-area fixed effects and characteristics of the pair (dummies for each combination of genders, parents' occupations, scholarship status, and country of birth). We also include the variables considered in the robustness analysis (social housing status of the pair, i's *brevet* score) and we exclude probable siblings from the sample.

who can enroll in an academic track are not influenced by their older neighbors who are in a vocational track. Rather, the neighbor effects uncovered in this paper are driven by pupils who are at the margin of passing the *Brevet* (exam score between 9 and 11) or those who are at the bottom of the grade distribution.

Figure 10 – Neighbor effects on track choices at the end of ninth grade - Equation (2) - By end-of-middle-school exam score



Source: MENJS DEPP, FAERE-Apprenants 2017

Note: The unit of observation is a pair of individuals living in the same catchment area. The sample is split into four subsamples based on the score obtained by the ninth grader (i) at the *Brevet* (end-of-middle-school national exam). Each point estimate is obtained from a separate regression. The different shades of gray correspond to the four subsamples. The first set of estimates measure the effect of living at the same address (distance = 0 m) relative to living between 0 and 100 meters apart (0 < distance $\leq$ 100 m). Standard errors are clustered at the catchment area level. Catchment areas which do not pass the sorting test are excluded from the sample. All estimations include catchment-area fixed effects and characteristics of the pair (dummies for each combination of genders, parents' occupations, scholarship status, and country of birth). We also include the variables considered in the robustness analysis (social housing status of the pair, i's *brevet* score) and we exclude probable siblings from the sample.

6 Conclusion

Using geo-coded administrative data from the French Ministry of Education, this paper analyzes the effect of neighbors' choices on individual track choices at the end of lower secondary education, which is a key turning point in the French school system, as pupils can either enroll in a vocational or in an academic track. To account for endogenous location decisions, we use the within-catchment-area variation in distance between neighbors, restricting our analysis to catchment areas for which we do not find evidence of spatial clustering along key dimensions (in particular, parents' socio-economic background). Our analysis relies on a pair-wise model, where each ninth grader is paired with every tenth grades residing within a one kilometer distance radius. Doing so ensures that we avoid the reflection problem, and enables us to precisely document the spatial scope of neighbor effects.

Our results show that older neighbors do have an impact on the probability to attend vocational education. The effect fades away very rapidly, as only same-building neighbors matter. These results are not driven by spatial sorting across neighborhoods, as they hold even when excluding catchment areas where spatial auto-correlation based on socio-demographic characteristics is likely. They are robust to controlling for academic achievement, social housing, or when we exclude probable siblings.

Importantly, we also document that neighbor effects are higher for pairs of boys, and they tend to increase the probability of pursuing an vocational track for pairs of peers from a lower socio-economic background. Overall, this is consistent with peers influencing each other rather than neighboring parents exchanging information. We also show that our findings are not due to closer pupils attending the same school, comforting the idea that we are capturing genuine neighbor effects, on top of schoolmate effects. Unsurprisingly, we find that this effect is driven by pupils who are at the bottom of the grade distribution.

The choice between an academic track and a vocational track at the end of ninth grade is an important stage, as these lead to very different educational paths and labor market opportunities. In France, the academic track is perceived as the most academically prestigious, whereas the vocational track is often considered a low-status track or a default option for pupils who do not perform well enough to pursue the academic track. If neighbor effects tend to increase the probability of enrolling in the vocational track for pairs of low-SES pupils, then it implies that neighbor effects tend to accentuate social segregation across high school tracks.

Overall, our findings suggest that place-based education policies relying on people's place of residence may be relevant, especially for the most socially and academically disadvantaged students. However, they should be designed carefully, taking into consideration potential effects on social segregation across education tracks.

References

Albert, D., J. Chein, and L. Steinberg (2013). Peer Influences on Adolescent Decision Making. *Current Directions in Psychological Science 22*(2), 114–120.

Altmejd, A., A. Barrios-Fernández, M. Drlje, J. Goodman, M. Hurwitz, D. Kovac, C. Mulhern, C. Neilson, and J. Smith (2021, 03). O Brother, Where Start Thou? Sibling Spillovers on College and Major Choice in Four Countries*. *The Quarterly Journal of Economics 136*(3), 1831–1886.

Authier, J.-Y. and J. Cayouette-Remblière (2021). Neighbourliness in france: an enduring—and selective—practice. *Population Societies 589*(5), 1–4.

Barrios-Fernández, A. (2022, July). Neighbors' Effects on University Enrollment. *American Economic Journal: Applied Economics 14*(3), 30–60.

Battiston, A., S. Hedges, T. Lazarowicz, and S. Speckesser (2020, January). Peer Effects and Social Influence in Post-16 Educational Choice. Research Discussion Paper 025 025, Centre for Vocational Education Research.

Bayer, P., S. L. Ross, and G. Topa (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy 116*(6), 1150–1196.

Bifulco, R., J. M. Fletcher, and S. L. Ross (2011). The effect of classmate characteristics on post-secondary outcomes: Evidence from the add health. *American Economic Journal: Economic Policy 3*(1), 25–53.

Boucher, V., M. Rendall, P. Ushchev, and Y. Zenou (2024). Toward a general theory of peer effects. *Econometrica 92*(2), 543–565.

Boutchenik, B., P. Givord, and O. Monso (2020, February). How do restrictive zoning and parental choices impact social diversity in schools?: A methodological contribution to the decomposition of segregation indices applied to France. Sciences Po publications 105, Sciences Po.

Brown, B. B. and J. Larson (2009). *Peer Relationships in Adolescence*, Chapter 3. John Wiley Sons, Ltd.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009, 10). Peer Effects and Social Networks in Education. *The Review of Economic Studies 76*(4), 1239–1267.

Cayouette-Remblière, J. and E. Charmes (2024). Social ties in and out of the neighbourhood: Between compensation and cumulation. *Urban Studies*, 00420980231212298.

Chetty, R. and N. Hendren (2018, 02). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects*. *The Quarterly Journal of Economics 133*(3), 1107–1162.

Chetty, R., N. Hendren, and L. F. Katz (2016, April). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review 106*(4), 855–902.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010). Identification of Social Interactions through Partially Overlapping Peer Groups. *American Economic Journal: Applied Economics 2*(2), 241–275.

Del Bello, C. L., E. Patacchini, and Y. Zenou (2015). Neighborhood effects in education.

DEPP (2020). L'état de l'école. Technical report, Direction de l'évaluation, de la prospective et de la performance - Ministère de l'Éducation nationale, de la Jeunesse et des Sports.

Deutscher, N. (2020, April). Place, peers, and the teenage years: Long-run neighborhood effects in australia. *American Economic Journal: Applied Economics 12*(2), 220–49.

Epple, D. and R. E. Romano (2011). Peer effects in education: A survey of the theory and evidence. In *Handbook of social economics*, Volume 1, pp. 1053–1163. Elsevier.

Fricke, H., J. Grogger, and A. Steinmayr (2018). Exposure to academic fields and college major choice. *Economics of Education Review 64*(C), 199–213.

Gibbons, S. and S. Telhaj (2016). Peer effects: Evidence from secondary school transition in england. *Oxford Bulletin of Economics and Statistics 78*(4), 548–575.

Goux, D., M. Gurgand, and E. Maurin (2017). Adjusting your dreams? high school plans and dropout behaviour. *The Economic Journal 127*(602), 1025–1046.

Goux, D. and E. Maurin (2007). Close neighbours matter: Neighbourhood effects on early performance at school*. *The Economic Journal 117*(523), 1193–1215.

Guyon, N. and E. Huillery (2020, 06). Biased Aspirations and Social Inequality at School: Evidence from French Teenagers. *The Economic Journal 131*(634), 745–796.

Helmers, C. and M. Patnam (2014). Does the rotten child spoil his companion? Spatial peer effects among children in rural India. *Quantitative Economics 5*(1), 67–121.

Hémet, C. and C. Malgouyres (2018). Diversity and Employment Prospects Neighbors Matter! *Journal of Human Resources 53*(3), 825–858.

Jonsson, J. O. and C. Mood (2008). Choice by contrast in swedish schools: How peers' achievement affects educational choice. *Social Forces 87*(2), 741–765.

Landaud, F., S. T. Ly, and É. Maurin (2020). Competitive schools and the gender gap in the choice of field of study. *Journal of Human Resources 55*(1), 278–308.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics 28*(4), 825–860.

Ly, S. T. and A. Riegert (2014, September). Persistent Classmates: How Familiarity with Peers Protects from Disruptive School Transitions. Working Paper 2013 – 21, Paris School of Economics. working paper or preprint.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Monarrez, T. E. (2023, July). School attendance boundaries and the segregation of public-schools in the us. *American Economic Journal: Applied Economics 15*(3), 210–237.

OECD (2019). PISA 2018 Results - Volume I-III. Technical report, OECD.

Patacchini, E., E. Rainone, and Y. Zenou (2017). Heterogeneous peer effects in education. *Journal of Economic Behavior and Organization 134*, 190 – 227.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*, Volume 3, pp. 249–277. Elsevier.

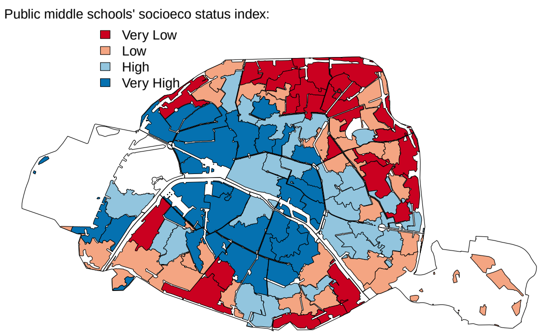
Silliman, M. and H. Virtanen (2022, January). Labor market returns to vocational secondary education. *American Economic Journal: Applied Economics 14*(1), 197–224.

Solignac, M. and M. Tô (2018). Do workers make good neighbours? The impact of local employment on young male and female entrants to the labour market. *Annals of Economics and Statistics* (130), 167–198.

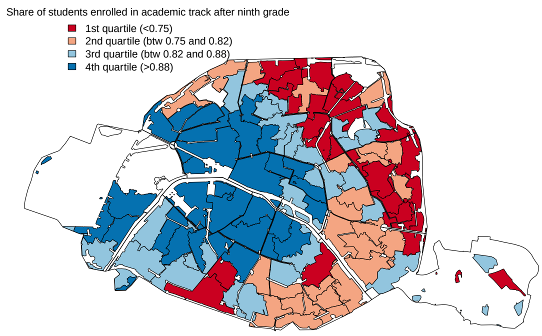
A Maps of share of pupils following an academic track

Figure A1 – Characteristics of Paris Middle schools' catchment areas

(a) Social background index (2020)



(b) Share of pupils in academic track (2017)



B No-fixed-effect estimates

Table B1 – **No fixed-effect** estimates of neighbors' effects on track choices - All catchment areas

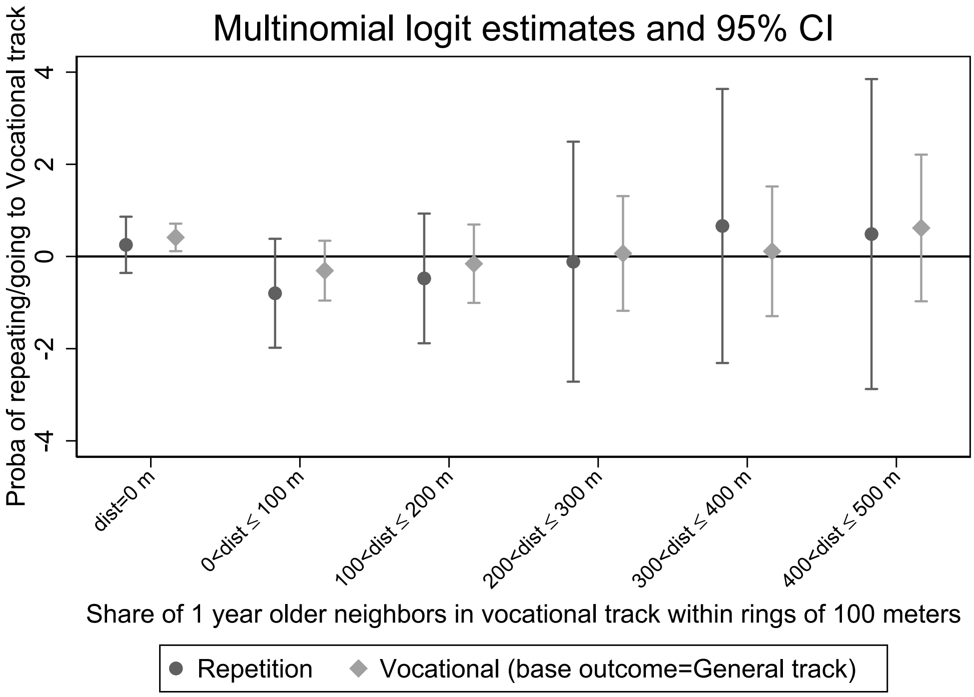
	i's track-j's track						
	Same track	Rep^{ion}-Gen^{al}	Rep^{ion}-Voc^{al}	Gen^{al}-Gen^{al}	Gen^{al}-Voc^{al}	Voc^{al}-Gen^{al}	Voc^{al}-Voc^{al}
Dist. btw neighbors (100 meters)	-0.0011	0.0009**	0.0000	0.0010	0.0002	0.0000	-0.0022***
	(0.0012)	(0.0004)	(0.0002)	(0.0016)	(0.0011)	(0.0008)	(0.0006)
Dist.^{2}	0.0004**	-0.0001**	-0.0000	0.0003	-0.0002	-0.0001	0.0001*
	(0.0002)	(0.0000)	(0.0000)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Intercept	0.7316***	0.0143***	0.0033***	0.6943***	0.1307***	0.1201***	0.0373***
	(0.0087)	(0.0026)	(0.0008)	(0.0110)	(0.0064)	(0.0069)	(0.0030)
Nbr obs	3,271,014	3,271,014	3,271,014	3,271,014	3,271,014	3,271,014	3,271,014
Nbr clusters	111	111	111	111	111	111	111

Source: MENJ, DEPP, FAERE-Apprenants 2017

Note: * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$). The unit of observation is a pair of individuals living in the same catchment area. Controls at the pair level = gender, SES, scholarship status, country of birth, and school choice. Standard errors in parentheses are clustered at the catchment area level.

C Estimation results - Non-linear model

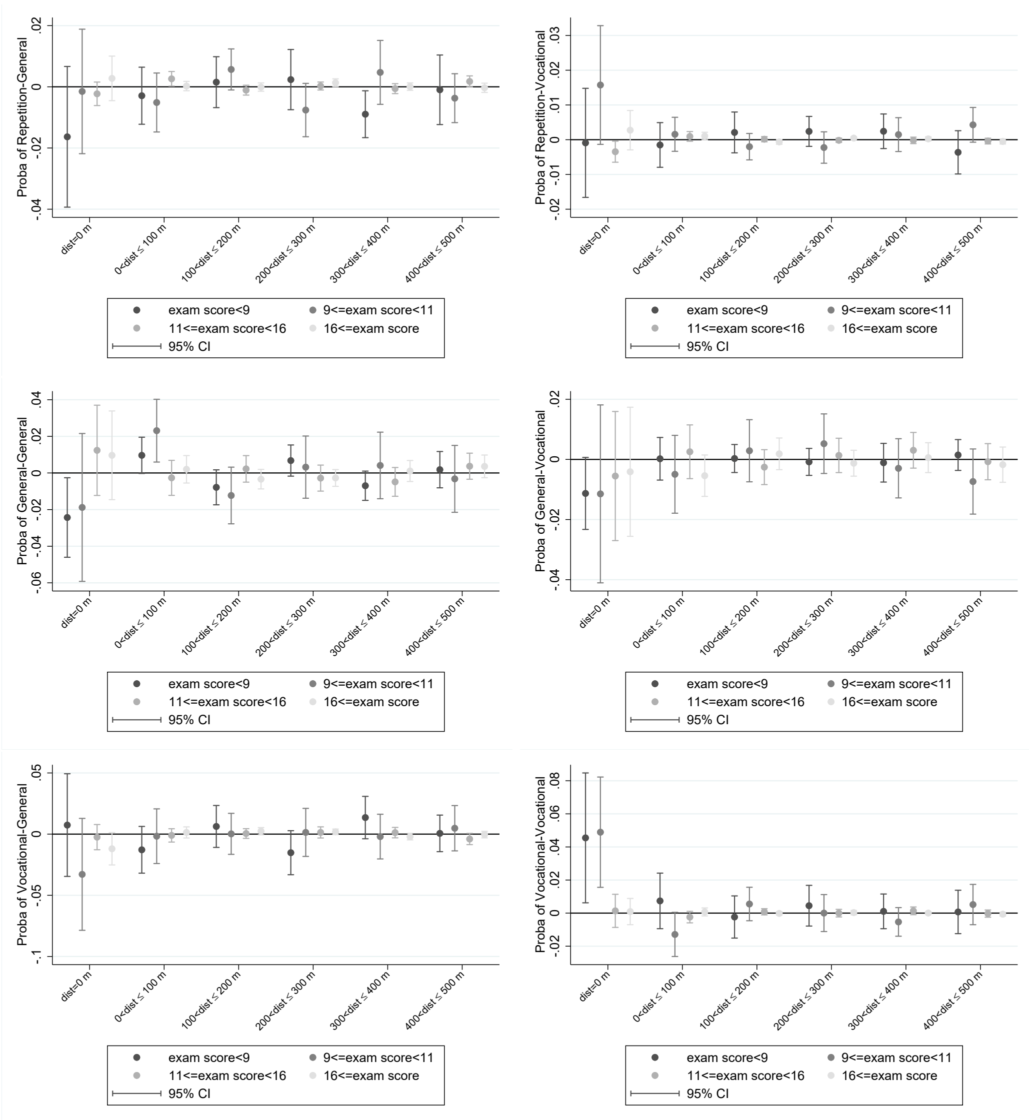
Figure C2 – Multinomial logit model



Source: MENJS DEPP, FAERE-Apprenants 2017, RPLS 2017

D Neighborhood peer effects on track choices at the end of ninth grade - By end-of-middle-school exam score

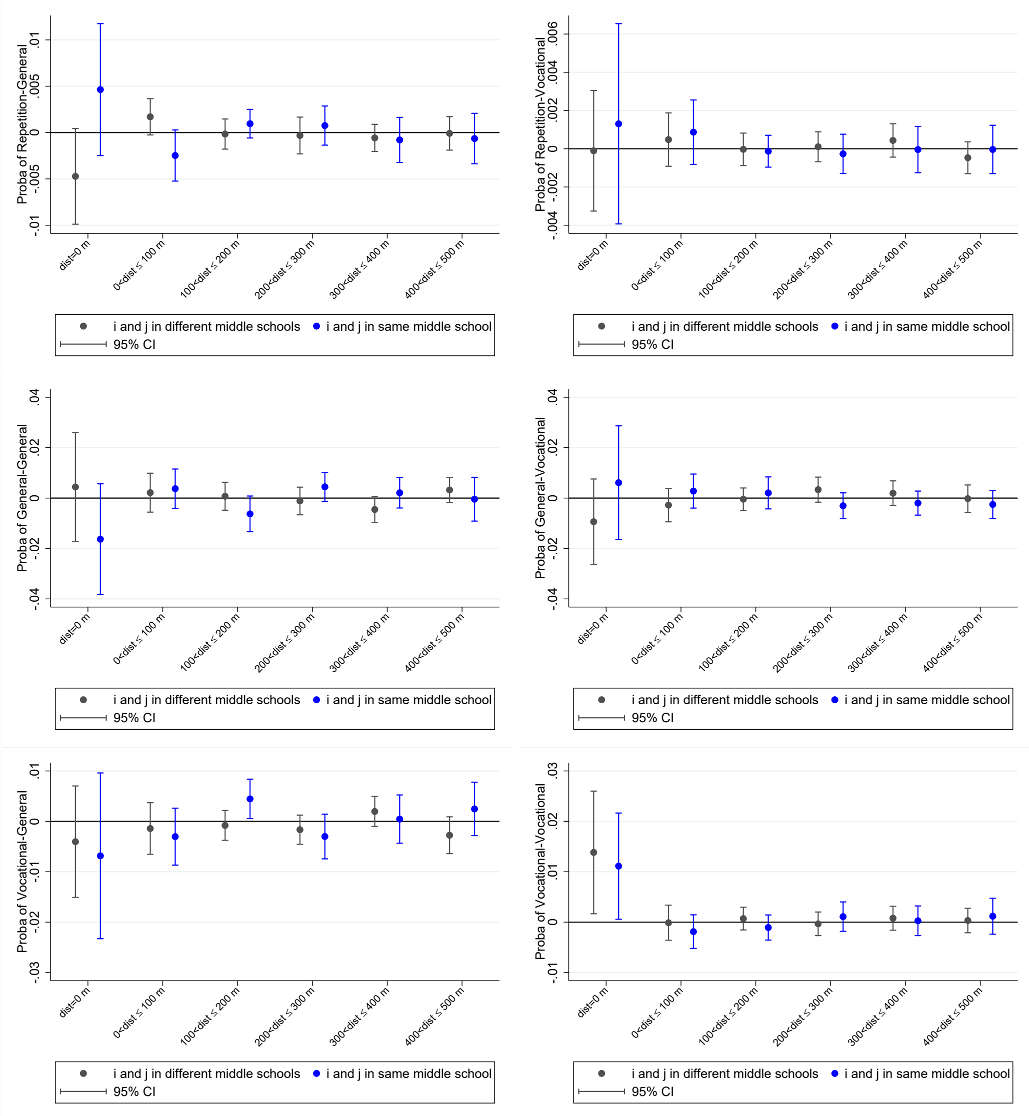
Figure D3 – Neighborhood peer effects on track choices at the end of ninth grade - Equation (2) - By end-of-middle-school exam score



Source: MENJS DEPP, FAERE-Apprenants 2017

E Neighborhood peer effects on track choices at the end of ninth grade - By same middle school

Figure E4 – Neighborhood peer effects on track choices at the end of ninth grade - Equation (2) - By same middle school



Source: MENJS DEPP, FAERE-Apprenants 2017