JMS | Journées de méthodologie statistique de l'Insee | 2018

Peer effects with peer and student heterogeneity: an assessment for French Baccalauréat

Béatrice BOUTCHENIK(*)(**), Sophie MAILLARD(*)

(*) INSEE – SSPLab
(**) Université Paris-Dauphine

beatrice.boutchenik@insee.fr

Keywords. achievement spillovers, peer effects, social interaction models, education and inequality, desegregation.

Résumé

Dans quelle mesure les notes au Baccalauréat d'un élève reflètent-elles le niveau de ses camarades de classe de Terminale ? Nous utilisons des données administratives sur les résultats au Baccalauréat (2010-2016) pour évaluer les effets de pairs. Nous nous appuyons sur la variabilité entre classes et entre cohortes au sein des lycées et nous restreignons l'analyse à un échantillon de lycées au sein desquels nous n'identifions pas de politique de classes de niveau. Nous autorisons l'effet des pairs à varier en fonction du niveau initial de l'élève, tel que mesuré par sa note au Brevet des collèges, et nous avons recours à une typologie de classes pour étudier l'effet de la composition globale de la classe. Nous montrons qu'une proportion élevée de bons élèves dans la classe est surtout profitable aux plus faibles, et peut même être défavorable pour les autres pairs de niveau élevé. A l'inverse, une proportion élevée d'élèves de faible niveau pénalise surtout les autres élèves fragiles, et moins les bons élèves. Ces conclusions restent valables quel que soit le sexe de l'élève considéré.

Abstract

To what extent are high school students impacted by their class composition? We use French administrative data on *Baccalauréat* scores (2010-2016) to assess peer effects. We rely on within-school, between-class and between-cohort variation to identify those effects, and focus on a subsample of schools for which we do not identify any tracking policy across classes. We allow peer effects to vary with students' own prior performance and resort to class clustering to investigate the heterogeneous effects of peers given the whole class composition. We show that a large proportion of high-ability peers is especially profitable to low-ability peers, and may even be harmful to high-ability peers. Conversely, a high share a low-ability peers is mostly detrimental for other low-ability peers, and less so to high ability ones. These conclusions hold regardless of pupils' gender.

Introduction

How much do students benefit or suffer from their high school classmates? Are within-class spillovers a zero-sum game? If not, who benefits most from which type of peers? These are key questions for educational policies. Indeed, if students exert spillover effects on their peers, and if these externalities do not linearly add up or cancel out when a student changes class, then fine-tuned assignment of students to classes can improve general education achievement at virtually no cost. To address this question, we study the impact of 12^{th} grade classroom peers on the score at *Baccalauréat*, the French high school final exam, from 2010 to 2016.

Peer effects are particularly difficult to identify in an educational framework without a natural experiment setting. One main issue is to disentangle exogenous peer effects from correlated effects, according to the terminology introduced by Manski (1995). Whereas exogenous peer effects measure to what extent students are actually impacted by their peers' individual characteristics- typically, the initial school ability of peers-, correlated effects capture the fact that peers may resemble each other in terms of individual characteristics. For instance, there may be sorting of pupils according to initial academic level and parental background. Disentangling these two channels is therefore made difficult in a context such as that of French high-schools where neither the assignment to school and major, nor the assignment to classrooms within schools and major, have reasons to be random. Here, we propose a methodology using panel data that allows us to address these issues. In line with Ammermueller and Pischke (2009), we use school and major fixed effects and rely on within-school, within-major variability in classroom peers, both between classes for a given cohort and between cohorts. This strategy allows us to identify exogenous peer effects under the assumption that class assignment process is not based on unobserved student ability. We therefore identify a subset of schools which exhibit no sorting behavior of pupils based on initial academic level, within a given major, across the seven cohorts considered. Ammermueller and Pischke (2009) and Burke and Sass (2013) also present this type of sample restriction strategy. Here, the analysis is all the more credible that we observe class formation for several subsequent years. If for a given year, a high school statistically appears to be sorting students across classes, it is hard to tell whether this is due to random variation or to a specific and willful behavior. However, if statistical sorting is observed year after year, there is good reason to believe this high school actually conducts a non-random allocation policy. We argue that the identified subset of high-schools corresponds to schools which are less subject to competition with other high-schools, so that there is less pressure for them to answer to parents' requests in terms of tracking.

Our paper contributes to a finer description of the impact of class composition in French schools. Little empirical work has been conducted on this topic for secondary education in France. Ly and Riegert (2014) study peer effects in secondary schools with the probability to repeat the first year of high school as the main outcome, depending on class composition (mainly peer ability and the number of persistent classmates from the previous year). Their results are very informative on the positive impact that persistent classmates can have in a disruptive school transition. However, the effect of peer ability on grade repetition cannot be interpreted as a "learning" peer effect, in an institutional context where ranking effects are very likely to be at stake (since repetition decisions are taken at the class level by the school board). Goux and Maurin (2007) focus on another aspect of social interactions considering neighborhood effects among middle-school aged children. Those kinds of phenomena, taking place possibly out of the school, are out of the scope of this work. However, we argue that the class level is very relevant in the French education system to capture the actual environment that students are exposed to in their everyday school life. Contrary to Davezies (2004), who uses panel data for the cohort entering 1^{st} grade in 1997 to measure the impact of peers' social background and school performance in primary schools, our work is dedicated to what happens among high school students. As in many other countries, the end of secondary education is a very important milestone for French students: end of high school tests are crucial for higher education pursuit. Hence, peer effects could be persistent, more than they can be at other educational stages. Moreover, the work undertaken on peer effects in French schools has mainly focused on the linear-in-means model, looking at the average effect of the mean of peer characteristics.

Hoxby and Weingarth (2005) have shown how fruitful peer effect specifications going beyond this linear-in-means model were. In these models, peer effects can vary with both the student's and the peers' performance. On the one hand, high and low achieving students may be impacted by the same peer composition to various extents. On the other hand, a student could be differently affected by a change in her average peers performance whether it is due to a general shift in performance distribution, or if it is accounted by the presence of a few very high or very low achieving peers. Measuring heterogeneous and non-linear peer effects is actually crucial if one wants to increase efficiency and equity in schools and classes. Indeed, in a linear-in-means model, moving a high achieving student from a class to another has zero cumulated effect since her positive effect on her peers is transferred from a class to another but is not quantitatively different. To improve efficiency what we rather need to know is in which type of class a high-achieving student has more spillover effects on her peers, and in which class a low-achieving one affects the class performance as little as possible or/and which class composition is the most favorable for her improvement. This kind of diagnosis is by construction impossible in the linear-in-means model.

A growing empirical literature has since focused on heterogeneous and non-linear peer effects. Several subsequent papers rely on specifications close to that of Hoxby and Weingarth (2005). For instance, Lavy, Silva, and Weinhardt (2012) examine the differentiated effects of having "bad" vs. "good" peers, interacted with pupils' own ability and gender. Their findings are consistent with a "bad apple" model, as coined by Hoxby and Weingarth (2005), where a few low-achievers have very negative spillovers on the rest of the class. Others have directly investigated the effect of heterogeneous abilities in the class, exploring whether students benefit from diverse abilities in their class, whatever their own ability level (the *rainbow* model in the terminology of (Hoxby and Weingarth, 2005), or if they are better off in a homogeneous class (the *focus* model). For instance, Bertoni, Brunello, and Cappellari (2017), Booij, Leuven, and Oosterbeek (2017), Lyle (2009) introduce directly in their peer effect specification a measure of pupil diversity (standard error, variance, interquartile range) in the classroom. Results are mixed but they rather point to a positive effect of diversity, for a given mean performance in the class: Rangvid (2007), Schneeweis and Winter-Ebmer (2007), Kiss (2013) find little or no significant effect (the first two papers focus on social heterogeneity) while Vigdor and Nechyba (2006) and Lyle (2009) assess positive effects of ability diversity in the classroom. The latter notices that as soon as a dispersion measure is included (variance or interquartile range), there is no more significant impact of the peer average ability.

While various measures of diversity have been introduced in peer effect estimations, to our knowledge there has been little concern for entirely describing the composition of classrooms and considering the effect of belonging to specific class types for students who are themselves of different types. We aim to bridge this gap in the literature by taking advantage of rich exhaustive administrative data for France, and resort to class clustering in order to go further than the baseline heterogeneous effect models based on interactions. Indeed, using interactions allowing each student to be differently influenced by her peers according to their performance has some limitations. In particular, while we argue the allocation of pupils *within* schools may be considered as random on the subsample of schools considered, the allocation of pupils *between* schools is strongly correlated to their initial academic level. Thus, in a school-fixed-effects framework, it is hard to distinguish heterogeneous effects depending on pupils' type from heterogeneous peer effects depending on the baseline classroom composition. For the sake of illustration, let us assume that we observe that peer quality is more beneficial to low achievers than to high achievers. In our framework, this could be due to the fact that high achievers are concentrated in good classes, whereas low achievers are often in classes with very few high achievers. The marginal effect of average peer academic level may be higher in classrooms where the baseline academic level is low. Hence, it is hard to disentangle heterogeneous peer effects from non-linear ones. We aim to overcome this interpretation issue for heterogeneous effects considering class composition is relevant as a whole to analyze student performance. Note that we do not seek to measure peer effects conditional of class type but directly to estimate the effect on achievement for a peer of a given initial level to be assigned to a specific kind of class. To do so, we reduce dimensionality using class clustering to identify a limited number of class types capturing the kind of class environment a student is confronted to.

We present evidence of heterogeneous peer effects in French schools at the *Baccalauréat* level. We show that high-ability pupils benefit from being in a class where there are not too many other pupils of their own type. On the contrary, being in a classroom environment with a large proportion of high-ability pupils may be highly beneficial to low-ability ones. A high share of low-ability pupils is detrimental to everyone, but especially so to low-ability students. This calls for more diversity in the composition of classroom, as it appears that the classroom type with a repartition of pupils similar to that observed in the general population, is rather beneficial to all types of pupils. The results obtained are rather close for boys and girls.

The remainder of this paper is as follows. Section 1 describes the French context and the data we use. Section 2 presents the model and discusses conditions for identification. Section 3 presents our preliminary results. Section 4 concludes.

1 Institutional context and data

1.1 The French secondary education system

In France, education from 11 to 18 is separated in two institutions: *collège* or middle school (from 6^{th} to 9^{th} grade), and *lycée* or high school (from 10^{th} to 12^{th} grade).[1] In both middle and high schools, students are assigned to the same class for all subjects and for the entire school year, with a few exceptions- optional classes or language classes can typically be composed of peers belonging to different classes. Hence, students spend most of the study week with the same peers: they change teachers for each subject, but the class composition is often identical. This feature makes the classroom one of the most relevant levels to investigate peer effects in the French secondary education system.

Secondary education is punctuated by two exams, one at the end of *collège* (9^{th} grade) -*Brevet* or DNB hereinafter- and the other at the end of *lycée* -*Baccalauréat* or *Bac* for short. The *DNB* score considered here depends on a national written exam in French, mathematics and history-geography. These tests are graded anonymously and by teachers from other schools. The *Bac* is the final exam of the high school education (12^{th} grade). Unlike *DNB*, it is only based on performance at final exams, most of them written and taken in June on the last high

[1]Early vocational tracks can be followed immediately after 8^{th} grade. In addition, school is only mandatory until sixteen year old (completed).

school year.[2] Final examination is taken for all subjects studied in high school, depending on the major. A student needs to pass the *Bac* to carry on in the French higher education system.

Middle school curriculum is rather uniform, and more so since the early 2000s due to the drop in repetition rate and the closing of some of the earliest vocational tracks, as evidenced by Caille (2014). At the end of middle school, students apply for either vocational or general studies, with the approval of their teachers. Among the pupils who started 6$^{\text{th}}$ grade in 2007, 61.7 % eventually got access to high school general track (versus 56.9 % for the 1995 cohort) - see Caille (2014) for more details on secondary education trajectories. For the general education high schools we focus on, students apply to schools and are selected based on their home address, their performance in middle school, and other criteria such as their socioeconomic background. The exact assignment procedure depends on the administrative region considered, especially regarding the importance of residential proximity: in several regions, the home address is associated with only one sector high school, and pupils have an absolute priority in this sector school; in others (such as Paris), a student is more likely to be accepted in a high school within a (wider) residential district.

In the first year of general high school (10$^{\text{th}}$ grade), students follow a common curriculum, even if they can choose a few optional courses, as in *collège*. However, at the end of the school year, they apply to both a track and a major- the final decision is taken by school boards. For the student, this decision determines the next two years, the examinations taken at the end of high school, and plays a large part in their upper education opportunities. The two main tracks are general and technological, though students can also be redirected to vocational studies. These tracks are divided in majors. For instance, the general track can be specialized in either science (S), economic and social sciences (ES) or literature (L). The technological track is split in six main tracks which are more narrowly defined : business administration sciences and technology (STMG), industrial sciences and technology (STI2D), design and arts technology (STD2A), laboratory sciences and technologies (STL), medical and social sciences and techniques (ST2S), and hotel management (STHR). Figure 1 plots in the top panel the share of each major among all general and technological *Bac* takers, whereas the bottom panel features the average *DNB* grades of students in each major across time. Tracks and majors are strongly segregated in terms of students' prior school performances and social origins: high achievers are overrepresented in the general track, in particular in the S major. We can see that the average percentile rank at *DNB* is at least 10 points higher among students who eventual take *Bac* in the scientific major, than among any other major. As a matter of fact, the S major is deemed to be the most demanding and the one offering the best opportunities to carry on in upper education. If students are unhappy with the school board ruling about their track and major, they can repeat grade 10 to have another chance to access the track and major they aspire to. Thus the repetition rate is 10 % that year, as compared to 5 % on average in middle school (Ly and Riegert, 2014). The success rates in the different *Bac* tracks and majors are eventually high, conditional on reaching the stage of taking the exam: in 2016 it was 88.5 % for all tracks (vocational included), 91.4 % for the general track, and 91.6 % for the S major of the general track.

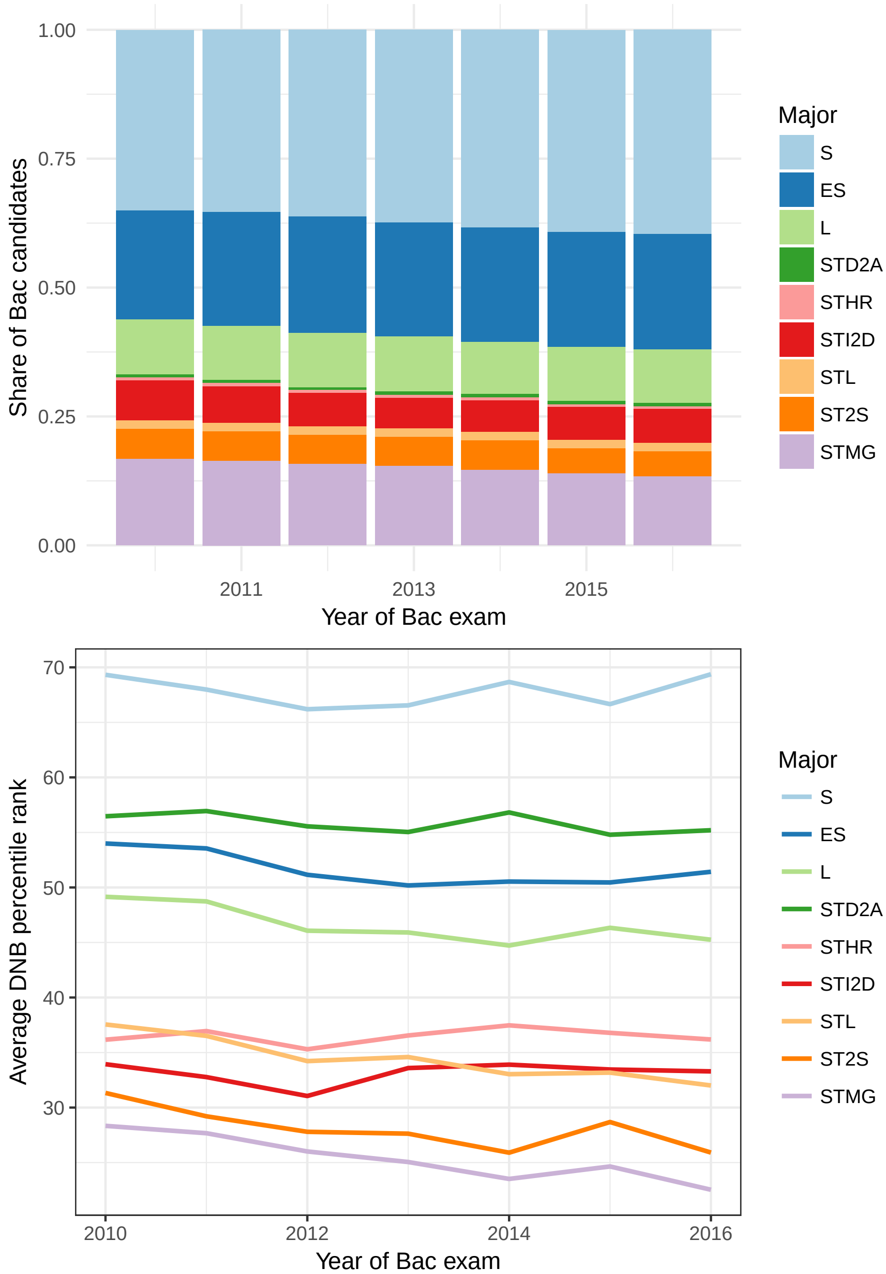
1.2 Data

We rely on exhaustive data from the French department of Education covering school tracks for all students in secondary education public and publicly-funded private schools: the *Fichiers Apprenants* dataset.[3] This exhaustive dataset retrieves for school years 2009-2010 to 2015-2016

[2]Except for a few subjects, such as French, that are taken a year before, or foreign languages, that can be the object of oral tests.

[3]In 2011-2012, 99.7 % of all primary and secondary education students were taught in public and publicly-funded private establishments (department of Education calculations).

Figure 1: Evolution of the distribution and initial academic level of pupils across majors



Sources: *Fichiers Apprenants*. Lecture: 35 % of general/technological *Baccalauréat* takers were enrolled in the scientific track in 2010, and their average DNB percentile rank was 69. Note: the names of some of majors in the technological tracks were modified in 2012, the names indicated here are the current ones.

various information on students, including the school they go to, their class, the curricula they are taught, their age and their socioeconomic background. We are also able to merge *Fichiers Apprenants* with the datasets keeping track of the two secondary education exams: *DNB* and *Bac*.

Hence, to study peer effects in 12$^{\text{th}}$ grade, we first extract from *Bac* files exam results obtained by students to the first test session (before catch-up exams but after a first round of grade adjustments) from 2010 to 2016. We are able to identify for each *Bac* taker her fellow classmates in the final year, and control for her own individual characteristics. Our main outcome is the percentile rank at the *Bac* exam. In our analysis we focus on students in all nine majors of the general and technological *Bac*.[4] Students' percentile rank at *DNB* is used as a measure of initial school ability when entering high school. We also use this variable to characterize the academic level of peers.[5]

The first column of table 1 displays some descriptive statistics of our original sample, which comprises 2,784,409 student observations. The share of girls among general and technological Bac takers is 54 %, and 23 % of the sample is one or more year late to take the exam (normal timing is to take Bac at 18), having repeated at least one grade. 70 % of pupils considered are enrolled in the general track (vs. technological), and 20 % attend a publicly-funded private school.

[4]The percentile rank does not distinguish between the different majors, however as indicated later, we will allow this outcome to depend flexibly on *DNB* rank depending on the major considered.

[5]We restrict our sample to students for which the *DNB* score is not missing. Especially, we do not recover *DNB* scores when more than five years have elapsed between *DNB* and *Bac*, that is, when the student has repeated a grade three times or more between *DNB* and *Bac*.

Table 1: Descriptive statistics for students taking *Baccalauréat* between 2010 and 2016

	All students		Exogenous class (EC) students		Bottom half students in EC		Top half students in EC	
	Mean	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*
				Individual variables				
Bac Score (percentile rank)	50.7	(28.8)	48.3	(28.1)	37.5	(24.2)	63.2	(26.3)
DNB Score (percentile rank)	50.4	(28.9)	45.0	(28.3)	24.2	(14.4)	73.8	(14.3)
Girl	0.54	(0.50)	0.54	(0.50)	0.53	(0.50)	0.55	(0.50)
Late student	0.23	(0.42)	0.27	(0.44)	0.38	(0.48)	0.12	(0.32)
Foreign student	0.02	(0.14)	0.02	(0.15)	0.03	(0.17)	0.01	(0.12)
High social status	0.33	(0.47)	0.29	(0.45)	0.21	(0.41)	0.39	(0.49)
- B social status	0.16	(0.37)	0.16	(0.37)	0.16	(0.36)	0.17	(0.37)
- C social status	0.26	(0.44)	0.26	(0.44)	0.28	(0.45)	0.24	(0.43)
- low social status	0.26	(0.44)	0.29	(0.45)	0.35	(0.48)	0.20	(0.40)
				Class-level variables				
General track	0.70	(0.46)	0.58	(0.49)	0.39	(0.49)	0.83	(0.37)
Private school	0.20	(0.40)	0.19	(0.39)	0.14	(0.34)	0.25	(0.43)
Class size	28.3	(6.0)	27.9	(5.7)	27.3	(5.8)	28.5	(5.5)
Average class DNB Score	50.4	(19.2)	45.0	(18.8)	36.0	(15.3)	57.5	(15.8)
Share of DNB low achievers	0.25	(0.24)	0.31	(0.25)	0.42	(0.24)	0.16	(0.17)
- DNB Q2 in the class	0.25	(0.13)	0.27	(0.12)	0.30	(0.12)	0.23	(0.12)
- DNB Q3 in the class	0.25	(0.13)	0.23	(0.13)	0.18	(0.13)	0.30	(0.11)
- DNB Q4 in the class	0.25	(0.23)	0.19	(0.21)	0.10	(0.14)	0.31	(0.22)
Share of A status	0.33	(0.20)	0.29	(0.19)	0.23	(0.16)	0.37	(0.21)
- B status in the class	0.16	(0.08)	0.16	(0.08)	0.16	(0.08)	0.16	(0.09)
- C status in the class	0.26	(0.11)	0.26	(0.11)	0.28	(0.11)	0.25	(0.11)
- D status in the class	0.26	(0.17)	0.29	(0.17)	0.33	(0.17)	0.23	(0.15)
Number of observations	2,784,409		1,203,870		699,352		504,518	

Source: *Fichiers Apprenants*. Lecture: on average, a student in final high school grade, has 25 % of peers in the bottom quarter of the DNB score distribution of all students taking a general or a technical *Baccalauréat*. This share goes up to 31 % if the student belongs to a high school passing the exogeneity test- 42 % if the student is himself a DNB low-achiever and 16 % in the other case.

2 Methodology

2.1 Identification strategy

We examine the effect of classroom peers' ability as measured by their *DNB* percentile rank on pupils' outcome at *Bac*: by focusing on the effect of the *initial* grade of peers, we do not have to handle the reflection problem (Manski, 1995) triggered by using the outcome of peers as the peer ability variable.

School × major fixed effects Our main issue here is therefore the problem of correlated effects: since pupils are not randomly allocated across majors and schools, we must be careful not to interpret as peer effects what actually reflects the correlation between one's initial level and the initial level of peers. Indeed, students appear to be segregated between high schools and majors, depending on their initial ability. For instance, pupils in the top half of the initial ability distribution are exposed, within their high school and major, to 62 % of pupils also belonging to the top half of the distribution. This figure is only 34 % for pupils who are in the bottom half of the distribution. In order to deal with this issue, we use school × major fixed effects. We therefore rely on intra-school, intra-major variability in peer ability, both between cohorts and between class groups. Several papers have a similar strategy, either focusing on the composition of the school × grade, hence using between cohort variation within schools;[6] or examining the effect of the class peer group, taking benefit from between-class within-school variability.[7] Ammermueller and Pischke (2009) use such a methodology for France among other countries for the primary school level. Vigdor and Nechyba (2006) compare the within-school, between-class methodology with results obtained exploiting changes in school composition associated with the redrawing of attendance zone boundaries, which is more likely to be random. The peer effects measured with the school-fixed-effects methodology are somewhat reduced when using this alternative source of variation, reflecting the correlated effects issue, but the impact of average peer level remains positive and significant.

Relying on between-classes variability may indeed prove problematic in the case of non-random allocation of pupils across classrooms, within a given school (and major): in this case, the school × major fixed effects introduced may not be sufficient to eliminate correlated effects, and the measured peer effect may capture the correlation of pupils' ability across classes. We may first underline that the largest part of academic sorting within a given high-school takes place across majors. This sorting takes place ahead of the final year of high school, that is in the 11^{th} grade. For pupils taking *Bac* in 2016 for instance, 47 % of average classroom *DNB* score variability is explained at the school level, and 94 % is explained at the school *and* the major level. Note also that what would be problematic here is the correlation of *unobserved* pupils' ability across classes, given that we already include several important observable characteristics that capture the criteria upon which the allocation of pupils to classes may be determined.

These controls include percentile rank at *DNB*, which we allow to impact individual *Bac* percentile rank in a flexible manner, sex, age (both age at *DNB* and number of years since *DNB*) and parental socioeconomic status, for both the mother and father and at a very detailed level. At the classroom level, we also control for classroom size. Finally, in order to control for potential trends in the relative selectivity of the different majors, we include fixed effects at the cohort × major × region level.[8] The baseline linear-in-means specification would therefore write

[6]See for example Black, Devereux, and Salvanes, 2013; Lavy and Schlosser, 2011; Schneeweis and Winter-Ebmer, 2007.

[7]Ammermueller and Pischke, 2009; Bertoni, Brunello, and Cappellari, 2017; Neidell and Waldfogel, 2010: the latter also introduce family fixed effects

[8]Here the cohort designates the group of pupils taking the *Bac* exam for one given year, regardless of

as follows:

$$Bac_i = b_0 + b_1^m DNB_i + b_2\overline{DNB}_c + b_3X_i + b_4X_c + \alpha_{ms} + \gamma_{ymr} + \epsilon_i \tag{1}$$

for each pupil i in class group c within major m and school s, in region r, and who is taking *Baccalauréat* in year y. Pupils are pooled across all cohorts from 2010 to 2016, and across all majors. Bac_i is the percentile rank at *Baccalauréat* for pupil i, and DNB_i her individual rank at DNB. Note that the relation between individual percentile rank at DNB and at Bac is allowed to depend on the major considered (for instance, the subjects taken at DNB matter differently for each Bac major considered). $\overline{DNB}_c$ is our baseline peer variable: that is, in the linear-in-means model, the average percentile rank at DNB among classroom peers in the final year of high-school.[9] X_i and X_c designate the control variables, other than the individual DNB percentile rank, respectively at the individual and classroom level. α_{ms} is the school × major fixed effect. It covers all classes c for a given major and high-school, so that we rely on variation across cohorts. γ_{ymr} is the cohort × region × major fixed effect mentioned above.

Restricting the sample to high schools with no statistical evidence of sorting
For the school-fixed-effects approach to be valid, we need to have students assigned randomly into classes *conditional* on their high school, their major, and observed characteristics. If not, our results may capture correlated effects rather than actual exogenous peer effects. To consolidate our identification strategy, we restrict the sample to a set of high schools which seem to assign students to classrooms in a random manner within a given major.[10] We proceed separately for each major, so that we consider it possible that a high-school has a non-random assignment process of pupils to classrooms for some majors only.

The spirit of this restriction is the same as in Ammermueller and Pischke (2009), however here we make use of the repeated information about class formation across the seven years available. The general idea is the following (the procedure is described in more details in the Appendix section): for each given school s × major m × cohort y, if there are at least two classes, we test the hypothesis H_0^{msy} of random allocation of students across classes. That is to say, we model students' probability in a given school × major × cohort msy comprising C classes ($C \geq 2$), to belong to a specific class c depending on their DNB score. With $class_i$ the variable defining the actual classroom of student i, (and omitting indexes msy for the sake of clarity), and taking class C as the reference, we thus estimate the following multinomial logit model:

$$P(class_i = c/DNB_i) = G(\alpha_c + \beta_c DNB_i), \forall c, \; c = 1...C-1$$

For school × major × cohort msy, the null hypothesis H_0^{msy} may therefore be written as follows :

$$H_0^{msy} : \; \forall \beta_c, \; c = 1...C-1, \; \beta_c = 0$$

H_0^{msy} is our hypothesis of random allocation of pupils across classrooms. For each year y, the decision to reject H_0^{msy} or not will depend on the actual allocation process of school s within

their previous academic trajectory. The region level considered is that of the *Académie*, an administrative unit at which level a large part of educational decisions are made. There are 26 academies in metropolitan France.

[9] In the chosen specification, the peer variable $\overline{DNB}_c$ includes pupil i's own score in order to be consistent with the subsequent approach in terms of class types. This does not matter here since we also control by the individual DNB percentile rank: using the class leave-out mean rather than the mean changes the coefficient on the peer variable by less than 2 %.

[10] By definition, this criterion eliminates high school with only one class within a given major.

major m, but also on natural variability. However, if H_0^{msy} is rejected year after year, there is good reason to think that school s has a general behavior of sorting pupils across classes (within major m). In the opposite case, if we fail to reject the null hypothesis year after year, then we may believe that there is no ability sorting behavior.[11] We therefore build a decision rule across cohorts, based on rejection decisions for each single year (cf. Appendix).[12] For instance, one decision rule could be to accept at most two rejections (out of the seven tests for years 2010...2016) at the 12.88 % level. We show that this corresponds to a 5 % type I error for the *general* (ie. across cohorts) random allocation hypothesis for the school × major considered. An alternative decision rule also corresponding to a 5 % type I error in the general test, is to accept at most only one rejection among the seven yearly tests, but at the 5.35 % level. That is, we dismiss from our sample the schools × majors for which H_0^{msy} is rejected twice or more at the 5.35 % level. Our sample of interest will be based on this latter decision rule.

The corresponding sample is described in column 2 of Table 1 – we refer to it as the *EC sample* (for exogenous class). 1,203,870 observations remain in the EC sample, out of 2,280,214 for which the random allocation test was relevant (there are 504,205 observations corresponding to a single class within a given school and major). The rejection rate of the null hypothesis of random allocation is 37 % among schools x majors, but the rejected schools × majors comprise a higher number of classes (3.2 vs. 2.5 in the EC sample). Finally, Table 1 shows that the number of students per class is slightly smaller in the EC sample than in the whole sample (27.9 vs. 28.3). The difference is sufficiently small to be confident that the non-rejection of the test in our EC sample is not due to a lack of statistical power.

The 37 % rejection rate varies greatly depending on the major considered: it is only around 20 % for the technological track, but as high as 58 % for the scientific major (and 31 to 35 % for the other general ones): the most prestigious major seems to resort more widely to academic tracking. This explains the lower share of general track pupils in the EC sample than in the whole population. In turn, this composition effect accounts for some differences between the two samples that may be noticed in Table 1: lower initial academic level, higher proportion of late students.

We aim to document the general factors which may lead a school to resort to non-random allocation. We argue that schools belonging to our subsample are more often characterized by the lack of school competition in the neighboring area, whether because they are located in sparser areas and/or because of the absence of private schools in the surroundings. The competition with nearby private schools is a rationale which has already been put forward in the literature (see Ly and Riegert, 2014). Indeed, private schools do not really have to sort pupils across classes, since they are already able to recruit their students selectively: from the perspective of parents, this already guarantees a certain homogeneity in academic level. However, since this is not the case for public schools, these may instead implement some kind of tracking in order to attract good pupils.[13] We focus on the sorting behavior for the scientific major (it is of specific interest since we saw that it is the more concerned by academic sorting). In Table 2 we regress the probability for a high school to form non-random classes in the sense of our exogeneity test on various characteristics: the distance to the closest private school[14], the median wage category

[11] This could also be due to a lack of statistical power. However as we will see, there is no substantial difference in class size between both types of schools.

[12] The spirit of this approach is somewhat similar to the Bonferroni correction used for multiple testing issues.

[13] An example of sorting mechanism is the presence of one particularly outstanding class, meant to prepare pupils for higher education (especially for *Classes préparatoires*).

[14] Here we only focus on public schools.

in the city, the population density category in the city. After controlling for the median wage in the high school city, we find the further away from a private competitor and the smaller population density, the lower chance a high school assigns students in classes based on their prior performance. These results are in line with the idea that tracking policies are related to a high level of school competition whether from private schools or from any alternative school options in the area because of low density of inhabitants.

Table 2: Random pupil allocation and high school environment for the S major

	Probability of academic sorting of pupils	
Distance to nearest private school	-0.009^{**}	(0.004)
Municipality in the 1st quartile for median income	0.193	(0.190)
Municipality in the 3rd quartile for median income	0.045	(0.187)
Municipality in the 4th quartile for median income	0.256	(0.205)
Municipality in the 1st quartile for density	-0.466^{**}	(0.194)
Municipality in the 3rd quartile for density	0.113	(0.201)
Municipality in the 4th quartile for density	-0.171	(0.205)
Constant	0.801^{***}	(0.184)
Observations	975	

Standard errors in parentheses - * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.2 Heterogeneous and non-linear effects

We wish to go further than the linear-in-means model 1. First, we introduce heterogeneous effects by allowing the effect of peers to be flexible depending on pupil i's initial academic level. Four types of pupils Q1 to Q4 are defined, depending on their position in the DNB score distribution *in the whole sample* (and not just the EC sample). Q1 corresponds to the lowest initial ability group and Q4 to the highest. The last two columns of Table 1 outline the difference of situation between pupils in the top and bottom half of the initial ability distribution (Q3 and Q4 vs. Q1 and Q2), when focusing on the EC sample. As expected, the proportion of pupils who have repeated a grade is much lower in the top half, as for the share of low social status pupils. What is of greater interest to us here is the classroom composition that pupils of different ability types are faced with, *across* schools and majors (if random allocation holds, there should be no systematic differences *within* schools and majors): high-ability pupils have 61 % high-achievers in their class, whereas pupils in the bottom half of the ability distribution only have 28 % high-achievers among their peers. Class size is slightly smaller in the bottom ability group.

In order to go beyond the effect of the average peer level, we also enrich the linear-in-means model with a direct measure of academic level diversity in the classroom (we use the standard-error, but the results are robust to the inclusion of other measure eg. entropy index, inter-decile or interquartile ratio). We also allow this measure to have heterogeneous effects depending on the ability group of the pupil impacted. We include the standard error and average level simultaneously in our model, since for instance a higher average level may be mechanically correlated with a lower dispersion (as is observed in our data). The inclusion of the standard error could also lead to issues related to class size: in smaller classes, the standard error will be much more sensitive to outliers (Lyle, 2009). This is taken care of since we are also controlling for class size. The corresponding heterogeneous effect model is therefore:

$$Bac_i = b_0^{m,q} + b_1^{m,q} DNB_i + b_2^q \overline{DNB}_c + b_3^q sd(DNB)_c + b_4 X_i + b_5 X_c + \alpha_{ms} + \gamma_{ymr} + \epsilon_i \quad (2)$$

for student i of ability type $q = 1...4$, with $sd(DNB)_c$ the standard deviation of DNB in class c and other notations identical to model 1. The heterogeneous effects model implies that both peer variables coefficients (b_2 and b_3) are allowed to depend on ability type q. Note that we also allow the relation between individual DNB and Bac (the b_1s) to vary with ability type q, as well as with major m.

Defining class types A model including the mean and standard error of peer ability does not inform us directly on the type of classroom composition from which a given pupil benefits most. Indeed, the marginal effect of a given peer measure may be non-linear and differ depending on the classroom type considered: for example, the marginal effect of the ability standard deviation may be low within classrooms that are already very diverse. In particular, heterogeneous effects depending on pupils' type must be interpreted with caution, since classroom composition differs when considering high-achievers and low-achievers *across* schools (1, see above). Thus, do effects differ with ability group because $Q1 - Q4$ pupils are actually impacted differently, or because they are distributed in different types of classrooms *across* schools?

To address this issue and identify the type of classroom environment most beneficial to each type of pupil, our proposed solution here is to develop a typology of classes and directly examine the effect of these different types of classroom environments on pupils with different ability levels. Compared to the canonical heterogeneous peer effects of Hoxby and Weingarth (2005), here we look at the effect of switching from one class type to another on pupil i depending on their ability, rather than looking at the marginal effect of pupils of a given ability depending on i's own ability. The method used for this classification of classes is that of partitioning around medoids ($k - medoids$), which is a more robust version of $k - means$. This non-supervised learning algorithm consists in identifying a fix number of clusters around representative points (medoids) such as to minimize within-cluster sum of distances. At each step of the algorithm, medoids are updated to be as representative as possible in their cluster, and points are reassigned to the new closest medoid. The process stops when there is no more gain to change medoids and cluster boundaries. Variables from which the partitioning is derived are the average level at DNB and the standard error of the classroom. Both classroom-level variables are standardized across all classrooms.

We do not use any information criteria to determine the number of clusters here (such as silhouette criterion): there is no reason to believe in a "natural" underlying clustering in our data, and we favor interpretable and balanced clusters. Once the class clusters $n = 1 \ldots N$ are defined, the estimation equation is the following, where the coefficients of interest to us are the $b_2^{q,n}$, which measure the effect of being in a classroom of cluster n (rather than the reference cluster) for a pupil of type $q = 1...4$:

$$Bac_i = b_0^{m,q} + b_1^{m,q} DNB_i + \sum_{\substack{n=1 \\ n \neq ref}}^{N} b_2^{q,n} \mathbb{1}_{type(c)=n} + b_3 X_i + b_4 X_c + \alpha_{ms} + \gamma_{ymr} + \epsilon_i \quad (3)$$

Here again, we consider pupil i in class c of type n, within major m, school s, region r and cohort y. The relation between percentile rank at DNB and percentile rank at Bac is again flexible depending on major m and ability quartile q. X_i and X_c are our control variables at the pupil and classroom level, α_{ms} is the $school \times major$ fixed effect and γ_{ymr} the $cohort \times major \times region$ fixed effect.

3 Results

3.1 From the linear-in-means model to heterogeneous effects

Table 3 presents results for several specifications of our baseline peer effects model. Columns (1) to (3) focus on our *EC* sample within which we suppose allocation to classrooms to be random. Column (1) shows estimates for an extension of the linear-in-means model 1, where the standard deviation of peer initial ability is also introduced. Both average peer ability and diversity of peer ability appear to have a positive impact on the result at *Baccalauréat*. The magnitude of the average peer effect (+0.083 *Bac* percentile rank for +1 peer average percentile rank at *DNB*), lies within the range of previous findings in the empirical literature: in the review carried out by Sacerdote (2011), the effect of a 1.0 point raise in average peer score ranges is most often comprised between 0.05 and 0.4 points. For a given average level of peers, class standard error in initial level also exerts a significant positive effect, although smaller. Control variables impact the *Bac* score in the direction that would be expected: the age at DNB as well as the lapse of time between DNB and *Bac*, which reflect the fact that pupils repeated a grade before or after DNB, both impact negatively the *Bac* score for a given DNB grade. Being female or a French national has a positive impact. Class size exerts a significant negative effect: reducing class size by one pupil has roughly the same effect as increasing peer average initial level by one percentile rank.

The effect of the average class level, and even more so the effect of the class diversity in terms of initial academic levels, vary widely depending on the type of pupil impacted, as shown in column (2) which introduces heterogeneous effects depending on pupil i's initial academic level. Column (2) therefore corresponds to specification 2. The effect of raising the peer average percentile rank at DNB by 1 point ranges from 0.029 for the high-ability pupils (Q4), to 0.140 for the lowest-ability pupils (Q1): low-ability pupils seem to be the ones who benefit the most from a raise in the ability level of their peers. For a given average level, the effect of diversity is negative for the lowest-ability pupils, whereas it is positive for the high-ability ones. Here we encounter the identification issue mentioned in Section 2: is diversity detrimental to low-ability pupils because they have low ability, or because they are distributed across schools in classes which may already have high levels of diversity? This will be addressed in the subsection 3.3.

Columns (3) and (4) respectively illustrate the need for including school $\times$ major fixed-effects, and for limiting our analysis to the *Exogenous class* sample, in order to tackle the sorting of pupils respectively between schools and majors, and between classes. When not including school fixed effects in column (3), the estimates for mean peer effects are much higher, suggesting strong positive sorting on unobservables across high schools, especially for high-ability students. Without fixed effects, the impact of diversity also appears more detrimental for low-ability pupils and neutral for high-ability ones. The bias induced by sorting between classrooms seems to go in the same direction, although it appears generally smaller (column (4)): here model 2 is estimated including school $\times$ major fixed effects, but is estimated using the whole sample, irrespective of whether schools passed the random allocation test or not.

Table 3: Introducing heterogeneous and diversity effects in the linear-in-means model

	(1)		(2)		(3)		(4)	
Class average level	0.0825***	(0.00417)						
Class standard error	0.0165*	(0.00703)						
Class average level # Q1			0.140***	(0.00643)	0.318***	(0.00544)	0.192***	(0.00402)
# Q2			0.124***	(0.00565)	0.317***	(0.00469)	0.195***	(0.00339)
# Q3			0.0983***	(0.00586)	0.306***	(0.00492)	0.205***	(0.00341)
# Q4			0.0287***	(0.00726)	0.249***	(0.00643)	0.175***	(0.00404)
Class std. error # Q1			-0.0892***	(0.0121)	-0.142***	(0.0118)	-0.126***	(0.00846)
# Q2			-0.0873***	(0.0126)	-0.196***	(0.0123)	-0.146***	(0.00822)
# Q3			-0.0327*	(0.0134)	-0.195***	(0.0130)	-0.0871***	(0.00822)
# Q4			0.203***	(0.0153)	-0.0137	(0.0144)	0.136***	(0.00852)
Female	2.462***	(0.0438)	2.468***	(0.0438)	2.475***	(0.0441)	1.855***	(0.0281)
Age at *DNB*	-6.375***	(0.0498)	-6.371***	(0.0498)	-6.507***	(0.0503)	-6.175***	(0.0336)
DNB-Bac time lapse	-4.139***	(0.0412)	-4.153***	(0.0412)	-4.228***	(0.0415)	-4.782***	(0.0285)
French national	4.212***	(0.134)	4.207***	(0.134)	4.276***	(0.135)	4.402***	(0.0911)
Class size	-0.0664***	(0.00468)	-0.0667***	(0.00468)	-0.112***	(0.00385)	-0.0505***	(0.00302)
Constant	145.3***	(9.663)	147.0***	(9.662)	118.8***	(1.078)	140.5***	(5.779)
School fixed effects	*Yes*		*Yes*		*No*		*Yes*	
Additional controls	*Yes*		*Yes*		*Yes*		*Yes*	
Sample	*EC*		*EC*		*EC*		*All*	
N	1,201,190		1,201,190		1,201,190		2,777,672	
adj. R^2	0.316		0.316		0.403		0.353	

Standard errors in parentheses - * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. On top of school × major fixed effects and cohort × major × region fixed effects, additional controls include parents' social background (32 groups), and initial DNB percentile rank interacted with major x DNB quartile Q1-Q4. Pupils of types Q1 to Q4 range from lowest to highest achieving in terms of *DNB* score.

3.2 Defining class types

We use K-medoid partioning to identify eight clusters of class compositions, among all classes of all schools in the general and technological tracks. Figure 2 characterizes those clusters in terms of the average *DNB* percentile rank, the *DNB* percentile rank standard error, and the share of each type of students $Q1...Q4$ in classes. Plotting the within-class standard deviation against the class average (Figure 2 top panel) allows us to see clearly how classes with a very high or very low average necessarily have a low standard deviation.

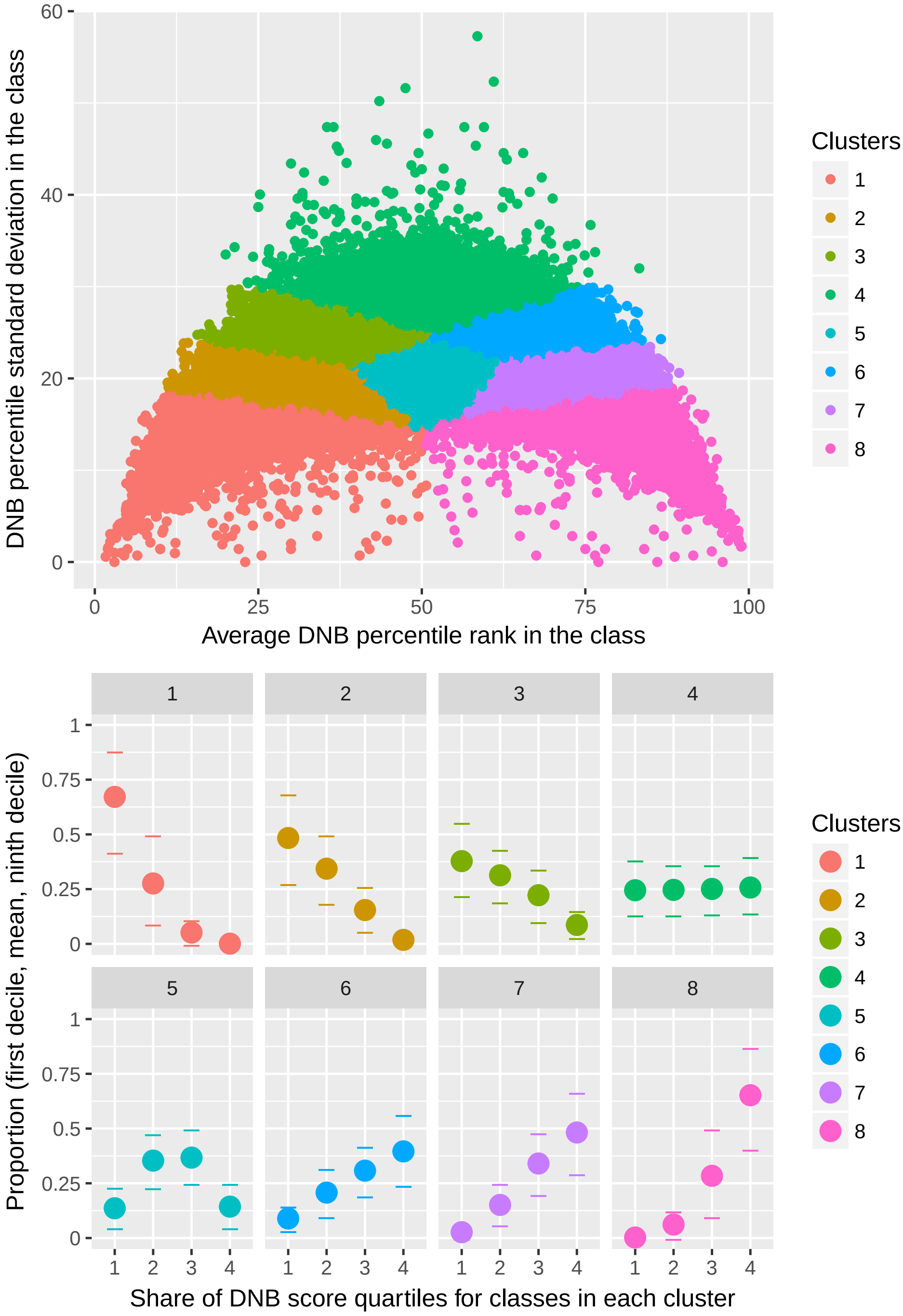
Cluster 5, which is central both in terms of average and standard deviation of DNB, will be our reference cluster in specification 3.[15] It comprises a large share of pupils from the center of the ability distribution, whereas pupils from both ends are underrepresented (Figure 2 bottom panel). Cluster 4 is also central in terms of average ability level within the classroom, however the within-class ability standard deviation is much higher. The bottom panel of Figure 2 shows that students of each $Q1-Q4$ type are almost equally represented in cluster 4 classes.

Clusters 1, 2 and 3 are all characterized by a rather low average ability, with an overrepresentation of pupils from Q1. From cluster 1 to cluster 3, average ability increases, and so does ability diversity. Indeed in cluster 1, a very high proportion of low-ability pupils (around 50 to 80 %) leads both to a very low average ability and a very low diversity of pupils. Moving to clusters 2 and 3, the proportion of $Q2$ remains roughly the same, whereas $Q1$ pupils are progressively swapped with pupils from the top half of the distribution.

Oppositely, clusters 6, 7 and 8 exhibit a relatively strong presence of high achieving students. Cluster 8 has the highest average ability, with 40 to 80 % Q1 pupils, which translates into a very low diversity: it is the counterpart of cluster 1 at the other end of the ability distribution.

[15] Partitions with 7 clusters or less do not provide such a "central" cluster, which is why we favored the 8-clusters partition in terms of interpretation.

Figure 2: Mean and standard error of DNB percentile rank, and share of students from each ability type, in K-medoid clusters



3.3 The heterogeneous effect of class types on pupils

Table 4 shows the set of coefficients $b_2^{q,n}, q = 1...4, n = 1...8, n \neq 5$ from model 3. As mentioned earlier, median cluster 5 is chosen as the reference group. The first two columns serve as reminders of the average characteristics of each cluster. Figure 3 plots the effects for each cluster and type of pupil impacted for easier comparison of the relative effects between clusters.

Effects of belonging to a given class type are clearly heterogeneous according to the student's own ability. Let us first consider the effect of moving from the median class (cluster 5, our reference) to a class where virtually the DNB score distribution was shifted to the right, that is to say with more *DNB* low achievers and less high achievers (cluster 1 to 3). Effects of such a rightward distortion of the distribution has negative and significant effects for Q1 and Q2 students. However, it has no significant effect for Q3 and Q4 students (although standard errors are large for the latter group, due to their under-representation especially in classes of clusters 1 and 2). We may relate this to the *Bad apple* model discussed in Hoxby and Weingarth (2005), which states that an increase in the number of bottom-achieving students has a negative effect on the achievement of students due to their disruptive behavior which in turn *"triggers disruptive behavior from children who would otherwise be attentive"*. Here, it appears that pupils are not affected equally, and that who it is pupils also at the bottom of the ability distribution who react the most to the disruptive behavior of their peers.

Oppositely, leftward modification in the class DNB score distribution is beneficial to Q1-Q3 students, who see their *Bac* result increase by 1 to 5 percentile ranks when switching from a class in cluster 5 to a class in cluster 6 to 8. This is in stark contrast with top-ability students who are negatively impacted by the change in class composition implied by switching from cluster 5 to clusters 7 and 8 (-0.562 and -1.939 *Bac* percentile ranks respectively). This result is in line with the idea that invidious comparison with peers can be detrimental to individual performance: high achievers can suffer from being in class with too high a proportion of other high achievers.

Finally, moving from cluster 5 to classes of cluster 4 with the highest diversity seems to be slightly beneficial for all types of students – except for the lowest-ability ones, for which the effect is not significant. The only other cluster which appears to be beneficial for all (or at least non-detrimental) is cluster 6. But contrary to cluster 4, starting from a sample of students representative of the whole population, the cluster 6 class type could not be achieved for all classes.

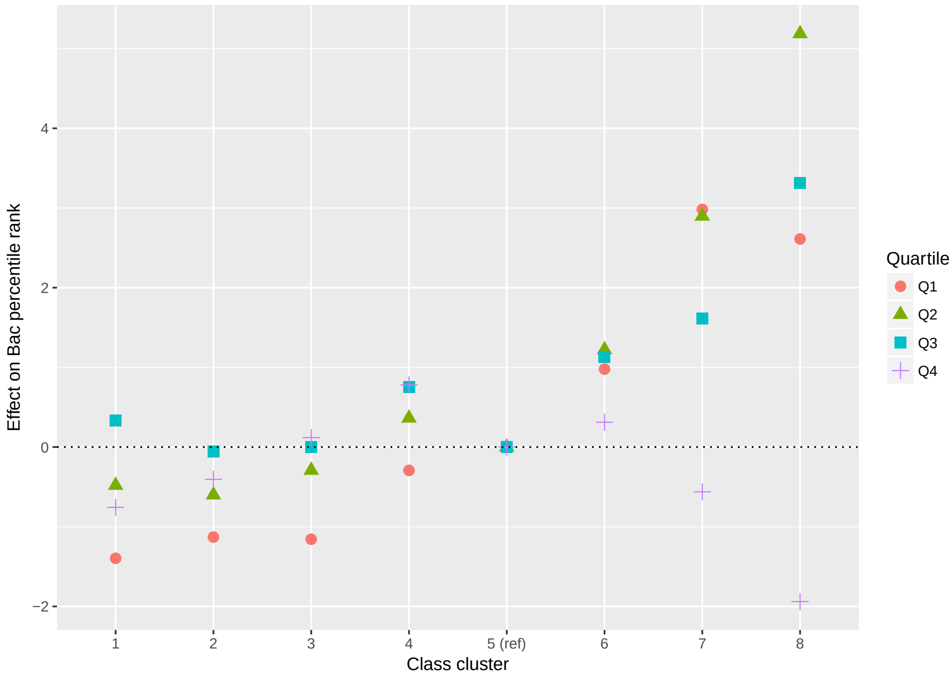
Table 5 displays the effect for a student of a given initial level (Q1 to Q4) to switch from a class of cluster 5 for each of the other seven types, depending on the student's gender. On the whole, effects are rather close for boys and girls, and previous comments remain valid. Two points are however worth noting. First, for pupils of type Q2 and Q3, the positive effect of moving to a more diverse class (from cluster 5 to cluster 4) is present only for girls. For boys, the benefit of being in a more diverse environment appears significant only for the top-ability student. Second, for the top quartile, the effect of belonging to a class with a large share of high achievers seems to be more detrimental for girls than for boys, implying that they would be more subject to the invidious comparison effect.

Table 4: Peer effects estimated with class type and student performance interactions

Cluster	Average score	Score Std error	Q1	Q2	Q3	Q4
1	− − −	−	-1.396***	-0.481*	0.332	-0.757
			(0.216)	(0.196)	(0.305)	(1.558)
2	−−	∼	-1.130***	-0.601***	-0.0583	-0.405
			(0.203)	(0.166)	(0.197)	(0.423)
3	−	+	-1.157***	-0.292	-0.000926	0.119
			(0.192)	(0.150)	(0.162)	(0.252)
4	∼	++	-0.293	0.363*	0.753***	0.779***
			(0.207)	(0.163)	(0.162)	(0.204)
5	∼	∼		*(ref)*		
6	+	+	0.978***	1.222***	1.132***	0.311
			(0.260)	(0.173)	(0.156)	(0.196)
7	++	∼	2.982***	2.897***	1.613***	-0.562**
			(0.425)	(0.209)	(0.170)	(0.207)
8	+ + +	−	2.611	5.187***	3.316***	-1.939***
			(2.023)	(0.411)	(0.237)	(0.243)
N			1,201,190			
adj. R^2			0.316			

Standard errors in parentheses - * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Controls are identical to those used in Table 3. Pupils of types Q1 to Q4 range from lowest to highest achieving in terms of *DNB* score.

Figure 3: Effects of belonging to cluster of a given type



Sources: *Fichiers Apprenants*. Lecture: Belonging to cluster 1 (rather than reference cluster 5) has a 0.332 effect on the percentile rank at *Bac* for pupils in Q3. Note: the displayed coefficients are those of Table 4.

Table 5: Peer effects estimated with class type and student performance interactions separately for girls and boys

	Q1		Q2		Q3		Q4	
Cluster	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
1	-1.361***	-1.387***	-0.317	-0.734*	0.249	0.305	-1.232	-1.694
	(0.296)	(0.316)	(0.267)	(0.288)	(0.409)	(0.458)	(1.896)	(2.664)
2	-1.060***	-1.079***	-0.510*	-0.650**	-0.0681	-0.0938	-1.177*	0.619
	(0.280)	(0.295)	(0.229)	(0.242)	(0.270)	(0.290)	(0.552)	(0.655)
3	-0.961***	-1.208***	-0.172	-0.315	0.206	-0.251	-0.0754	0.456
	(0.263)	(0.283)	(0.199)	(0.227)	(0.210)	(0.252)	(0.309)	(0.425)
4	-0.194	-0.192	0.678**	0.0469	1.094***	0.311	0.691**	0.816*
	(0.279)	(0.311)	(0.212)	(0.253)	(0.205)	(0.259)	(0.249)	(0.350)
5				*ref*				
6	0.917*	1.035**	1.382***	1.042***	1.360***	0.905***	0.0829	0.793*
	(0.362)	(0.377)	(0.233)	(0.258)	(0.202)	(0.242)	(0.242)	(0.330)
7	2.857***	2.809***	2.835***	2.708***	1.717***	1.441***	-0.743**	-0.0367
	(0.637)	(0.578)	(0.294)	(0.300)	(0.227)	(0.258)	(0.259)	(0.341)
8	3.259	2.418	5.165***	4.758***	3.485***	2.953***	-2.049***	-1.434***
	(2.992)	(2.760)	(0.628)	(0.551)	(0.335)	(0.341)	(0.315)	(0.388)
N	649,452	551,738	649,452	551,738	649,452	551,738	649,452	551,738
adj. R^2	0.349	0.271	0.349	0.271	0.349	0.271	0.349	0.271

Standard errors in parentheses - * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Controls are identical to those used in Table 3. Pupils of types Q1 to Q4 range from lowest to highest achieving in terms of *DNB* score.

4 Conclusion

We use French exhaustive administrative data through 2010 to 2016 and exploit within school and major variation to assess peer and class composition effects on *Baccalauréat*. After restricting our analysis to a subsample for which the school fixed effect specification hypotheses are credible, we identify eight class types to capture class composition as a whole. This strategy allows us to disentangle heterogeneity and non-linearity of peer effects.

We find peer and class composition effects vary greatly according to the student's ability. Contrary to what could be expected, top-ability students do not remain unaffected by their class environment. As a matter of fact, we find that students are negatively impacted by classes when their own ability level is over-represented, compared to an average class.

Hence, a large proportion of high-ability peers is especially profitable to low-ability peers, whereas it may even be detrimental to other high prior performance students. Conversely, a high share a low-ability peers is mostly harmful for other low-ability students, and less so to high ability ones. These general conclusions hold regardless of pupils' gender.

These results contrast partially with the ones we obtain using our baseline specification, where the effect of average and standard deviation peer prior performance are interacted with the student's ability. Estimates from this baseline indicate that within-class peer diversity is only beneficial for high ability students. This finding seems to arise from the fact that without taking into account the whole class composition, controlling for the standard error in the class mainly consists for low ability students to consider situations where they are surrounded by many other low achievers. Hence, using class types in the specification allows us to distinguish upward and downward diversity and assess that poor ability students actually benefit from very good classes.

References

AMMERMUELLER, A., AND J.-S. PISCHKE (2009): "Peer effects in European primary schools: Evidence from the progress in international reading literacy study," *Journal of Labor Economics*, 27(3), 315–348.

BERTONI, M., G. BRUNELLO, AND L. CAPPELLARI (2017): "Parents, Siblings and Schoolmates: The Effects of Family-School Interactions on Educational Achievement and Long-Term Labor Market Outcomes," IZA Discussion Papers 11200.

BLACK, S., P. DEVEREUX, AND K. G. SALVANES (2013): "Under Pressure? The Effect of Peers on Outcomes of Young Adults," *Journal of Labor Economics*, 31(1), 119 – 153.

BOOIJ, A. S., E. LEUVEN, AND H. OOSTERBEEK (2017): "Ability Peer Effects in University: Evidence from a Randomized Experiment," *The Review of Economic Studies*, 84(2), 547–578.

BURKE, M. A., AND T. SASS (2013): "Classroom Peer Effects and Student Achievement," *Journal of Labor Economics*, 31(1), 51 – 82.

CAILLE, J.-P. (2014): "Les transformations des trajectoires au collège: des parcours plus homogènes mais encore très liés au passé scolaire et à l'origine sociale," *Éducation & formations*, (85), 6.

DAVEZIES, L. (2004): "Influence des caractéristiques du groupe des pairs sur la scolarité élémentaire," *Éducation et formations*, 72, 171.

GOUX, D., AND E. MAURIN (2007): "Close neighbours matter: Neighbourhood effects on early performance at school," *The Economic Journal*, 117(523), 1193–1215.

HOXBY, C. M., AND G. WEINGARTH (2005): "Taking race out of the equation: School reassignment and the structure of peer effects," Discussion paper, Working paper.

KISS, D. (2013): "The impact of peer achievement and peer heterogeneity on own achievement growth: Evidence from school transitions," *Economics of Education Review*, 37(C), 58–65.

LAVY, V., AND A. SCHLOSSER (2011): "Mechanisms and Impacts of Gender Peer Effects at School," *American Economic Journal: Applied Economics*, 3(2), 1–33.

LAVY, V., O. SILVA, AND F. WEINHARDT (2012): "The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools," *Journal of Labor Economics*, 30(2), 367–414.

LY, S. T., AND A. RIEGERT (2014): "Persistent Classmates: How Familiarity with Peers Protects from Disruptive School Transitions," *PSE Working Papers*, (2013-21).

LYLE, D. S. (2009): "The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point," *American Economic Journal: Applied Economics*, 1(4), 69–84.

MANSKI, C. (1995): *Identification Problems in the Social Sciences*. Harvard University Press.

NEIDELL, M., AND J. WALDFOGEL (2010): "Cognitive and Noncognitive Peer Effects in Early Education," *The Review of Economics and Statistics*, 92(3), 562–576.

Rangvid, B. S. (2007): “School composition effects in Denmark: quantile regression evidence from PISA 2000,” *Empirical Economics*, 33(2), 359–388.

Sacerdote, B. (2011): “Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?,” vol. 3 of *Handbook of the Economics of Education*, pp. 249 – 277. Elsevier.

Schneeweis, N., and R. Winter-Ebmer (2007): “Peer effects in Austrian schools,” *Empirical Economics*, 32(2), 387–409.

Vigdor, J., and T. Nechyba (2006): “Peer effects in North Carolina public schools,” in *In Schools and the equal opportunities problem, ed. Ludger Woessmann and Paul.* MIT Press.

Appendix

Description of the test for exogenous assignment of students in classes within schools

A single-year test does not take advantage of the repeated information we have across years regarding class formation in a given high school. We therefore consider a test which relies on information across the seven years of observations, in order to determine if schools generally exhibit sorting practices, ie. form classes endogenously year after year.

For a given high school s and major m:

- We define the *general* assumption H_0^{ms} that school s has a policy of random allocation of students to classes for the major m.
- We may also define H_0^{msy} for a particular year y, as the testable assumption that allocation of students to classes is random for year y.
- For each year y, and for each high school and major ms we estimate the multinomial regression described in section 2 and define H_0^{msy} for each year accordingly. That is:
- H_0^{msy} : $\forall \beta_c^y,\ c = 1...C-1,\ \beta_c^y = 0$
- H_0^{ms} then corresponds to: $\forall y = 1...7, \forall \beta_c^y, c_y = 1...C_y - 1,\ \beta_c^y = 0$
- Let us suppose that we want P(reject H_0^{ms} / H_0^{ms} true) = 0.05
- We want to base the rejection decision for H_0^{ms} on the 7 separate rejection decisions for the single-year tests H_0^{msy}, $y = 1...7$
- Suppose we define the rejection rule for H_0^{ms} as "rejecting H_0^{msy} twice or more out of the 7 single-year tests". Then how should we set the significance level α for the rejection decision of H_0^{msy} (α =P(reject H_0^{msy} / H_0^{msy} true), such that P(reject H_0^{msy} twice or more / H_0^{ms} true) = 0.05 ?
  - **Suppose H_0^{ms} true.** Then H_0^{msy} is true $\forall y = 1...7$ (high school s practices random assignment in any given year, for major m) and $\forall y = 1...7$, P(reject H_0^{msy})=α and we have :

- $P(H_0^{msy}$ rejected 0 out of 7 trials$) = (1-\alpha)^7$
- $P(H_0^{msy}$ rejected once out of 7 trials$) = 7\alpha(1-\alpha)^6$
- So $P(H_0^{msy}$ rejected twice or more out of 7 trials$) = 1-(1-\alpha)^7 - 7\alpha(1-\alpha)^6$
- We are therefore looking for α such that $1-(1-\alpha)^7 - 7\alpha(1-\alpha)^6 = 0.05$. The corresponding numerical value for α is $\alpha = 0.0534$. so that we can keep high schools for which H_0^{msy} is only rejected 0 or 1 time out of 7 at the 0.0534 % level.

α must be set the following way depending on the number of H_0^{msy} rejections out of 7 that we want to accept (with a general type I error set at 5 % in any case):

	Corresponding α for H_0^{msy}, depending on number of years Y available				
Decision rule	Y=7	Y=6	Y=5	Y=4	Y=3
Refuse any H_0^{msy} rejection	0.0073	0.0085	0.0102	0.0127	0.0170
Accept 1 H_0^{msy} rejection at most	0.0534	0.0628	0.0764	0.0976	0.1354
Accept 2 H_0^{msy} rejections at most	0.1288	0.1532	0.1893	0.2486	0.3684