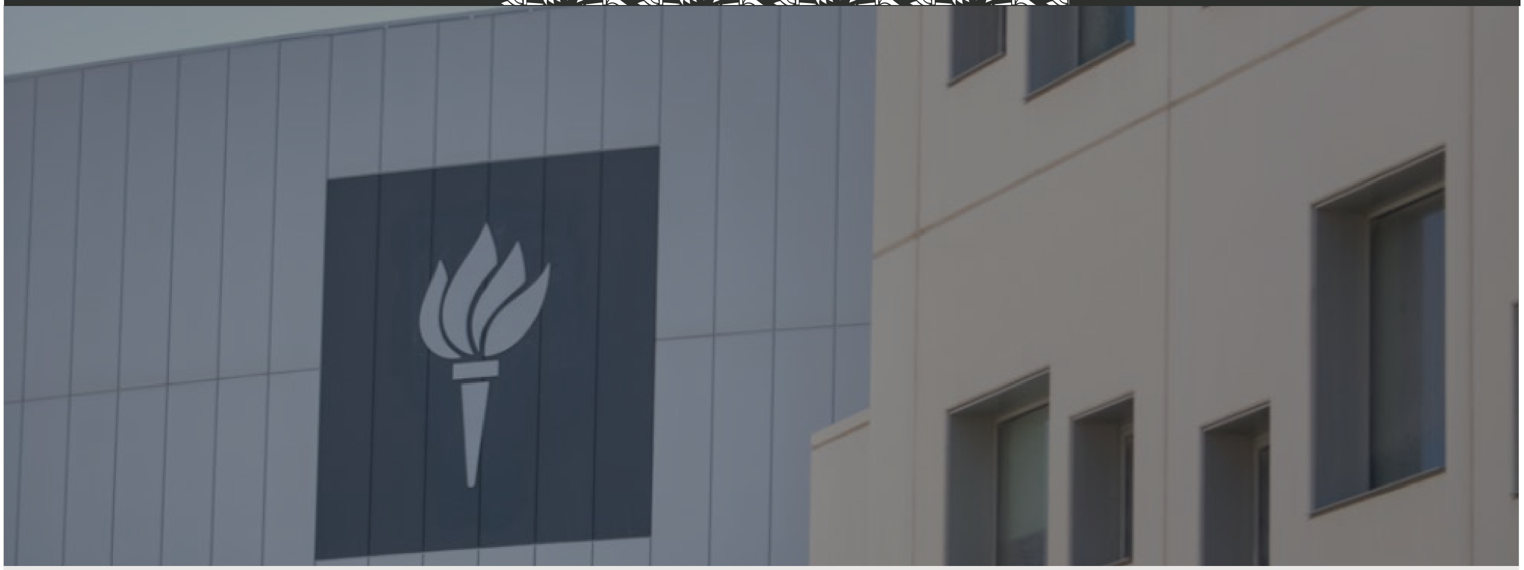

Endogenous Tracking: Sorting and Peer Effects

Aleksei Chernulich, Romain Gauriot, and Daehong Min

Working Paper # 0084
January 2023

Division of Social Science Working Paper Series
New York University Abu Dhabi, Saadiyat Island P.O Box 129188, Abu Dhabi, UAE
http://nyuad.nyu.edu/en/academics/academic-divisions/social-science.html

Endogenous Tracking: Sorting and Peer Effects*

This version: January 9, 2023
The latest version is available here

Aleksei Chernulich [a,†] Romain Gauriot [b,‡] Daehong Min [a,§]

[a] *Division of Social Science, New York University Abu Dhabi*
[b] *Department of Economics, Deakin University*

Abstract

We show that, when the educational choice is costly, the motive of seeking positive peer effects can result in ability grouping. In particular, high-achieving students self-sort by choosing costly courses, which we refer to as "endogenous tracking." We demonstrate the implications of endogenous tracking using the data from French middle schools, where ability grouping officially is not allowed. Instead, students are grouped together to study all courses in the standardized curriculum based on their choices between studying Spanish or a more effort-costly German. We find that costly language choices result in groups that significantly differ in terms of academic performance. Furthermore, we exploit regional differences in the effort costs of learning German to confirm that larger costs of choosing German result in more selective endogenous tracking. Finally, we identify peer effects that, together with sorting, generate inequality in educational outcomes. Such inequality, combined with observed inequality in socioeconomic status between the formed groups, works against egalitarian educational policies.

Keywords: Peer effects, sorting, educational tracking, socioeconomic inequality.
JEL Classification: H75, I21, I28, J24.

*We thank Olivier Bochet, Celeste Carruthers, Julien Combe, Hervé Crès, Matt Eliott, Raquel Fernandez, Ben Golub, Maia Guell, John Ham, Jean Imbs, Moritz Janas, Francis Kramarz, Jérémy Laurent-Lucchetti, Tanguy Le Fur, Max Mihm, Nikos Nikiforakis, Ernesto Reuben, Sevi Rodriguez Mora, Georg Schaur, John Singleton, Tayfun Sönmez, Olivier Tercieux, William Thomson, Utku Ünver, Etienne Wasmer, John Wooders, Myrna Wooders, Bumin Yenmez, Ulf Zölitz, and participants of the SOLE 2022, the PET 2022, the Australian Education Markets seminar, the AASLE 2022, and seminars at Boston College and NYUAD for useful comments. Aleksei Chernulich is grateful for financial support from Tamkeen under the NYUAD Research Institute award for Project CG005.

† *E-mail:* Aleksei.Chernulich@nyu.edu.
‡ *E-mail:* Romain.Gauriot@deakin.edu.au.
§ *E-mail:* Daehong.Min@nyu.edu.

1 Introduction

When students choose schools, educational tracks, and courses, they not only choose to accumulate human capital or signal about it, but they also choose peers. Peers are important for the academic progress of students, and peer effects in education are well documented (Hoxby, 2000b; Sacerdote, 2001; Ammermueller and Pischke, 2009; Booij et al., 2017; Bursztyn et al., 2019). To leverage these effects, schools can use tracking and put students in groups by ability, for example, based on grade point average (GPA). But many countries opt out of early tracking because it generates inequality and possibly harmful effects for lower-achieving students (Hanushek and Wößmann, 2006). We show that tracking nevertheless can emerge *endogenously*, without restrictions on GPA. Students anticipate the importance of peer effects, and high-achieving students take effort-costly courses to study with other high-achieving peers. The effects of endogenous tracking within schools have not been previously studied in the literature, unlike the effects of standard "*exogenous*" tracking (Figlio and Page, 2002; Zimmer, 2003; Lefgren, 2004; Duflo et al., 2011). We use data from French middle schools to demonstrate that endogenous tracking leads to inequality in educational outcomes due to peer effects and also results in grouping by socioeconomic status. Notably, identified endogenous tracking in French schools occurs earlier than official tracking in the final years of middle school.

Endogenous tracking – self-sorting of students by ability – can emerge when the choice of a course defines peers, and a more effort-costly course attracts higher-achieving students who seek positive peer effects. Educational choices usually involve considering costs and prospective peers, among other factors, but we provide the following example from the French secondary education system where other factors have limited presence. Most students in France enter middle school when they are 11 years old and receive a diploma after four years of secondary education. Upon entry, the vast majority of students choose English as their first foreign language to study and then choose between German and Spanish as their second foreign language class. German is a costly choice–while the returns to studying German are limited compared to Spanish, German is harder to learn for French-speaking students since French and Spanish belong to the Romance family of languages, while German belongs to the Germanic family.[1] Importantly, while there is no tracking at this point, the language class choice is pivotal in determining peers and their effects because class composition for *all* school courses is often based on a second foreign language choice to avoid schedule clashes. In this way, since students are usually taught by the same teachers following the statutory curriculum, peers and their effects on academic progress are the only driving force of endogenous tracking within schools. Our theoretical model predicts that, in equilibrium, academically better-performing students endogenously track themselves into German.

To empirically investigate the effects of endogenous tracking, we perform an observational study of representative panel data on French students collected by the French Ministry of Education. We use the information on more than 22,500 students who entered middle school in 2007 and graduated in 2012. The panel provides information on the academic progress of students with test results before and after the second foreign language class choice. The panel also reports students' socioeconomic status and zip codes, which are sufficient to recover information on geographical regions of schools that

[1] Only fragmented evidence on returns on learning German and Spanish in France exists. Using data from the European Community Household Panel Survey (1994-2001), Williams (2006) shows that speaking Spanish has a larger positive effect on income than speaking German.

we use to exploit the variation in the effort cost of studying German.

We find evidence of endogenous tracking in French middle schools. A simple comparison of Spanish-taking and German-taking students' academic performances in 2007 shows predicted by the model sorting effects – students who choose German over Spanish have significantly better test scores before they choose a language class by approximately 10%. As a signaling motive can only play a limited role in the identified differences due to observability of the academic progress that makes information on language choice irrelevant for predicting the abilities of students, we can attribute identified differences to a peer-seeking motive.[2] We use various background characteristics, including income and parents' level of education, to show inequality in socioeconomic status between the formed groups – students choosing German belong to richer and more educated households.

Our second finding is that larger effort cost differences in courses lead to more selective sorting, in line with theoretical predictions. We use regional variation in school locations and exploit the fact that the relative cost of choosing German over Spanish is higher near Spain than near Germany. This variation is due to the higher probability of having Spanish-speaking relatives in regions close to Spain and German-speaking relatives in regions close to Germany, and the possibility of commuting to work in corresponding countries in the future. We compare sorting effects in regions of France that share a border with Germany with sorting in regions of France that share a border with Spain. We find that significantly more able students sort themselves into German-learning classes in regions close to Spain when compared to German-learning students in regions close to Germany, confirming a comparative statics prediction derived from our model of endogenous tracking.

Finally, we identify positive peer effects, supporting the central assumption of the theoretical model. We demonstrate that students' progress is positively affected by those peers who endogenously track themselves by choosing a German language course. To identify possible peer effects, we match similar students choosing different classes using propensity score matching from Rosenbaum and Rubin (1983). Using students' observable characteristics before they choose their second foreign language course, we match students based on academic performance and various background characteristics. Propensity score matching shows a positive effect of German class composition on performance for many school courses that are included in the National Diploma assessment. As a robustness check, we also perform matching directly on observables using the genetic matching estimator of Diamond and Sekhon (2013) and confirm the result. We use the rich set of observable characteristics that are directly relevant to academic progress, and we, therefore, expect matching to eliminate differences in unobservable characteristics, allowing us to attribute the identified effect to peer group composition.

Our results are relevant to debates on developing effective policies to address inequalities in education. Inequality in education and skill acquisition is associated with inequality in wages (Oakes, 2005; Taber and Vejlin, 2020). While the focus is usually on inequalities generated by formal tracking, our paper highlights the issues generated by endogenous tracking within schools. Identifying endogenous tracking effects in French schools is particularly relevant for a string of recent reforms tackling inequalities in education, including the middle school reform of 2016 aimed at eliminating the language-based grouping rule. Notably, inequality in French middle schools is high, and French students

[2]Notably, the choice of German does not appear to be a signaling device for students who want to signal their propensity to learn languages since only 5% of German takers later choose a language-learning-intensive Literature track.

were the most affected by their socioeconomic status among all OECD countries, according to the Programme for International Student Assessment (PISA) study of 2015. Understanding hidden forces such as endogenous tracking are instrumental for the effective implementation of egalitarian educational policy.

The theoretical model of endogenous tracking introduced in our paper contributes to the literature on costly sorting. In sorting models, externalities are generated by a group, and there is some form of cost for joining groups that results in sorting. Papers on the costly sorting cover, inter alia, tournament participation (Morgan et al., 2018; Azmat and Möller, 2018), club participation (Windsteiger, 2021), and residential choices (Tiebout, 1956; Rothstein, 2006).[3] We provide a simple theoretical model that generates sorting using an increasing differences condition for cost and a standard linear-in-means peer effects model (Manski, 1993; Bramoullé et al., 2009). We also establish an observational equivalence between our model and the signaling model of Spence (1973) that is widely used to rationalize costly investments in education (Altonji, 1995; Weiss, 1995). In this way, we extend the domain cases where costly educational choices lead to sorting to situations of observable abilities by replacing the signaling motive with the peer-seeking motive. The intuition captured by our model may go beyond the educational framework and provide additional explanations for various observed stylized facts about self-sorting in situations where peer *composition* matters. For example, one may rationalize high fees to enter elite golf clubs using our model, where wealthy players have preferences to play with wealthy peers and separate themselves from other not-so-wealthy golf players.[4]

Our empirical analysis that identifies peer effects in self-sorted groups of students contributes to the large literature on peer effects in education, see Epple and Romano (2011) and Sacerdote (2011) for an overview. Peers can affect their classmates directly by generating student-to-student spillovers and indirectly by influencing teacher efforts and choice of instruction levels, and our estimates are likely to combine both types of effects. Students' academic progress may be affected by having both high-achieving peers (Lavy et al., 2012; Card and Giuliano, 2016) and disruptive peers (Lazear, 2001; Carrell et al., 2018) in their class. Evidence on the effects of the average ability composition of peers on academic performance is mixed. Booij et al. (2017) show positive peer effects for low-ability students, while Carrell et al. (2013) and Feld and Zölitz (2017) show negative effects; Duflo et al. (2011), Bui et al. (2014), Abdulkadiroğlu et al. (2014), and Tangvatcharapong (2020) do not find direct peer effects. We hypothesize that ability grouping, arising from endogenous tracking, results in peer effects because many social ties and friendships, which are shown to play an important role in academic context (Ly and Riegert, 2014; Behaghel et al., 2017; Beuchert et al., 2018), are preserved in this case. This hypothesis is consistent with results in Carrell et al. (2013) and recent evidence from a framed field experiment in Kiessling et al. (2021) that shows peer effects are larger in self-selected groups.

[3] An interesting example of experimental work on costly sorting is Aimone et al. (2013), where the authors use the voluntary contribution mechanism setup to show how agents can use sacrifice to endogenously sort into groups of more cooperative agents.

[4] One may use the proposed model to explain why talented specialists join startups instead of better-paying jobs at established companies, even without stock option motivation programs – they are willing to sacrifice a fraction of their paycheck to enjoy peer effects from those who are also willing to do the same. Data on the earnings of employees in the US (Ouimet and Zarutskie, 2014) and in Denmark (Burton et al., 2018) suggest startup employees earn 5% less than comparable employees at established firms.

2 Theoretical Predictions

In this section, we model endogenous tracking as a game of incomplete information. In the model, students make a language choice without observing the other students' abilities and choices. We assume that the peer effect is an average of the group members' abilities and that there is a relative cost of joining a German-learning group.[5] Based on the theoretical model, we derive three hypotheses that we later empirically validate using the data: (i) German learners and Spanish learners differ in terms of ability composition, (ii) larger cost leads to more selective sorting, and (iii) educational experience for German-learners and Spanish-learners differ, not due to ability differences, but due to peer effects.

2.1 Model

Consider n players who decide which language they want to learn. There is a commonly known distribution, $F(a)$, over the set of possible abilities, $A = [\underline{a}, \overline{a}]$, with a density $f(a)$ positive everywhere. Each player observes her own ability, a_i, drawn from $F(a)$. Then, each player i chooses one of two languages, G or S. Hence, each player's strategy is $s_i : A \to \{G, S\}$.[6] Players who choose the same language belong to the same group. Denote by g_l the group of players who choose language $l \in \{G, S\}$. Each member of a group bears the cost of learning the chosen language, but also enjoys a benefit from her peers' abilities, i.e., the peer effect.

The peer effect is an average of the abilities of other players who belong to the same group. We denote by $P_i^e(g_l)$ the peer effect that player i in group g_l experiences. The learning cost of each language l depends on an intrinsic difficulty of learning, λ_l, and a player's ability, a_i. We denote by $c_l(\lambda_l, a_i)$ the learning cost of a player whose ability is a_i. In particular, we are interested in the difference in two learning costs, $\Delta c(\lambda, a_i) := c_G(\lambda_G, a_i) - c_S(\lambda_S, a_i)$, where $\lambda = h(\lambda_G - \lambda_S)$ measures the difference in the intrinsic difficulties and strictly increases in $\lambda_G - \lambda_S$. Assumption 1 articulates the learning costs and peer effects.

Assumption 1. (Learning Costs and Peer Effects)

(a) *$c_l(\lambda_l, a_i)$ is continuously differentiable in λ_l and a_i,*

(b) *$\Delta c(\lambda, a_i) > 0$, $\partial \Delta c(\lambda, a_i)/\partial \lambda > 0$, and $\partial \Delta c(\lambda, a_i)/\partial a_i < 0$ for all $a_i \in A$,*

(c) *The peer effect for player i in group g_l is*

$$P_i^e(g_l) \;=\; \begin{cases} \frac{1}{|g_l \backslash \{i\}|} \left(\sum_{j \in g_l \backslash \{i\}} a_j \right) & \text{if } \; |g_l \backslash \{i\}| > 0, \\ 0 & \text{if } \; |g_l \backslash \{i\}| = 0, \end{cases}$$

where $|g_l \backslash \{i\}|$ is the cardinality of $g_l \backslash \{i\}$.

[5] There are two ways to define this average for a given student: (i) the average of abilities of all group members, including herself, and (ii) the average of abilities of all group members except herself. In the main model, we use the second definition, which is widely accepted in the literature. As a complementary analysis, in Appendix A, we provide a complete information model with peer effects modeled as the average of abilities of all group members, including oneself. In that model, we provide an alternative mechanism of sorting that, unlike the model from the main text, is not based on the ability-dependent costs of studying languages. Instead, separation happens because low-ability students may decrease average peer effects by joining the group to the extent that makes them worse off.

[6] We do not consider randomizations over $\{G, S\}$.

Assumption 1(a) is for convenience. Assumption 1(b) concerns the properties of $\Delta c(\cdot)$. First, it is strictly positive, that is, given a realized ability, a_i, every player finds it more costly to choose G than S. Second, it is increasing in λ, that is, an increase in the difference between the intrinsic difficulties makes the cost difference higher for any $a_i \in A$. Third, it is strictly decreasing in a_i, that is, a player with a higher ability experiences less increase in the learning cost when changing her choice from S to G. These decreasing differences allow us to derive an equilibrium *à la* Spence (1973): high types isolate themselves from low types by bearing an additional cost of learning German that is only bearable for high types.[7] Last, Assumption 1(c) says that player i's peer effect in group g_l is the average of the realized abilities of other players in that group, excluding her own ability. We assume a zero peer effect for a player if she is the only player in a group.

Each player's payoff is her net educational benefit – a positive peer effect minus the cost of the chosen language. Thus, each player's payoff increases in the peer effect she experiences in g_l and decreases in the learning cost of the chosen language.[8] We assume that the payoff function is linear in these two factors. Given a realized ability $a_i \in A$ and a strategy profile, $\boldsymbol{s} = (s_1, \cdots, s_n)$, if a player chooses G, then her payoff is

$$u_i(G|a_i, \boldsymbol{s}_{-i}) = E[P_i^e(g_G)|a_i, \boldsymbol{s}_{-i}] - c_G(\lambda_G, a_i),$$

while, if a player chooses S, then her payoff is

$$u_i(S|a_i, \boldsymbol{s}_{-i}) = E[P_i^e(g_S)|a_i, \boldsymbol{s}_{-i}] - c_S(\lambda_S, a_i).$$

We focus on a symmetric equilibrium in which every player employs the same "threshold strategy." Consider a strategy profile, $\hat{\boldsymbol{s}} = (\hat{s}, \cdots, \hat{s})$, where

$$\hat{s}(a_i) \quad = \quad \begin{cases} G & \text{if } a_i \geq \hat{a}, \\ S & \text{otherwise.} \end{cases}$$

Given that every player plays $\hat{s}$ and a realization of player i's ability a_i, consider an event in which there are k players whose ability realizations are greater than $\hat{a}$, and $n-k$ number of players whose ability realization is less than $\hat{a}$. Denote this event by $\mathcal{E}_k$, and its probability is

$$P[\mathcal{E}_k] = \binom{n-1}{k} F(\hat{a})^{n-k}(1-F(\hat{a}))^k.$$

Note that random variable $\mathcal{E}_k$ follows the binomial distribution with a "success" probability of $(1-F(\hat{a}))$. Given that event $\mathcal{E}_k$ happened, the expected peer effect for player i for choosing G is

$$\begin{aligned} E[P_i^e(g_G)|\hat{\boldsymbol{s}}_{-i}, a_i; \mathcal{E}_k] \quad &= \quad \frac{1}{k} E\left[\sum_{j \in g_G \setminus \{i\}} a_j \;\middle|\; a_j \geq \hat{a} \right] \\ &= \quad E[a|a \geq \hat{a}] = \frac{\int_{\hat{a}}^{\bar{a}} af(a)da}{1-F(\hat{a})}. \end{aligned}$$

[7] We elaborate the observational equivalence between the separating equilibrium outcomes in Spence (1973) and Sorting outcomes in our model under certain conditions in Appendix M.

[8] Here, we are abstract from the preferences for within-group ranking that may affect decisions to join more competitive or less competitive groups. See Villeval (2020) for a recent overview of the existing evidence on the effects of providing information on relative ranking.

Then, the expected peer effect is simply

$$E[P_i^e(g_G)|\hat{\boldsymbol{s}}_{-i}, a_i] = E_{\mathcal{E}_k}[E[P_i^e(g_G)|\hat{\boldsymbol{s}}_{-i}, a_i, \mathcal{E}_k]] = \sum_{k=0}^{n-1} P[\mathcal{E}_k]E[a|a \geq \hat{a}] = E[a|a \geq \hat{a}].$$

In the same manner, the expected peer effect for player i when joining g_S is

$$E[P_i^e(g_S)|\hat{\boldsymbol{s}}_{-i}, a_i] = E[a|a < \hat{a}].$$

Denote by $\Delta P_i^e(\hat{a})$ the difference between $E[P_i^e(g_G)|a_i, \hat{\boldsymbol{s}}_{-i}]$ and $E[P_i^e(g_S)|a_i, \hat{\boldsymbol{s}}_{-i}]$. That is,

$$\begin{aligned}
\Delta P_i^e(\hat{a}) &= E[a|a \geq \hat{a}] - E[a|a < \hat{a}] \\
&= \int_{\hat{a}}^{\bar{a}} a\frac{f(a)}{1-F(\hat{a})}da - \int_{\underline{a}}^{\hat{a}} a\frac{f(a)}{F(\hat{a})}da \\
&= \frac{1}{1-F(\hat{a})}\int_{\hat{a}}^{\bar{a}} af(a)da - \frac{1}{F(\hat{a})}\left\{\int_{\underline{a}}^{\bar{a}} af(a)da - \int_{\underline{a}}^{\hat{a}} af(a)da\right\} \\
&= \frac{\int_{\hat{a}}^{\bar{a}} af(a)da - (1-F(\hat{a}))a_m}{F(\hat{a})(1-F(\hat{a}))},
\end{aligned}$$

where $a_m = \int_{\underline{a}}^{\bar{a}} af(a)da$. An immediate observation is that $\Delta P_i^e(\hat{a}) > 0$ for any $\hat{a} \in (\underline{a}, \bar{a})$. Furthermore, given that every player employs $\hat{s}(a_j)$, $\Delta P_i^e(\hat{a})$ is identical across all players. More importantly, it is independent of a_i, a realized ability of player i. Lemma 1 discusses the properties of $\Delta P_i^e(\hat{a})$.

Lemma 1. $\Delta P_i^e(\hat{a})$ *is continuously differentiable in* $\hat{a} \in (\underline{a}, \bar{a})$. *In addition,*

$$\lim_{\hat{a}\to\underline{a}} \Delta P_i^e(\hat{a}) = a_m - \underline{a} \quad \textit{and} \quad \lim_{\hat{a}\to\bar{a}} \Delta P_i^e(\hat{a}) = \bar{a} - a_m,$$

where $a_m = \int_{\underline{a}}^{\bar{a}} af(a)da$.

The proof of Lemma 1 is relegated to the Appendix B. Note that $\Delta P_i^e(\hat{a})$ is continuous as far as $f(a)$ is continuous. Furthermore, while $\Delta P_i^e(\hat{a})$ is not defined at $\hat{a} = \underline{a}$ or $\hat{a} = \bar{a}$, it approaches two different limits as either $\hat{a} \to \underline{a}$ or $\hat{a} \to \bar{a}$.

Proposition 1. *If (i)* $\Delta c(\lambda, \underline{a}) > a_m - \underline{a}$ *and (ii)* $\Delta c(\lambda, \bar{a}) < \bar{a} - a_m$, *there exists a sorting equilibrium in which every player chooses* G *if* $a_i \geq a^*$ *and chooses* S *otherwise, where* $a^* \in (\underline{a}, \bar{a})$.

Proof. Suppose every player follows a threshold strategy, $\hat{s}(a_i|\hat{a})$, in which she chooses G if $a_i \geq \hat{a}$ and S otherwise, where $\hat{a} \in (\underline{a}, \bar{a})$. Consider type $\hat{a}$ of a player. By choosing G, this type would get

$$u_i(G|\hat{a}, \boldsymbol{s}_{-i}) = E[a|a \geq \hat{a}] - c_G(\lambda_G, \hat{a}),$$

and, by choosing S, this type would get

$$u_i(S|\hat{a}, \boldsymbol{s}_{-i}) = E[a|a < \hat{a}] - c_S(\lambda_S, \hat{a}).$$

Then, the difference between these two payoffs, denoted by $\Delta u_i(\hat{a})$, is

$$\Delta u_i(\hat{a};\lambda) = \Delta P_i^e(\hat{a}) - \Delta c(\lambda, \hat{a}).$$

Note that $\Delta u_i(\hat{a};\lambda)$ is continuous in $\hat{a} \in (\underline{a}, \bar{a})$, and, by Lemma 1,

$$\lim_{\hat{a}\to\underline{a}} \Delta u_i(\hat{a};\lambda) = a_m - \underline{a} - \Delta c(\lambda, \underline{a}) \quad \text{and} \quad \lim_{\hat{a}\to\bar{a}} \Delta u_i(\hat{a};\lambda) = \bar{a} - a_m - \Delta c(\lambda, \bar{a}).$$

By the assumptions, $\Delta c(\lambda, \underline{a}) > a_m - \underline{a}$ and $\Delta c(\lambda, \bar{a}) < \bar{a} - a_m$. Thus, we have $\lim_{\hat{a}\to\underline{a}} \Delta u_i(\hat{a};\lambda) < 0$ and $\lim_{\hat{a}\to\bar{a}} \Delta u_i(\hat{a};\lambda) > 0$. As $\Delta u_i(\hat{a};\lambda)$ is continuous in $(\underline{a}, \bar{a})$, there must exist $a^* \in (\underline{a}, \bar{a})$ such that $\Delta u_i(\hat{a} = a^*;\lambda) = 0$ by the intermediate value theorem. Let every player follow $\hat{s}(a_i|\hat{a} = a^*)$. Then, every type of each player has the identical $\Delta P_i^e(a^*)$ since $\Delta P_i^e(\cdot)$ is independent of a_i. Thus, the payoff difference of each type of player i is

$$\Delta u_i(a_i;\lambda) = \Delta P_i^e(a^*) - \Delta c(\lambda, a_i).$$

For $a_i = a^*$, $\Delta u_i(a_i;\lambda) = \Delta P_i^e(a^*) - \Delta c(\lambda, a^*) = 0$ by the definition of a^*. For any $a_i < a^*$, $\Delta u_i(a_i;\lambda) = \Delta P_i^e(a^*) - \Delta c(\lambda, a_i) < 0$, and, for any $a_i > a^*$, $\Delta u_i(a_i;\lambda) = \Delta P_i^e(a^*) - \Delta c(\lambda, a_i) > 0$ since $\Delta c(\lambda, a_i)$ is strictly decreasing in a_i. Thus, $\hat{s}(a_i|a^*)$ is a best response to $\hat{\boldsymbol{s}}_{-i} = (\hat{s}(a_1|a^*), \cdots, \hat{s}(a_{i-1}|a^*), \hat{s}(a_{i+1}|a^*), \cdots, \hat{s}(a_n|a^*))$. Thus, $\boldsymbol{s}^* = (\hat{s}(a_1|a^*), \cdots, \hat{s}(a_n|a^*))$ is an equilibrium. □

Proposition 1 tells us the sufficient conditions for a (non-trivial) sorting equilibrium to exist. The sufficient conditions are intuitive; they simply impose the lower bound for the highest cost difference, $\Delta c(\lambda, \underline{a})$, and the upper bound for the lowest cost difference, $\Delta c(\lambda, \bar{a})$. Note that if $\Delta c(\lambda, \underline{a})$ is too low, no one would stay in g_S with a lower peer effect; likewise, if $\Delta c(\lambda, \bar{a})$ is too high, no one would stay in g_G and afford the learning cost.

Note that Proposition 1 is silent about the uniqueness of the sorting equilibrium. The proof shows that, by continuity of $\Delta P_i^e(\hat{a})$ and $c(\lambda, \hat{a})$ and the assumptions on the boundary values of $\Delta c(\lambda, \hat{a})$, there must be a^* which makes the difference between two functions zero. Then we can construct a sorting equilibrium by making all players employ the threshold strategy with a^*, i.e., $\hat{s}(a_i|a^*)$. Hence, if there exists more than one solution for $\Delta P_i^e(\hat{a}) - \Delta c(\lambda, \hat{a}) = 0$, we can construct multiple sorting equilibria. The multiplicity of the sorting equilibria obscures the comparative statics.[9] However, with the assumptions in Proposition 1, we can state the following comparative statics results. Denote by A^* the set of equilibrium threshold values. Let $\underline{a}^* = \min A^*$ be the smallest equilibrium threshold and $\bar{a}^* = \max A^*$.

Proposition 2. *Suppose that (i) $\Delta c(\lambda, \underline{a}) > a_m - \underline{a}$ and (ii) $\Delta c(\lambda, \bar{a}) < \bar{a} - a_m$. Given $\lambda > 0$, if $\partial \Delta P_i^e(a)/\partial a \neq \partial \Delta c(\lambda, a)/\partial a$ at $a = \underline{a}^*$ and $a = \bar{a}^*$, then $\underline{a}^*(\lambda)$ and $\bar{a}^*(\lambda)$ are increasing in λ.*

Proof. Both $\underline{a}^*$ and $\bar{a}^*$ are solutions for $\Delta u_i(a;\lambda) = \Delta P_i^e(a) - \Delta c(\lambda, a) = 0$. Consider $\underline{a}^*$ first. By the implicit function theorem, there exists $\underline{a}^*(\lambda)$ and its derivative is

$$\frac{d\underline{a}^*(\lambda)}{d\lambda} = -\frac{\partial \Delta u_i(\underline{a}^*;\lambda)/\partial\lambda}{\partial \Delta u_i(\underline{a}^*;\lambda)/\partial a}.$$

[9] For example, suppose that we have two sorting equilibria. Suppose also that we have a new equilibrium with a different threshold which is between two sorting equilibria we start with. Then, we cannot distinguish whether the threshold has been increased or decreased.

Now note that

$$\begin{aligned}\frac{\Delta u_i(\underline{a}^*;\lambda)}{\partial\lambda} &= -\frac{\partial\Delta c(\underline{a}^*,\lambda)}{\partial\lambda} < 0, \quad \text{and}\\ \frac{\Delta u_i(\underline{a}^*;\lambda)}{\partial a} &= \frac{\partial\Delta P_i^e(\underline{a}^*)}{\partial a} - \frac{\partial\Delta c(\underline{a}^*,\lambda)}{\partial a} \neq 0.\end{aligned}$$

Thus, $d\underline{a}^*(\lambda)/d\lambda$ is well-defined and its sign entirely depends on the sign of $\partial\Delta u_i(\underline{a}^*;\lambda)/\partial a$. Now suppose that $d\underline{a}^*(\lambda)/da < 0$. Then, we have

$$\frac{\Delta u_i(\underline{a}^*;\lambda)}{\partial a} = \frac{\partial\Delta P_i^e(\underline{a}^*)}{\partial a} - \frac{\partial\Delta c(\underline{a}^*,\lambda)}{\partial a} < 0, \quad \text{which is equivalent to}$$

$$\frac{\partial\Delta P_i^e(\underline{a}^*)}{\partial a} < \frac{\partial\Delta c(\underline{a}^*,\lambda)}{\partial a} < 0.$$

As $\Delta P_i^e(a)$ is decreasing faster than $\Delta c(a,\lambda)$ at $a = \underline{a}^*$, there exists an $\epsilon > 0$ such that

$$\Delta P_i^e(\underline{a}^* - \epsilon) > \Delta c(\underline{a}^* - \epsilon, \lambda), \quad \text{which implies} \quad \Delta u_i(\underline{a}^* - \epsilon;\lambda) > 0.$$

By the assumption that $\Delta c(\underline{a},\lambda) > a_m - \underline{a} = \lim_{a\to\underline{a}} \Delta P_i^e(a)$, we must have $\Delta u_i(\underline{a};\lambda) < 0$. Since $u_i(a,\lambda)$ is continuous in a, we must have $a' \in (\underline{a}, \underline{a}^* - \epsilon)$ such that $\Delta u_i(a = a';\lambda) = 0$ by the intermediate function theorem. This is a contradiction to the definition of $\underline{a}^*$ since there exists $a' < \underline{a}^*$ which solves $\Delta u_i(a;\lambda) = 0$. Hence, $d\underline{a}^*(\lambda)/d\lambda \geq 0$. By a similar argument, one can show that $d\bar{a}^*(\lambda)/d\lambda \geq 0$. □

Proposition 2 is reminiscent of Topkis's Monotonicity Theorem. The comparative statics results are strong because they are valid under a large set of ability distributions, $F(a)$. However, they are weak since they are silent about the intermediate equilibrium thresholds. Furthermore, with the multiple values of a^* given λ, it is possible to construct a decreasing $a^*(\lambda)$. For example, given $\lambda' < \lambda''$, one can select $a^*(\lambda') \in A^*(\lambda')$ which is greater than $a^*(\lambda'') \in A^*(\lambda'')$.

Example. Consider a cohort of students with abilities distributed on the interval $[0, 20]$ following truncated normal with a mean of 10 and a standard deviation of 4.5. Further assume that $\Delta c(a_i) = C - 0.5a_i$. First, consider a situation where $C = 15$. Equilibrium cutoff ability a^* is equal to 14.71; the equilibrium is illustrated in the left panel of Figure 1. Note that the socially optimal cutoff is 20, where all students choose Spanish language. Second, consider a variation of C between 10 and 20 – following Proposition 1, sorting is an equilibrium outcome in this case. The right panel of Figure 1 demonstrates how equilibrium ability cutoffs a^* vary with the cost. One can see that, with the increase in the cost, average abilities in *both* groups increase.

2.2 Conjectures

We use the theoretical model of the language class choice to formulate three conjectures concerning sorting effects, cost increase effects, and peer effects.

Conjecture 1. *On average, students who choose German have higher abilities than students who choose Spanish.*

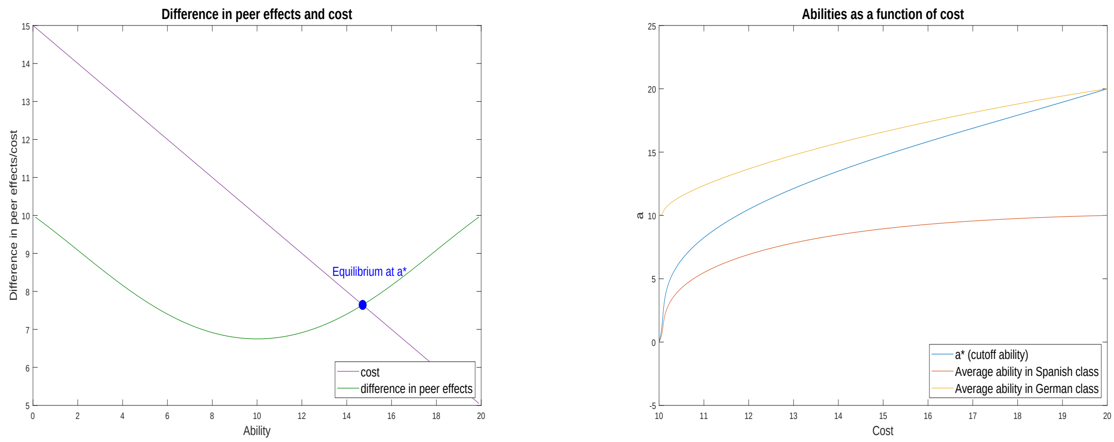

Figure 1: Left panel: an illustration of the equilibrium following the example. Right panel: an illustration of varying costs following the example.

Conjecture 1 naturally follows from Proposition 1 if one assumes the relative cost of languages supports sorting in equilibrium for a given ability curve. For empirical testing, we use a somewhat weaker implication of the model and focus on using average values of abilities in the groups to account for the possible non-equilibrium behavior of agents near the cutoff value on the ability curve.

Conjecture 2. *An increase in the cost of taking German leads to an increase in the selectivity of endogenous tracking.*

For the last conjecture, we further assume the linearity of the cost function and refer to changes in cost as changes in the value of the intercept – the component of the cost that is independent of ability. Following the example provided in the text, it is easy to see that a larger cost leads to a larger cutoff ability and that it increases the average abilities in both groups.

Conjecture 3. *On average, students who choose German experience larger positive peer effects than students who choose Spanish.*

Conjecture 3 is based on Conjecture 1 and the definition of peer effects. Peer effects play a prominent role in sorting effects, and identifying those effects with the data is an important part of the process of validating the model's assumptions.

3 Background, Data, and Empirical Strategy

In this section, we provide a brief overview of the relevant features of the French educational system. We then describe our main source of data – the panel collected by the French Ministry of Education. Finally, we explain our empirical strategy to test three theoretical conjectures introduced in Section 2.

3.1 French Middle Schools

French secondary education has two stages – middle schools for children aged 11 to 15 and high school for children aged 15 to 18. Tracking takes place at the high school level, where students can choose from science, economics and social sciences, and literature tracks, but ability tracking is officially forbidden in middle schools. The French educational system is very centralized; to harmonize the educational process, it has legal requirements to teach a curriculum designed by the French Ministry of Education and not to track students by ability, leaving no room for individual schools to apply their own rules.

Nevertheless, as we argue in the paper, tracking emerges endogenously in French middle schools, and high-achieving students choose German instead of Spanish as a second foreign language because class compositions for all school courses are usually based on this choice. The fact that optional courses, such as a second language or Latin language, appear to be a mechanism for within-school sorting of students has been mentioned in a few reports and papers on the French education system (Duru-Bellat and Mingat, 1997; Ly et al., 2014). In particular, Herbaut et al. (2019) noted the following tendency: "Most notably, the choice of German as a first foreign language, Latin (an option from grade 7th), of a 'bilingual stream' from grade 6th, are typically chosen by upper-class and good-performing students." Note that the optional choice of Latin, unlike the second foreign language choice, does not affect the class composition, but rather allows students to get extra Diploma points. The choice of German as a first foreign language has been historically popular but currently accounts for approximately 5% of choices.

The implications of a second foreign language for students are understood by students themselves and their parents, who, at this stage, play an important role in children's decisions. There are ample discussions among anxious parents on online forums seeking advice on which second language their child should choose. For example, in March 2013, on an online forum for teachers, one parent wrote: "My youngest son is due to start middle school next year. I am told that the only option that guarantees him a class of a good level is German!!! A 'bilingual' German-English class." One of the replies stated that "[...] it allows sorting students without saying it."[10]

3.2 Data

	Before	Language Choice	After
Year	2007-08	2009-10	2011-12
Data			
Family Survey	✓		
National Standardized Test	✓		
Specific Standardized Test	✓		✓
National Diploma			✓

Table 1: The overview of the data used in the study. The data available *before* the language choice are used to identify sorting effects and to perform matching. The data available *after* the language choice is used to identify peer effects.

The French Ministry of Education collects information on students' educational experiences approximately every decade. We use the latest available and the most detailed panel data of $35,000$ French students, following students through their schooling DEPP

[10]The discussion is available here: https://www.neoprofs.org/t56937p25-pourquoi-l-allemand-est-il-encore-critere-de-selection.

(2018).[11] These rich panel data have been collected by the Ministry of Education to investigate students' learning pathways. The sample is large and representative of the French student population. Out of the 7,004 junior high schools operating in France in 2007, 98% (i.e., 6,857) had at least one student represented in the panel.[12] Each junior high school included in the data has, on average, five students participating in the study.

The panel starts in 2007 when students enter middle school (usually at 11 years old) and stops in 2014 when they finish high school and go to university or start working. Table 1 shows the data from the panel we use in the empirical analysis. In particular, we use the results of the Family Survey performed in 2007, when families answered an extensive survey about their living situation. At the beginning of the 2007-2008 school year, students took standardized tests in French and Mathematics. Those tests were administered to all students entering middle school, and we refer to them as *National Standardized Tests*. At the end of the 2007-2008 and of the 2011-2012 school year, students participating in this study took other standardized tests in French and Mathematics. Those tests were administered only to students participating in this study, and we refer to them as *Specific Standardized Tests*. At the end of 2012, students received the National Diploma (*"Brevet des collèges"*), which collects marks for all courses, including graduation exams taken by the end of middle school. The grades for different tests are given on different scales and, to harmonize the grades and allow for direct comparison, we rescale some grades to a scale of 0 to 20, which is the typical grade scale in France. We use the standardized tests and the exam results to measure students' abilities before and after studying in the groups formed based on their second language choices.

At the beginning of middle school, students choose their first foreign language class: 92% choose English, 6% choose German, and 1% choose Spanish. We exclude those students who choose German or Spanish as their first foreign language class. Students who follow the usual path make a choice of a second foreign language in 2010-2011.We also exclude observations that are not suitable for the study because marks for the National Diploma are missing. Reasons for missing observations include: students have gone to pre-vocational junior high school, the exam data are missing, students retook a year, etc. Finally, we exclude those students who chose neither German nor Spanish as their second foreign language class. In this way, we exclude 3,970 students choosing Italian, Portuguese, Arabic, etc. We are left with 26,508 students, which represents 75.8% of the original sample. Out of these 22,538 students, 3,696 (16%) studied German and 22,812 (84%) studied Spanish. Descriptive statistics for all variables used are presented in Appendix C.

3.3 Method

Conjecture 1 concerns the sorting effects, stating that better-performing students are expected to prefer German over Spanish. To test such a conjecture in Section 4.1, we use exam scores as a proxy for academic abilities. In this case, the analysis is straightforward – we compare average values for German learners and Spanish learners for various exam scores. Here, we do not investigate the determinants of language choice, but check if

[11] More information about this data set can be found here: http://www.progedo-adisp.fr/enquetes/XML/lil.php?lil=lil-0955

[12] Students of middle schools that are part of the *Ambition Réussite* program (the program for disadvantaged schools) have been oversampled, as this program was of particular interest to the researchers who collected these data. For schools in the *Ambition Réussite* program, the sample rate is 1/8 compared to 1/23 for the other schools.

sorting patterns align with theoretical predictions. We introduce a binary variable that reflects if the student chooses German, and use the Ordinary Least Squares (OLS) method to estimate the corresponding coefficient and assess its statistical significance. In a similar way, we investigate the effects of sorting on socioeconomic inequality between groups.

We estimate the following model of academic performance that helps highlight the differences between those who chose German and those who chose Spanish:

$$O_i = \alpha + \beta \times \mathbb{1}_{i\ chose\ German} + \epsilon_i, \tag{1}$$

where O_i is an outcome variable, $\mathbb{1}_{i\ chose\ German}$ is a dummy variable equal to 1 if the student is studying German and 0–otherwise, and ϵ_i is iid noise.

We perform the following exercise to investigate if the identified differences can be attributed to the peer-seeking motive rather than to the signaling motive. If a student tries to signal his ability by choosing a German course, then such a course choice should have a predictive power of abilities given observables related to abilities. To evaluate this predictive power, we predict middle school exit exam scores using entry exam scores and a language course choice. Results, available in Appendix L, demonstrate that German course choice does not have statistically significant predictive power, suggesting the limited presence of the signaling motive.

To address Conjecture 2, we use a similar approach and compare sorting outcomes separately for two regions – one region is close to Spain, and another is close to Germany. We run the following regression:

$$O_i = \alpha + \beta \times \mathbb{1}_{i\ chose\ German} + \gamma \times \mathbb{1}_{i\ close\ to\ Germany} + \delta \times \mathbb{1}_{i\ chose\ German} \times \mathbb{1}_{i\ close\ to\ Germany} + \epsilon_i, \tag{2}$$

where O_i is an outcome variable, $\mathbb{1}_{i\ chose\ German}$ is a dummy variable equal to 1 if student i is studying German and 0–otherwise, $\mathbb{1}_{i\ close\ to\ Germany}$ is a dummy variable equal to 1 if the student lives in "Close to Germany" region and 0–if student i lives in "Close to Spain" region, and ϵ_i is iid noise.

Conjecture 3 relates to peer effects, and it states that peer effects are larger for those students who choose German. One possibility to illuminate peer effects would be to show the difference in academic progress between those who choose German and those who choose Spanish using Difference-in-Difference (DID) estimation. However, based on the significant differences between sorted students established in Section 4.1, one can suspect that the parallel trend assumption, which is key for DID, does not hold. Moreover, not only do we not have data on performances sufficient to validate the assumption on a parallel trend, but we also cannot construct some synthetic analog as proposed, for example, as in Arkhangelsky et al. (2021).

Instead, we use matching to disentangle peer effects from sorting effects. We match students by the propensity score and complement the analysis with a robustness check using genetic matching directly on observables.[13] Using students' observable characteristics before choosing which second foreign language to study, we match students who chose German to comparable students who chose Spanish. We achieve the balance using the nearest matching estimator with a replacement that gives us pairs of students, where

[13]Genetic matching is based on an evolutionary search of weights for observables that are used for matching. It has propensity score matching and Mahalanobis distance-based matching as limit cases. For examples of utilizing genetic matching estimator, see Hopkins (2010) and Frymer and Grumbach (2021).

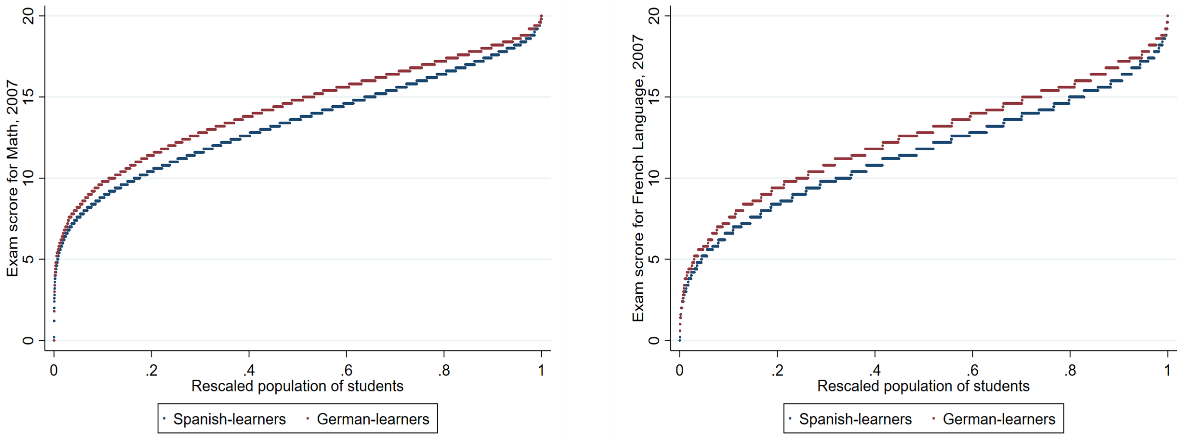

Figure 2: The graphs show the distribution of exam scores for the 2007 National Test. Populations of students are rescaled to be between zero and one; all scores are shown in ascending order; each dot represents one student, with Spanish learners in blue and German learners in red. The left panel depicts the distribution of exam scores in Mathematics. The right panel depicts the distribution of exam scores in a French language course.

one student represents the control group (Spanish learners) and another student represents the treatment group (German learners). We perform matching with replacements, leaving a possibility for one Spanish-learning student to be matched with several German-learning students. Results obtained through the matching with replacements have the desirable property of being invariant to the order of matching. We use the consistent estimator from Abadie and Imbens (2006), as using the bootstrap for calculation of the standard errors is invalid, as shown in Abadie and Imbens (2008).

For the matching method to work, we need two assumptions to hold: common support and conditional independence (for a detailed discussion of the issue, see, for example, Imbens (2015)). The validity of the common support assumption is ensured as we impose the common support restriction during matching and generate a sufficient number of matched pairs. In particular, we estimate the effect of class composition using only those German-taking students with propensity scores within the bounds of the propensity scores estimated for Spanish-taking students. The conditional independence assumption (also known as ignorability or unconfoundedness) ensures that treatment and control groups receive treatment assignments randomly once controlled for all observables. While this assumption is not refutable, we argue that because we collect many relevant observables, the likelihood of missing the variable that may be responsible for the estimated treatment effect is small. In particular, we match students based on their academic performance, socioeconomic status, and parents' involvement. We suggest the following possible sources of treatment randomization for students who are matched based on all observables. First, in the spirit of our model, students face different and random ability curves of their school cohort that define their relative position and, as a result, their optimal language class choice. Second, students are likely to have idiosyncratic preferences for languages to be learned or for competition to be faced, as in Falk and Knell (2004), that define whether they prefer to be in a more competitive environment of German-learners or in a less pressing environment of Spanish-learners.

	National Test		Special Test				
	Mathematics	French Language	Mathematics	French Language			
				Treatment of Incomplete Sentences	Understanding	Lexicon	Reasoning
German class	0.908***	0.933***	1.036***	1.201***	0.870***	0.822***	0.912***
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
Constant	13.365***	11.549***	10.850***	9.031***	13.102***	11.272***	10.546***
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
N	18,055	18,184	20,548	20,548	20,548	20,548	20,548

Probability p in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Difference in results between students who choose German and Spanish for courses in 2007 National test and 2008 Special test. Variable *German* is a binary variable that is equal to one if a student takes a German class and zero–if Spanish.

4 Results

In this section, we discuss the results of the empirical analysis and report on the evidence of three types of effects: sorting effects, cost difference effects, and peer effects. We show that sorting effects are present in the data – students who choose German have better grades than those who choose Spanish. We also show that sorting is associated with socioeconomic inequality – students choosing different languages differ across many background characteristics. We then show that the strength of sorting varies as the cost of choosing German varies.[14] Using matching, we identify the peer effects that lead to significantly better academic performance of those who choose German.

4.1 Sorting

To demonstrate sorting effects, we show that students choosing German have better academic performance. As measures of academic performance, we use the test scores for various courses. When entering middle school, students take National Standardized Tests in French and Mathematics. Table 2 shows that students who chose to study German scored 0.9 points higher (out of 20) in both French and Mathematics than the ones studying Spanish. We further illustrate the sorting effects with the graphs of the distributions of the scores in Figure 5. One can clearly see that the grades of German-learning students first-order stochastically dominate the grades of Spanish-learning students.

As part of the data collection for the panel, students passed Specific Standardized Test at the end of the 2008 and 2011 school years. This test measured ability in Mathematics (calculations and numbers, quantities and measurements, organization of logical data, geometry) and French language (lexicon and comprehension). We use OLS to estimate Model (1), where the outcome variable is test scores for pre-treatment tests in 2007 and 2008. Table 2 reports the results – those students who chose German did better in all sections of 2007 National and 2008 Special Tests. All identified differences across both tests are statistically significant at less than 0.1%-level, positive, and have a similar amplitude of approximately 10%.

[14] In Appendix I, we report on the structural estimation of language choices allowing for heterogeneity in the cost of choosing German over Spanish across regions of France. In line with the intuition on the nature of these costs, costs are smaller for regions closer to Germany.

We further explore the differences between students choosing different second foreign languages and show that students sorting into German have a higher socioeconomic status. We use data from the family survey completed in 2007 (i.e., before the second language was chosen) and find various proxies for socioeconomic status. We employ the following variables reported: family monthly income, number of rooms in the house, whether the child has his own room, whether the parents studied after high school, whether the child was born in France, and whether the parents were born in France.

We use Model (1) and employ various socioeconomic indicators as O_i. Table 3 reports the results of the OLS estimation. Students from more wealthy and educated families choose German over Spanish – corresponding variables are positive and highly statistically significant for related outcome variables. Parents of students studying German earn 236 euros more than parents of students studying German per month, their houses are larger by a quarter of a room, and their children are 2.3% more likely to have their own room.

Interestingly, while we do not find any effect of France being the birthplace of students, we find positive effects of France being the birthplace of their parents. We explain this finding with the inter-generational information spillover effects. Parents who have already gone through the French educational system are more likely to be aware of the strategic implications of language choice and, therefore, may strongly advise their children to take a German class in the hope that their children will experience positive peer effects from their classmates. Some may argue that involved parents strive to make sure their kids study together with "good" peers, while the notion of "good", in this case, may include socioeconomic status alongside academic excellence. The model introduced in Section 2 can accommodate these extensions simply by replacing the ability curve with the "propensity to be a good peer" curve that reflects the combination of academic performance and socioeconomic status.

We further explore the possible role of parents in choices and show that students with more involved parents choose German more often than Spanish. In particular, we highlight the indirect evidence that parents play an active role in the language choice of their children. We use binary variables that indicate whether a parent serves as a class representative and is part of the Parents' Association as a proxy for involvement in a students' schooling. The last two columns in Table 3 show that, in line with the intuition presented, students choosing German are about 3% more likely to have parents who are actively involved in the scholarly life of their children.[15]

4.2 Varying Cost and Selectivity of Sorting

One of the key components of our theoretical model is the difference in costs between taking German and Spanish classes. Conjecture 2 gives a clear prediction over changes in average abilities in French-learning and German-learning classes as the cost increases the selectivity of sorting increases. France shares borders with both Germany and Spain. We use the variation in the proximity to these countries as a source of variation in the cost. We assume that the cost associated with studying German is lower near the German border and higher near the Spanish border.[16] This assumption is based on the fact

[15]One may think that parents of students choosing German also put more effort into advancing the academic progress of their children. While this may sometimes be the case, it is not always true, as parents of children attending better schools have been shown to put in less effort in Pop-Eleches and Urquiola (2013).

[16]In Appendix I, we provide a structural estimation of cost in each department of France separately to illustrate how the cost increases as departments become more distant from the border with Germany.

	Socioeconomic Status								Parental Involvement	
	Income	N room	Own room	M Uni	F Uni	Born in F	M Born in F	F Born in F	Representative	Association
German class	236***	0.260***	0.023**	0.069***	0.075***	0	0.016**	0.023***	0.029***	0.040***
	(< 0.001)	(< 0.001)	(0.004)	(< 0.001)	(< 0.001)	(1.000)	(0.022)	(0.001)	(< 0.001)	(< 0.001)
Constant	2,988***	5.198***	0.780***	0.308***	0.281***	0.973***	0.833***	0.825***	0.098***	0.135***
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
N	11,895	20,839	21,369	20,092	18,204	20,449	20,654	18,665	21,299	21,307

Probability p in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Difference in background characteristics between students who study German and Spanish. Socioeconomic status indicators include a difference in monthly income in euros (*Income*), number of rooms in the house (*N Rooms*), whether the child has their own room (*Own Room*), whether the mother has a university degree *(M Uni*), whether the father has a university degree (*F Uni*), whether the child was born in France (*Born in F*), whether the mother was born in France (*M Born in F*), and whether the father was born in France (*F Born in F*). Parental involvement indicators include the probability that one of the parents is a class representative (*Representative*) and is part of the Parents' Association (*Association*).

that, in a border area, students are more likely to have family members speaking the corresponding foreign language and to have an opportunity to commute to work across the border in the future. By comparing the strength of sorting effects in the region close to Germany with sorting in the region close to Spain, we test Conjecture 2.

The available data allow us to identify the department – the unit of the administrative divisions of France – where the observed student attends a school. In particular, we identify the department in which students study using the department where they lived in 2007 at the start of the panel. There are 100 departments in our data, and we exclude four departments where we observe zero students studying German. We identify departments that belong to two regions that we refer to as "Close to Germany" and "Close to Spain"; Figure 7 illustrates the division. The "Close to Germany" region includes Lorraine and Alsace, the "Close to Spain" region includes Aquitane, Midi Pyrénées, and Languedoc.

Table 4 provides the results of the OLS estimation of Model (2). The coefficient for *Close* is negative and statistically significant for the majority of courses. It demonstrates that students studying Spanish living close to Germany have worse academic progress compared to students studying Spanish living close to Spain. The coefficient for German class×Close to Germany demonstrates the same effect for students studying German – those who live close to Germany perform worse than those who live close to Spain. These findings confirm predicted comparative statics in Conjecture 2. Moreover, sorting effects in the "Close to Germany" region disappear for almost half of the tests – German class+German class×Close to Germany is not statistically significant for some French Language tests and is statistically significant for the Mathematics part of the National Test at 5%.

4.3 Peer Effects

The previous section has shown that students studying German have better grades before making the choice of a second foreign language, but this is also true if one considers exams at the end of middle school. In this section, we show that the performance of German-learning students compared to Spanish-learning students improved more between

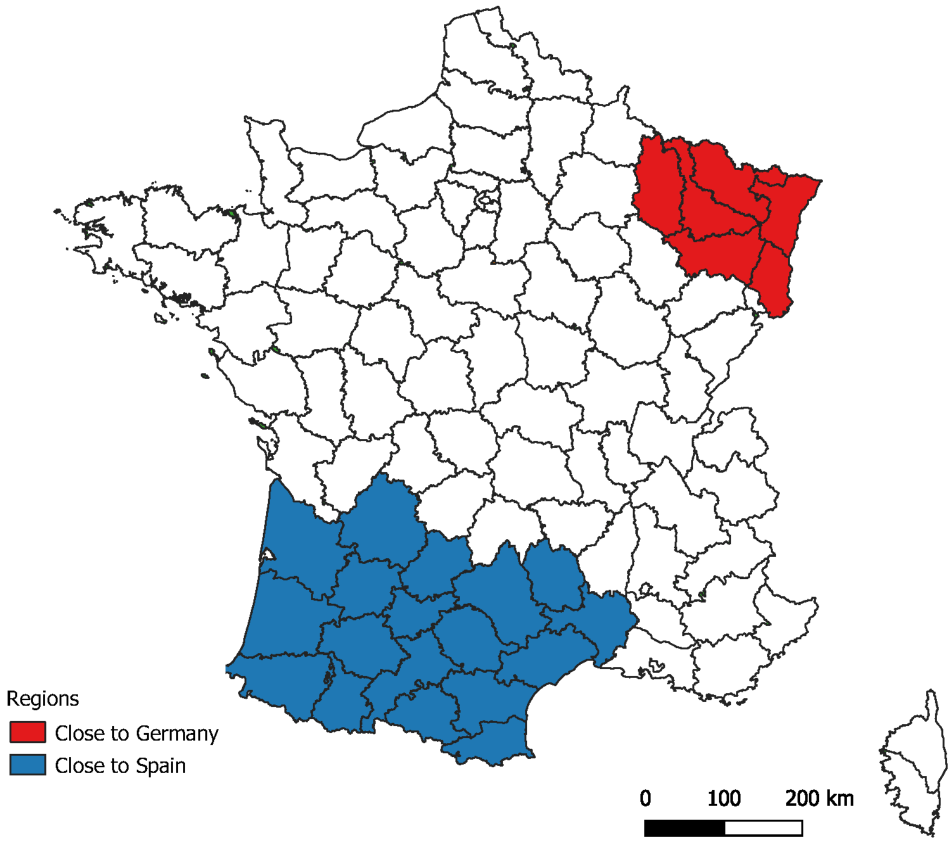

Figure 3: Map of France with "Close to Germany" and "Close to Spain" regions indicated.

2009 and 2011, which suggests that peer effects play an important role in the academic performance of students.

We perform propensity score matching on all observables listed in Tables 2 and 3. Matching achieves balance, and results are reported in Appendix H. Results of matching models that use only subsets of observables are available in Appendix G. Hereafter, we focus on the results that are obtained based on matching with all observables. In this case, the matched sample contains almost 7,000 matches.

When it comes to the Specific Standardized Test, the results are not crystal clear. While the effect is always positive, it is rarely significant when we do the matching using all the observables (results are available in Appendices G and D). We attribute this non-significance of the results to two factors. First, students study together for only one year before taking the test, which is possibly too little time for peer effects to reveal themselves. Second, the specific test was not incentivized, unlike the National Diploma, which plays an important role in students' life paths. As evidence of the importance of the National Diploma, we provide an illustrative graph National Diploma exam scores in Appendix F. In this graph, effects of bunching can be observed – there are spikes in exam scores around important cut-offs, such as minimum score (10 out of 20) and high distinction (see Diamond and Persson (2016) for an analysis of exam scores bunching in high-stakes tests).

Table 5 contains the results of the PSM-based estimation of peer effects for a number of courses from the National Diploma. The results suggest students studying German tend to demonstrate better academic performance in the National Diploma than similar students studying Spanish. The National Diploma includes two types of grades: contin-

	National Test		Special Test				
	Mathematics	French Language	Mathematics	French Language			
				Treatment of Incomplete Sentences	Understanding	Lexicon	Reasoning
German class	1.221***	1.325***	1.832***	1.560***	1.355***	1.363***	1.574***
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
Close to Germany	−0.559***	−0.778***	−0.495**	−0.989***	−0.501**	−0.531***	−0.336
	(0.004)	(< 0.001)	(0.013)	(< 0.001)	(0.021)	(0.001)	(0.140)
German class×Close to Germany	−0.696*	−0.976**	−1.117***	−0.811*	−0.960**	−0.858***	−1.382***
	(0.063)	(0.016)	(0.004)	(0.073)	(0.021)	(0.007)	(0.002)
Constant	13.835***	12.085***	11.702***	9.595***	13.509***	12.146***	10.856***
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
German class+ German class×Close to Germany	0.525**	0.349	0.715***	0.749***	0.396	0.505***	0.192
	(0.019)	(0.149)	(0.002)	(0.006)	(0.117)	(0.009)	(0.469)
	3255	3278	3538	3538	3538	3538	3538

Probability p in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Difference in results between students who choose German and Spanish and regions (close to Germany or close to Spain) for courses that were part of the 2007 National Test and the 2008 Specific Test. Estimates and confidence intervals are obtained using the OLS.

uous assessments and exams. Continuous assessments are grades given during the school year by the professor–those are not anonymous and might depend on the level of the classmates. For this reason, we mostly focus on exam grades that are anonymous, while exams are unified across schools. Table 5 reports estimated effects for all exams for the Diploma.[17] The Final Grade is a weighted average of all National Diploma courses, where scores for Mathematics and French Language are the most impactful, and the effect on grade is positive and significant at 10%. Effects for course exams are positive, but significant at 5% only for Mathematics. Interestingly, the estimated effect on the French language exam score is both small in amplitude and insignificant, speaking against the hypothesis that German is taken by those who are (or want to signal about being) good at learning languages. Further, the effect is large and statistically significant at 1% for the History and Geography exam. We also consider two courses with continuous assessment marks. First, we include Physics and Chemistry, where one can expect an education production function that strongly depends on peers. Indeed, we find a positive effect that is large and significant. Second, we include Sports as a placebo test, as it presumably has the least peer-dependent educational production function. The placebo test is passed with no differences in grades for Sports.

We perform a robustness check by matching students directly on observables using the Generic Algorithm (GA). PSM-based and GM-based estimates are very similar across courses included in National Diploma, with GM-based effects being statistically significant but more conservative with slightly smaller sizes of estimated effects (see Table 18

[17]We investigate the implications of the variation in the strength of sorting effects, established in Section 4.2, on peer effects. For this, we estimate peer effects separately for three regions, expecting the comparative statics of peers effects to follow sorting effects. Since splitting the sample into three sub-samples already substantially reduces the sample size for each region, we aim to minimize the loss of observations. For that reason, we allow for one difference in the estimation process and do not include revenue among socioeconomic characteristics, as there are many missing observations for this variable. The results are under-powered due to sample splitting indicating, that for French Language and History and Geography (but not so much for Mathematics) exams, the predicted pattern may hold – the further away from Germany, the stronger the peer effects, as sorting of the peers is more intense (see Figure 7 in Appendix J).

	Final Grade	Math	French Language	History & Geography	Physics & Chemistry	Sports
	Exam & Cont.	Exam	Exam	Exam	Cont.	Cont.
German class	0.21*	0.391**	0.128	0.438***	0.404***	0.076
	(0.062)	(0.042)	(0.35)	(0.002)	(0.008)	(0.453)
N	6980	6913	6918	6706	6970	6884

Probability p in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Differences in grades for National Diploma courses between matched students. Exam grades (Exam) and continuous assessment grades (Cont.) are given for different courses.

in Appendix K for an illustrative comparison).

We assume that there is no difference in unobservable characteristics between students studying German and students they have been matched with. In that case, we can interpret those results as evidence of peer effects. Here, we do not distinguish between different channels through which peers affect performance, which may include learning from peers, the overall speed of class progress, and teachers tailoring lectures, among others.[18] The results suggest that students choosing German perform better in many important courses because of the class composition, which is a product of endogenous tracking. One can treat the identified differences in peer effects as evidence of the inequality in the educational experience of students choosing different second foreign languages.[19]

5 Discussion

Many countries around the world have shaped their educational systems to harmonize students' middle school experience. Over the last half-century, many European countries, including France, have abandoned early tracking to tackle inequality in education (Betts, 2011). This paper shows that educational system features, for example, fixing class composition for all courses based on the second foreign language choice, may lead to endogenous tracking and result in ability grouping. We identify sorting effects within schools that are less evident than, for example, widely studied sorting between schools and corresponding residential areas (Black, 1999; Rothstein, 2006; Bayer et al., 2007). Nevertheless, we show that within-school endogenous tracking is associated with sorting, not only by ability but also by socioeconomic status, working against building an inclusive learning environment for students.

The most natural policy recommendation to address the inequalities generated by endogenous tracking would be to abandon the class formation rule that is based on

[18] In French schools, the average class is 21 students, and there is no evidence that German-learning and Spanish-learning classes systematically differ in sizes (see OECD (2015) for an overview). Moreover, existing evidence on the effects of class size on academic performance is mixed. While many papers, including Hoxby (2000a) and Urquiola (2006), report positive effects of smaller classes, more recent studies such as Angrist et al. (2017) and Angrist et al. (2019) find little to no effects.

[19] There might be more long-term consequences on both academic performance and future earnings, as in Carrell et al. (2018), where authors demonstrate negative long-run labor market effects from having disruptive peers at school. Unfortunately, we do not have data to match labor market outcomes with endogenous tracking effects.

costly course choices. However, we would like to point out that such a recommendation does not take into account probable general equilibrium effects that would follow such an intervention. In particular, one can anticipate that eliminating sorting mechanisms within schools may amplify between-schools sorting, leading to an increase in the number of selective and private schools.[20] In that case, students of different academic performances and socioeconomic backgrounds are less likely to study together, even in primary schools. Such an early selectivity would have detrimental effects on the equality of educational opportunities (Wößmann et al., 2007). Note that the 2016 French education reform designed to tackle inequalities, including those generated by within-school grouping rules at the middle school level, was mainly abolished in 2017.

We use variation in the ability composition of classes to identify peer effects, and reported effects are likely to encompass both direct and indirect effects. Importantly, our results suggest that students consider those peer effects when making their strategic educational choices and correctly anticipate other students doing so. While we show the role of the peer-seeking motive in within-school sorting, the same forces are likely to influence the choices of schools and universities, affecting the behavior of both sides of the market. In this way, our results are consistent with Abdulkadiroğlu et al. (2020), and show that peer preferences play a non-negligible role in educational choices supporting the relevance of matching models that incorporate peer preferences (Dutta and Massó, 1997; Echenique and Yenmez, 2007; Leshno, 2021; Pycia and Yenmez, 2021) for designing school and college admission systems.

References

Abadie, A., Imbens, G.W., 2006. Large sample properties of matching estimators for average treatment effects. Econometrica 74, 235–267.

Abadie, A., Imbens, G.W., 2008. On the failure of the bootstrap for matching estimators. Econometrica 76, 1537–1557.

Abdulkadiroğlu, A., Angrist, J., Pathak, P., 2014. The elite illusion: Achievement effects at boston and new york exam schools. Econometrica 82, 137–196.

Abdulkadiroğlu, A., Pathak, P.A., Schellenberg, J., Walters, C.R., 2020. Do parents value school effectiveness? American Economic Review 110, 1502–39.

Aimone, J.A., Iannaccone, L.R., Makowsky, M.D., Rubin, J., 2013. Endogenous group formation via unproductive costs. Review of Economic Studies 80, 1215–1236.

Altonji, J.G., 1995. The effects of high school curriculum on education and labor market outcomes. The Journal of Human Resources 30, 409.

Ammermueller, A., Pischke, J.S., 2009. Peer effects in european primary schools: Evidence from the progress in international reading literacy study. Journal of Labor Economics 27, 315–348.

[20] Bertola (2017) analyzes private and public schools in France and reports that, at present, students admitted to public schools have *higher* abilities than those admitted to private schools.

Angrist, J.D., Battistin, E., Vuri, D., 2017. In a small moment: Class size and moral hazard in the Italian Mezzogiorno. American Economic Journal: Applied Economics 9, 216–49.

Angrist, J.D., Lavy, V., Leder-Luis, J., Shany, A., 2019. Maimonides' rule redux. American Economic Review: Insights 1, 309–24.

Arkhangelsky, D., Athey, S., Hirshberg, D.A., Imbens, G.W., Wager, S., 2021. Synthetic difference-in-differences. American Economic Review 111, 4088–4118.

Azmat, G., Möller, M., 2018. The distribution of talent across contests. The Economic Journal 128, 471–509.

Bayer, P., Ferreira, F., McMillan, R., 2007. A unified framework for measuring preferences for schools and neighborhoods. Journal of Political Economy 115, 588–638.

Behaghel, L., De Chaisemartin, C., Gurgand, M., 2017. Ready for boarding? the effects of a boarding school for disadvantaged students. American Economic Journal: Applied Economics 9, 140–64.

Bertola, G., 2017. France's almost public private schools. Labour 31, 225–244.

Betts, J.R., 2011. The economics of tracking in education, in: Handbook of the Economics of Education. Elsevier. volume 3, pp. 341–381.

Beuchert, L., Humlum, M.K., Nielsen, H.S., Smith, N., 2018. The short-term effects of school consolidation on student achievement: Evidence of disruption? Economics of Education Review 65, 31–47.

Black, S.E., 1999. Do better schools matter? parental valuation of elementary education. The Quarterly Journal of Economics 114, 577–599.

Booij, A.S., Leuven, E., Oosterbeek, H., 2017. Ability peer effects in university: Evidence from a randomized experiment. The Review of Economic Studies 84, 547–578.

Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. Journal of Econometrics 150, 41–55.

Bui, S.A., Craig, S.G., Imberman, S.A., 2014. Is gifted education a bright idea? Assessing the impact of gifted and talented programs on students. American Economic Journal: Economic Policy 6, 30–62.

Bursztyn, L., Egorov, G., Jensen, R., 2019. Cool to be smart or smart to be cool? Understanding peer pressure in education. The Review of Economic Studies 86, 1487–1526.

Burton, M.D., Dahl, M.S., Sorenson, O., 2018. Do start-ups pay less? ILR Review 71, 1179–1200.

Card, D., Giuliano, L., 2016. Can tracking raise the test scores of high-ability minority students? American Economic Review 106, 2783–2816.

Carrell, S.E., Hoekstra, M., Kuka, E., 2018. The long-run effects of disruptive peers. American Economic Review 108, 3377–3415.

Carrell, S.E., Sacerdote, B.I., West, J.E., 2013. From natural variation to optimal policy? The importance of endogenous peer group formation. Econometrica 81, 855–882.

Dehejia, R., 2005. Practical propensity score matching: a reply to Smith and Todd. Journal of Econometrics 125, 355–364.

Dehejia, R.H., Wahba, S., 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. Journal of the American Statistical Association 94, 1053–1062.

Dehejia, R.H., Wahba, S., 2002. Propensity score-matching methods for nonexperimental causal studies. Review of Economics and Statistics 84, 151–161.

DEPP, A., 2018. Panel d'élèves du second degré, recrutement 2007 - 2007-2013. Ministère de l'Éducation.

Diamond, A., Sekhon, J.S., 2013. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. Review of Economics and Statistics 95, 932–945.

Diamond, R., Persson, P., 2016. The long-term consequences of teacher discretion in grading of high-stakes tests. Technical Report. National Bureau of Economic Research.

Duflo, E., Dupas, P., Kremer, M., 2011. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. American Economic Review 101, 1739–74.

Duru-Bellat, M., Mingat, A., 1997. La constitution de classes de niveau dans les collèges; les effets pervers d'une pratique à visée égalisatrice. Revue française de sociologie , 759–789.

Dutta, B., Massó, J., 1997. Stability of matchings when individuals have preferences over colleagues. Journal of Economic Theory 75, 464–475.

Echenique, F., Yenmez, M.B., 2007. A solution to matching with preferences over colleagues. Games and Economic Behavior 59, 46–71.

Epple, D., Romano, R.E., 2011. Peer effects in education: A survey of the theory and evidence, in: Handbook of Social Economics. Elsevier. volume 1, pp. 1053–1163.

Falk, A., Knell, M., 2004. Choosing the Joneses: Endogenous goals and reference standards. Scandinavian Journal of Economics 106, 417–435.

Feld, J., Zölitz, U., 2017. Understanding peer effects: On the nature, estimation, and channels of peer effects. Journal of Labor Economics 35, 387–428.

Figlio, D.N., Page, M.E., 2002. School choice and the distributional effects of ability tracking: does separation increase inequality? Journal of Urban Economics 51, 497–514.

Frymer, P., Grumbach, J.M., 2021. Labor unions and white racial politics. American Journal of Political Science 65, 225–240.

Hanushek, E.A., Wößmann, L., 2006. Does early tracking affect educational inequality and performance? differences-in-differences evidence across countries. Economic Journal 116, 63–76.

Herbaut, E., Barone, C., Ichou, M., Vallet, L.A., 2019. Tracking and sorting in the french educational system. DIAL Working Paper Series 13/2019 .

Hopkins, D.J., 2010. Politicized places: Explaining where and when immigrants provoke local opposition. American Political Science Review 104, 40–60.

Hoxby, C.M., 2000a. The effects of class size on student achievement: New evidence from population variation. Quarterly Journal of Economics 115, 1239–1285.

Hoxby, C.M., 2000b. Peer effects in the classroom: Learning from gender and race variation.

Imbens, G.W., 2015. Matching methods in practice: Three examples. Journal of Human Resources 50, 373–419.

Kiessling, L., Radbruch, J., Schaube, S., 2021. Self-selection of peers and performance. Management Science .

LaLonde, R.J., 1986. Evaluating the econometric evaluations of training programs with experimental data. American Economic Review , 604–620.

Lavy, V., Silva, O., Weinhardt, F., 2012. The good, the bad, and the average: Evidence on ability peer effects in schools. Journal of Labor Economics 30, 367–414.

Lazear, E.P., 2001. Educational production. The Quarterly Journal of Economics 116, 777–803.

Lefgren, L., 2004. Educational peer effects and the chicago public schools. Journal of Urban Economics 56, 169–191.

Leshno, J., 2021. Stable matching with peer-dependent preferences in large markets: Existence and cutoff characterization. Available at SSRN 3822060 .

Ly, S., Maurin, É., Riegert, A., 2014. La mixité sociale et scolaire en Île-de-France: le rôle des établissements. Ph.D. thesis. Institut des politiques publiques (IPP).

Ly, S.T., Riegert, A., 2014. Persistent classmates: How familiarity with peers protects from disruptive school transitions .

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

Morgan, J., Sisak, D., Várdy, F., 2018. The ponds dilemma. The Economic Journal 128, 1634–1682.

Oakes, J., 2005. Keeping track: How schools structure inequality. Yale University Press.

OECD, 2015. Education at a glance 2015 URL: https://www.oecd-ilibrary.org/content/publication/eag-2015-en, doi:https://doi.org/https://doi.org/10.1787/eag-2015-en.

Ouimet, P., Zarutskie, R., 2014. Who works for startups? the relation between firm age, employee age, and growth. Journal of Financial Economics 112, 386–407.

Pop-Eleches, C., Urquiola, M., 2013. Going to a better school: Effects and behavioral responses. American Economic Review 103, 1289–1324.

Pycia, M., Yenmez, M.B., 2021. Matching with externalities. University of Zurich, Department of Economics, Working Paper .

Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

Rothstein, J.M., 2006. Good principals or good peers? Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions. American Economic Review 96, 1333–1350.

Sacerdote, B., 2001. Peer effects with random assignment: Results for Dartmouth roommates. The Quarterly journal of Economics 116, 681–704.

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. Elsevier. volume 3, pp. 249–277.

Smith, J.A., Todd, P.E., 2001. Reconciling conflicting evidence on the performance of propensity-score matching methods. American Economic Review 91, 112–118.

Smith, J.A., Todd, P.E., 2005. Does matching overcome lalonde's critique of nonexperimental estimators? Journal of Econometrics 125, 305–353.

Spence, M., 1973. Job market signaling. Quarterly Journal of Economics 87, 355–374.

Taber, C., Vejlin, R., 2020. Estimation of a roy/search/compensating differential model of the labor market. Econometrica 88, 1031–1069.

Tangvatcharapong, M., 2020. The impact of school tracking and peer quality on student achievement: Regression discontinuity evidence from Thailand. Mimeo .

Tiebout, C.M., 1956. A pure theory of local expenditures. Journal of Political Economy 64, 416–424.

Train, K.E., 2009. Discrete choice methods with simulation. Cambridge University Press.

Urquiola, M., 2006. Identifying class size effects in developing countries: Evidence from rural Bolivia. Review of Economics and Statistics 88, 171–177.

Villeval, M.C., 2020. Performance feedback and peer effects: A review. Available at SSRN 3543371 .

Weiss, A., 1995. Human capital vs. signalling explanations of wages. Journal of Economic Perspectives 9, 133–154.

Williams, D.R., 2006. The economic returns to multiple language usage in Western Europe. IRISS Working Paper Series .

Windsteiger, L., 2021. Monopolistic supply of sorting, inequality, and welfare. Journal of Public Economic Theory 23, 801–821.

Wößmann, L., Lüdemann, E., Schütz, G., West, M.R., 2007. School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003 .

Zimmer, R., 2003. A new twist in the educational tracking debate. Economics of Education Review 22, 307–315.

APPENDIX

A Model with Complete Information and Alternative Peer Effects Production Function

In this section, we model endogenous tracking in the simultaneous move game of complete information. We study conditions under which the self-sorting of students, when better-performing students choose German class and the rest – Spanish, constitute a Nash Equilibrium of the game. In the model's setup, we follow two main assumptions – there is a sunk cost of choosing German over Spanish, and there are peer effects.

Consider n-players simultaneous-move game. The game is of complete information– player's utility functions, payoffs, strategies, "types" and other model components are common knowledge. Each player chooses one of two languages, $l \in \{g, s\}$. Each player is endowed with her own "ability" denoted by a_i.[21] Without loss of generality, assume that players are indexed from the highest to lowest ability: $a_1 > a_2 > \cdots > a_n$. We will refer to the graph of these ranked abilities as the "ability curve". Each language has an (opportunity) cost denoted by c_l: e.g., c_g represents the cost associated with learning the language g.

Assumption 2. $c_g > c_s$ for all player i.

Assumption 2 allows us to derive a separating equilibrium.

Denote by G_l the group of players who chooses language l. Denote by $p_i(G_l)$ the peer effect of player i in G_l. $p_i(G_l)$ is a mapping from the set of all possible groups to a real number: i.e., $p_i : \{G_l\} \to \mathbb{R}$.

Assumption 3. We assume that

$$p_i(G_l) = \frac{1}{|G_l|-1} \sum_{j \in G_l} a_j.$$

Assumption 3 says that the player i's peer effect is the average of the endowed abilities of other players in the group to which player i belongs. This assumption is standard in the literature on peer effects in education. Note that assumption 3 implies that $p_i(G_l)$ is increasing in the player's index i: given any G_l, for any $k < m$,

$$p_k(G_l) = \frac{1}{|G_l|-1} \sum_{j \in G_l \setminus \{k\}} a_j < p_m(G_l) = \frac{1}{|G_l|-1} \sum_{j \in G_l \setminus \{m\}} a_j.$$

In other words, the lower a_i is, the higher $p_i(G_l)$ is: the lower your ability is, the higher benefit from your peer is.

Each player in each group enjoys a positive peer effect and bears the cost of the chosen language. Each player's payoff is her educational benefit. Each player's educational benefit increases in her own ability a_i and the peer effect she experiences in G_l *but*

[21] We mainly refer to ability in the context of academic performance. At the same time, one may extend the model to absorb more students' characteristics and study "propensity to be a good peer" instead.

decreases in the cost of choosing a language. We assume that the educational benefit is linear in these three factors. Hence, the payoff of player i in G_l,

$$u_i(l, G_l) = a_i + p_i(G_l) - c_l.$$

Given any strategy profile, we denote by $\bar{a}_l$ the highest ability in group G_l for $l = g, s$. We call the player with $\bar{a}_l$ player l for $l = g, s$. We denote by $p_l(G_l)$ the player l's peer effect for $l = g, s$.

Lemma 2. *Given any group composition, if the player with the highest ability in each group does not want to deviate to the other group, it is a Nash equilibrium outcome.*

Denote by $\bar{a}_l$ the highest ability in group G_l for $l = g, s$. We call the player with $\bar{a}_l$ player l for $l = g, s$. Consider any group composition, G_g and G_s. Player i in G_g does not want to deviate if and only if

$$\begin{aligned} & a_i + p(G_g) - c_g \leq a_i + p(G_s \cup \{a_i\}) - c_s \\ \iff \quad & p(G_g) - p(G_s \cup \{a_i\}) \leq c_g - c_s. \end{aligned}$$

Since $p(G_s \cup \{a_i\})$ decreases in a_i, we have

$$p(G_g) - p(G_s \cup \{\bar{a}_g\}) \leq p(G_g) - p(G_s \cup \{a_i\}) \text{ for any } a_i \in G_g.$$

In the same manner, player j in G_s does not want to deviate if and only if

$$\begin{aligned} & a_j + p(G_s) - c_s \leq a_j + p(G_g \cup \{a_j\}) - c_g \\ \iff \quad & c_g - c_s \leq p(G_g \cup \{a_j\}) - p(G_s). \end{aligned}$$

Again $p(G_g \cup \{a_j\})$ decreases in a_j. Hence, we have

$$p(G_g \cup \{\bar{a}_s\}) - p(G_s) \leq p(G_g \cup \{a_j\}) - p(G_s) \text{ for any } \ a_j \in G_s.$$

Hence, it is sufficient to check whether the player with the highest ability in each group wants to deviate to the other group.

Consider the following group composition in which the players are sorted in the ascending order of their abilities:

$$G_s = \{a_1, \ldots, a_k\} \text{ and } G_g = \{a_{k+1}, \ldots, a_n\}.$$

We refer to such a group composition as a monotonic sorting outcome k as the k-th student is pivotal. For a monotonic sorting outcome k, let us also denote an average ability of students taking Spanish class by $\overline{a}_S^k$ and by $\overline{a}_G^k$–those who take German. Denote by C by the cost difference between language g and l.

Proposition 1. *A monotonic sorting outcome k is a Nash equilibrium outcome if and only if* $\dfrac{a_k + .. + a_N}{N - k + 1} - \overline{a}_S^k \leq C \leq \overline{a}_G^k - \dfrac{a_1 + .. + a_k + a_N}{k + 1}$.

The proof of Proposition 1 is available further in the text. Proposition 1 naturally applies Lemma 1 to derive necessary and sufficient conditions for equilibrium. The player whose ability is the highest in each group benefits the least from her peers. If the player

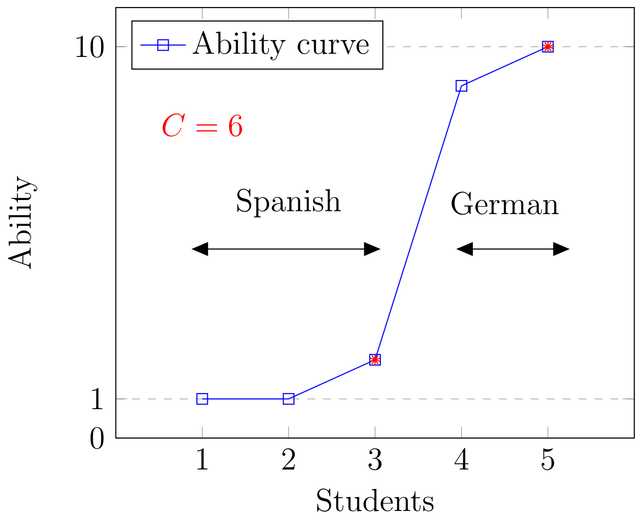

Figure 4: An example of the sorting equilibrium in a cohort of students with abilities $\{1, 1, 2, 9, 10\}$ and relative cost of choosing German language is $C = 6$. Students with highest ability in each group are marked with red starts–their incentives not to deviate are ensured by the relative cost supporting an equilibrium.

who benefits the least from her peers does not have an incentive to deviate, the other players who benefit more than her would not deviate either. Intuitively, the proposition formalizes the idea that for sorting outcome to be supported with a given cost, there should be both enough homogeneity of students within groups and enough heterogeneity between groups.

Example. Consider the cohort of students with the following set of abilities: $\{1, 1, 2, 9, 10\}$. Further assume that the relative cost of choosing German language is $c_g - c_s = C = 6$. Using Proposition 1, we can check if sorting outcome when a student with abilities $\{1, 1, 2\}$ choose Spanish class and student with abilities $\{9, 10\}$ choose German class constitutes a Nash equilibrium of the game.

Using Proposition 1 we can check if sorting outcome with $k = 3$ is a Nash equilibrium:

$$\frac{2+9+10}{3} - \frac{1+1+2}{3} \leq 6 \leq \frac{9+10}{2} - \frac{1+1+2+10}{4}$$

$$\iff \quad 5.6 \leq 6 \leq 6.$$

Hence, $G_g = \{9, 10\}$ and $G_s = \{1, 1, 2\}$ is an equilibrium sorting outcome.

A.1 Proof of Proposition 1

□ Following Lemma 1, it is sufficient to show that neither a_k would deviate and join German class nor a_N would deviate and join Spanish class.

$$\begin{cases} \dfrac{a_1 + .. + a_k}{k} \geq \dfrac{a_k + .. + a_N}{N-k+1} - C \\ \dfrac{a_k + 1 + .. + a_N}{N-k} - C \geq \dfrac{a_1 + .. + a_k + a_N}{k+1} \end{cases}$$

which can be rewritten as:

$$\frac{a_k + .. + a_N}{N - k + 1} - \frac{a_1 + .. + a_k}{k} \le C \le \frac{a_k + 1 + .. + a_N}{N - k} - \frac{a_1 + .. + a_k + a_N}{k + 1}.$$

We use previously introduced notations $\overline{a}_G^k$ and $\overline{a}_S^k$ to further simplify the expression:

$$\frac{a_k + .. + a_N}{N - k + 1} - \overline{a}_S^k \le C \le \overline{a}_G^k - \frac{a_1 + .. + a_k + a_N}{k + 1}.$$

The final expression is a condition used in Proposition 1. ■

B Omitted Lemmas and Proofs

B.1 Lemma B.1

Lemma B.1. *Consider any equilibrium, $s^* = (s_1^*, \cdots, s_n^*)$. Then, s_i^* is a threshold strategy for all $i = 1, , n$.*

Proof. Consider $s_{-i}^* = (s_1^*, \cdots, s_{i-1}^*, s_{i+1}^*, \cdots, s_n^*)$. Then s_{-i}^* determines $E[P_i^e(g_G)|s_{-i}^*]$ and $E[P_i^e(g_S)|s_{-i}^*]$ which are constants and independent of player i's realized ability a_i. Given a_i, player i's payoff difference between two actions, G and S, is

$$\Delta u_i(a_i) = \Delta P_i^e(s_{-i}^*) - \Delta c(a_i),$$

and it increases in a_i because $\Delta c(a_i)$ decreases in a_i. Suppose that $\Delta u_i(\underline{a}) < 0$ and $\Delta u_i(\bar{a}) > 0$. Then, there exists $\tilde{a}$ such that

$$\Delta u_i(\tilde{a}) = \Delta P_i^e(s_{-i}^*) - \Delta c(\tilde{a}) = 0$$

because $\Delta u_i(\cdot)$ is continuous. Thus $s_i^*(a_i) = G$ for $a_i \ge \tilde{a}$ and $s_i^*(a_i) = S$ for $a_i < \tilde{a}$. Now suppose that $\Delta u_i(\bar{a}) \le 0$. Then $s_i^*(a_i)$ is a threshold strategy with $\tilde{a} = \bar{a}$, i.e., $s_i^*(a_i) = G$ for $a_i \ge \bar{a}$ and $s_i^*(a_i) = S$ for $a_i < \bar{a}$. Lastly, suppose that $\Delta u_i(\underline{a}) \ge 0$. Then $s_i^*(a_i)$ is a threshold strategy with $\tilde{a} = \underline{a}$. □

B.2 Proof of Lemma 1

Proof. $\Delta P_i^e(\hat{a})$ is continuous as $f(a)$ is continuous. Note that

$$\Delta P_i^e(\hat{a}) = \frac{\int_{\hat{a}}^{\bar{a}} af(a)da - (1 - F(\hat{a}))a_m}{F(\hat{a})(1 - F(\hat{a}))}$$

is indeterminate at either $\hat{a} = \underline{a}$ or $\hat{a} = \bar{a}$ because its numerator and denominator are both zero at $\hat{a} = \underline{a}$ or $\hat{a} = \bar{a}$. We can apply L'Hôpital's rule. The derivative of the numerator with respect to $\hat{a}$ is

$$\frac{\partial \int_{\hat{a}}^{\bar{a}} af(a)da}{\partial \hat{a}} - \frac{\partial (1-F(\hat{a}))a_m}{\partial \hat{a}} = \underbrace{-\hat{a}f(\hat{a})}_{\text{By the Leibniz integral rule}} + a_m f(\hat{a}) = f(\hat{a})(a_m - \hat{a}).$$

The derivative of the denominator with respect to $\hat{a}$ is

$$\frac{\partial F(\hat{a})(1-F(\hat{a}))}{\partial \hat{a}} = f(\hat{a})(1-F(\hat{a})) - F(\hat{a})f(\hat{a}) = f(\hat{a})(1-2F(\hat{a})).$$

Then,

$$\begin{aligned}
\lim_{\hat{a}\to\bar{a}} \Delta P_i^e(\hat{a}) &= \lim_{\hat{a}\to\bar{a}} \frac{f(\hat{a})(a_m - \hat{a})}{f(\hat{a})(1-2F(\hat{a}))} = \frac{a_m - \bar{a}}{-1} = \bar{a} - a_m, \text{ and} \\
\lim_{\hat{a}\to\underline{a}} \Delta P_i^e(\hat{a}) &= \lim_{\hat{a}\to\underline{a}} \frac{f(\hat{a})(a_m - \hat{a})}{f(\hat{a})(1-2F(\hat{a}))} = \frac{a_m - \underline{a}}{1} = a_m - \underline{a}.
\end{aligned}$$

□

C Descriptive statistics

Variable	Mean	SD	Min	Max	N
German	0.16	0.37	0	1	22,538
French 2007	11.70	3.63	0	20	18,184
Mathematics 2007	13.52	3.31	0	20	18,055
Mathematics 2008	11.48	3.54	0	20	20,548
French: Sentences 2008	9.23	4.21	0	20	20,548
French: Understanding 2008	13.25	3.91	0	20	20,548
French: Lexicon 2008	13.97	4.20	0	20	20,548
French: Reasoning 2008	10.70	4.03	0	20	20,548
monthlyrevenue	3,027.48	2,016.64	110	80,000	11,895
nbrroom	5.24	1.55	1	18	20,839
ownroom	0.78	0.41	0	1	21,369
motheruniversity	0.32	0.47	0	1	20,092
fatheruniversity	0.29	0.46	0	1	18,204
borninfrance	0.97	0.16	0	1	20,449
motherborninfrance	0.84	0.37	0	1	20,654
fatherborninfrance	0.83	0.38	0	1	18,665
parentsrepresentative	0.10	0.30	0	1	21,299
parentsinassociation	0.14	0.35	0	1	21,249

Table 6: Descriptive statistics.

D Matching with different sets of observables

Table 8, 9, and 10 show the results for the Brevet des Colleges grades and Table 7 shows the results for the standardized tests. We performed five different matchings for each outcome variable, where we gradually included more variables in the matching. The fifth column of the tables does the matching taking into account the 2007 National Standardized test, 2008 Specific Standardized test, the socioeconomic characteristics, and the parents' involvement. The more variables are included in the matching, the more credible the matching is. Therefore, the last column in each table shows the most credible estimates.

			Mathematics		
German class	0.306***	0.111†	0.200*	0.080	0.168
p-value	(< 0.001)	(0.067)	(0.020)	(0.384)	(0.114)
N	15617	14916	7970	6258	6229
N treated	2685	2574	1400	1116	1112
			French		
			Treatment of Incomplete Sentences		
German class	0.380***	0.148*	0.169†	0.277**	0.276*
p-value	(< 0.001)	(0.027)	(0.074)	(0.008)	(0.010)
N	15597	14900	7958	6251	6222
N treated	2685	2574	1400	1116	1112
			Understanding		
German class	0.287***	0.053	0.190	0.113	0.241†
p-value	(< 0.001)	(0.531)	(0.117)	(0.395)	(0.068)
N	15587	14887	7951	6242	6214
N treated	2682	2570	1398	1115	1111
			Lexicon		
German class	0.400***	0.131*	0.303***	0.336***	0.181*
p-value	(< 0.001)	(0.014)	(< 0.001)	(< 0.001)	(0.030)
N	15623	14920	7968	6255	6227
N treated	2690	2578	1402	1117	1113
			Reasoning		
German class	0.118	−0.112	−0.073	0.096	0.258*
p-value	(0.102)	(0.159)	(0.515)	(0.474)	(0.045)
N	15564	14867	7932	6231	6203
N treated	2683	2571	1396	1113	1109
National Standardized test	✓	✓	✓	✓	✓
Specific Standardized test	✗	✓	✓	✓	✓
Parent's Monthly Revenue	✗	✗	✓	✓	✓
Other socioeconomic	✗	✗	✗	✓	✓
Parent's involvement	✗	✗	✗	✗	✓

Table 7: Difference in standardized test in 2011. Genetic Matching with a population size of 50,000.

	Overall Average				
	Continuous Assessment and Exam				
German class	0.332***	0.259***	0.154*	0.189*	0.156†
p-value	(< 0.001)	(< 0.001)	(0.045)	(0.018)	(0.068)
N	17950	17039	9047	7014	6980
N treated	2992	2856	1547	1225	1220
	Continuous Assessment				
German class	0.338***	0.248***	0.214**	0.204**	0.134†
p-value	(< 0.001)	(< 0.001)	(0.003)	(0.010)	(0.098)
N	17950	17039	9047	7014	6980
N treated	2992	2856	1547	1225	1220
	Mathematics				
	Continuous Assessment				
German class	0.318***	0.186*	0.007	0.081	0.076
p-value	(< 0.001)	(0.028)	(0.953)	(0.529)	(0.556)
N	17914	17007	9032	7007	6973
N treated	2991	2855	1546	1224	1219
	Exam				
German class	0.387***	0.255**	0.238*	0.208	0.353*
p-value	(< 0.001)	(0.004)	(0.044)	(0.123)	(0.011)
N	17645	16771	8921	6946	6913
N treated	2957	2827	1527	1213	1208
	French				
	Continuous Assessment				
German class	0.353***	0.217***	0.141†	0.261**	0.239*
p-value	(< 0.001)	(< 0.001)	(0.085)	(0.004)	(0.011)
N	17915	17008	9032	7005	6971
N treated	2991	2855	1546	1224	1219
	Exam Average				
German class	0.204***	0.150*	0.161†	0.323**	0.147
p-value	(< 0.001)	(0.023)	(0.075)	(0.001)	(0.149)
N	17675	16796	8929	6951	6918
N treated	2957	2826	1527	1213	1208
	Dictation Exam				
German class	0.239**	0.145	0.167	0.357**	0.196
p-value	(0.001)	(0.103)	(0.169)	(0.008)	(0.153)
N	17669	16793	8928	6950	6917
N treated	2956	2825	1527	1213	1208
	Essay Exam				
German class	0.229**	0.106	0.189	0.268†	0.269†
p-value	(0.002)	(0.254)	(0.130)	(0.064)	(0.061)
N	17661	16786	8927	6950	6917
N treated	2956	2825	1527	1213	1208
National Standardized test	✓	✓	✓	✓	✓
Specific Standardized test	✗	✓	✓	✓	✓
Parent's Monthly Revenue	✗	✗	✓	✓	✓
Other socioeconomic	✗	✗	✗	✓	✓
Parent's involvement	✗	✗	✗	✗	✓

Table 8: Difference in grade in Brevet des College exam in 2011. Genetic Matching with a population size of 50,000.

	History and Geography				
	Continuous Assessment				
German class	0.332***	0.300***	0.326***	0.250*	0.271**
p-value	(< 0.001)	(< 0.001)	(< 0.001)	(0.014)	(0.008)
N	17261	16385	8683	6739	6706
N treated	2900	2766	1494	1178	1174
	Exam				
German class	0.204***	0.070	0.137	0.296**	0.154
p-value	(< 0.001)	(0.300)	(0.123)	(0.003)	(0.138)
N	17675	16796	8929	6951	6918
N treated	2957	2826	1527	1213	1208
	Physic and Chemistry				
	Continuous Assessment				
German class	0.377***	0.227**	0.222*	0.406***	0.329**
p-value	(< 0.001)	(0.003)	(0.033)	(< 0.001)	(0.007)
N	17902	16996	9029	7004	6970
N treated	2988	2852	1544	1223	1218
	Biology				
	Continuous Assessment				
German class	0.420***	0.254***	0.373***	0.256*	0.200†
p-value	(< 0.001)	(< 0.001)	(< 0.001)	(0.016)	(0.064)
N	17902	16996	9027	7001	6967
N treated	2990	2854	1546	1224	1219
	Technology				
	Continuous Assessment				
German class	0.301***	0.177*	0.163†	0.091	0.104
p-value	(< 0.001)	(0.010)	(0.076)	(0.351)	(0.292)
N	17872	16965	9012	6992	6959
N treated	2982	2846	1542	1222	1217
National Standardized test	✓	✓	✓	✓	✓
Specific Standardized test	✗	✓	✓	✓	✓
Parent's Monthly Revenue	✗	✗	✓	✓	✓
Other socioeconomic	✗	✗	✗	✓	✓
Parent's involvement	✗	✗	✗	✗	✓

Table 9: Difference in grade in Brevet des College exam in 2011. Genetic Matching with a population size of 50,000.

	Plastic Art				
	Continuous Assessment				
German class	0.236***	0.189*	0.129	0.298**	0.116
p-value	(< 0.001)	(0.010)	(0.185)	(0.008)	(0.274)
N	17866	16961	9012	6991	6957
N treated	2984	2848	1543	1222	1217
	Music				
	Continuous Assessment				
German class	0.389***	0.348***	0.355***	0.241*	0.390***
p-value	(< 0.001)	(< 0.001)	(< 0.001)	(0.026)	(< 0.001)
N	17800	16901	8982	6968	6934
N treated	2974	2840	1535	1217	1212
	Sport				
	Continuous Assessment				
German class	−0.016	−0.040	0.016	0.007	−0.001
p-value	(0.752)	(0.545)	(0.856)	(0.943)	(0.990)
N	17644	16759	8911	6918	6884
N treated	2957	2824	1531	1213	1208
	Civic Education				
	Continuous Assessment				
German class	0.284***	0.144†	0.229*	0.196†	0.225†
p-value	(< 0.001)	(0.063)	(0.033)	(0.098)	(0.061)
N	15401	14616	7763	6034	6006
N treated	2600	2479	1348	1060	1056
	Conduct Mark				
	Continuous Assessment				
German class	0.324***	0.345***	0.087	0.107	0.105
p-value	(< 0.001)	(< 0.001)	(0.411)	(0.345)	(0.359)
N	17917	17010	9035	7006	6972
N treated	2990	2854	1546	1224	1219
National Standardized test	✓	✓	✓	✓	✓
Specific Standardized test	✗	✓	✓	✓	✓
Parent's Monthly Revenue	✗	✗	✓	✓	✓
Other socioeconomic	✗	✗	✗	✓	✓
Parent's involvement	✗	✗	✗	✗	✓

Table 10: Difference in grade in Brevet des College exam in 2011. Genetic Matching with a population size of 50,000.

E Check for balance in observables

	National Evaluation 2007				
	French				
Difference	0.000	-0.001	-0.025	0.003	-0.002
t-test p-value	(0.911)	(0.930)	$(0.085^{\dagger})$	(0.915)	(0.942)
KS stat	0.002	0.015	0.014	0.020	0.020
KS bootstrapped p-value	(0.911)	(0.930)	(0.085)	(0.915)	(0.942)
	Mathematics				
Difference	-0.000	0.007	-0.014	-0.024	-0.003
t-test p-value	(0.876)	(0.474)	(0.683)	(0.674)	(0.927)
KS stat	0.001	0.009	0.016	0.018	0.031
KS bootstrapped p-value	(0.876)	(0.474)	(0.683)	(0.674)	(0.927)
	Specific Evaluation 2008				
	Mathematics				
Difference	0.254	0.015	0.037	0.007	0.081
t-test p-value	$(< 0.001^{***})$	$(0.070^{\dagger})$	(0.009^{**})	(0.826)	$(< 0.001^{***})$
KS stat	0.042	0.009	0.013	0.028	0.027
KS bootstrapped p-value	$(< 0.001^{***})$	(0.070)	(0.009)	(0.826)	(0.001)
	French: Treatment of Incomplete Sentences				
Difference	0.270	0.000	0.016	0.061	0.052
t-test p-value	(0.001^{**})	(1.000)	(0.519)	(0.118)	(0.277)
KS stat	0.046	0.019	0.021	0.030	0.018
KS bootstrapped p-value	$(< 0.001^{***})$	(1.000)	(0.519)	(0.118)	(0.277)
	French: Understanding				
Difference	0.256	-0.111	0.013	-0.089	-0.108
t-test p-value	(0.002^{**})	(0.022^{*})	(0.524)	(0.035^{*})	$(0.073^{\dagger})$
KS stat	0.037	0.022	0.016	0.023	0.034
KS bootstrapped p-value	$(< 0.001^{***})$	(0.022)	(0.524)	(0.035)	(0.073)
	French: Lexicon				
Difference	0.311	0.004	0.009	0.037	0.044
t-test p-value	$(< 0.001^{***})$	(0.628)	(0.523)	(0.049^{*})	(0.035^{*})
KS stat	0.072	0.011	0.016	0.026	0.020
KS bootstrapped p-value	$(< 0.001^{***})$	(0.628)	(0.523)	(0.049)	(0.035)
	French: Reasoning				
Difference	0.251	-0.011	0.021	-0.161	-0.028
t-test p-value	(0.004^{**})	(0.473)	(0.810)	(0.020^{*})	(0.522)
KS stat	0.052	0.015	0.024	0.038	0.025
KS bootstrapped p-value	$(< 0.001^{***})$	(0.473)	(0.810)	(0.020)	(0.522)
National Standardized test	✓	✓	✓	✓	✓
Specific Standardized test	✗	✓	✓	✓	✓
Parent's Monthly Revenue	✗	✗	✓	✓	✓
Other socioeconomic	✗	✗	✗	✓	✓
Parent's involvement	✗	✗	✗	✗	✓

Table 11: Balance check with genetic Matching. Population size of 10,000.

	Socioeconomic				
			monthlyrevenue		
Difference	35.800	5.434	56.181	-20.220	92.235
t-test p-value	(0.610)	(0.935)	(0.365)	(0.779)	(0.119)
KS stat	0.020	0.039	0.021	0.022	0.036
KS bootstrapped p-value	(0.610†)	(0.935)	(0.365)	(0.779)	(0.119)
			nbrroom		
Difference	0.168	0.143	0.154	0.023	0.023
t-test p-value	(< 0.001***)	(< 0.001***)	(0.003**)	(0.011*)	(0.021*)
KS stat	0.043	0.034	0.036	0.007	0.014
KS bootstrapped p-value	(< 0.001***)	(0.001*)	(0.003†)	(0.011)	(0.021)
			ownroom		
Difference	-0.010	-0.010	-0.010	-0.007	-0.033
t-test p-value	(0.328)	(0.362)	(0.471)	(0.005**)	(0.014*)
			motheruniversity		
Difference	0.016	0.004	0.014	0.022	0.021
t-test p-value	(0.186)	(0.763)	(0.384)	(0.009**)	(0.054†)
			fatheruniversity		
Difference	0.025	0.018	0.013	0.026	0.012
t-test p-value	(0.054†)	(0.162)	(0.431)	(0.028*)	(0.404)
			borninfrance		
Difference	-0.002	-0.005	-0.005	0.000	0.000
t-test p-value	(0.581)	(0.256)	(0.453)	(1.000)	(1.000)
			motherborninfrance		
Difference	-0.003	-0.017	-0.010	0.000	0.000
t-test p-value	(0.716)	(0.066†)	(0.419)	(1.000)	(1.000)
			fatherborninfrance		
Difference	0.005	-0.002	-0.016	-0.002	-0.007
t-test p-value	(0.591)	(0.867)	(0.211)	(0.480)	(0.560)
			parentsrepresentative		
Difference	0.025	0.023	-0.005	0.011	0.019
t-test p-value	(0.003**)	(0.009**)	(0.671)	(0.428)	(0.005**)
			parentsinassociation		
Difference	0.029	0.032	0.002	-0.001	0.006
t-test p-value	(0.003**)	(0.001**)	(0.886)	(0.957)	(0.161)
National Standardized test	✓	✓	✓	✓	✓
Specific Standardized test	✗	✓	✓	✓	✓
Parent's Monthly Revenue	✗	✗	✓	✓	✓
Other socioeconomic	✗	✗	✗	✓	✓
Parent's involvement	✗	✗	✗	✗	✓

Table 12: Balance check with genetic Matching. Population size of 10,000.

F Exam scores in the National Diploma

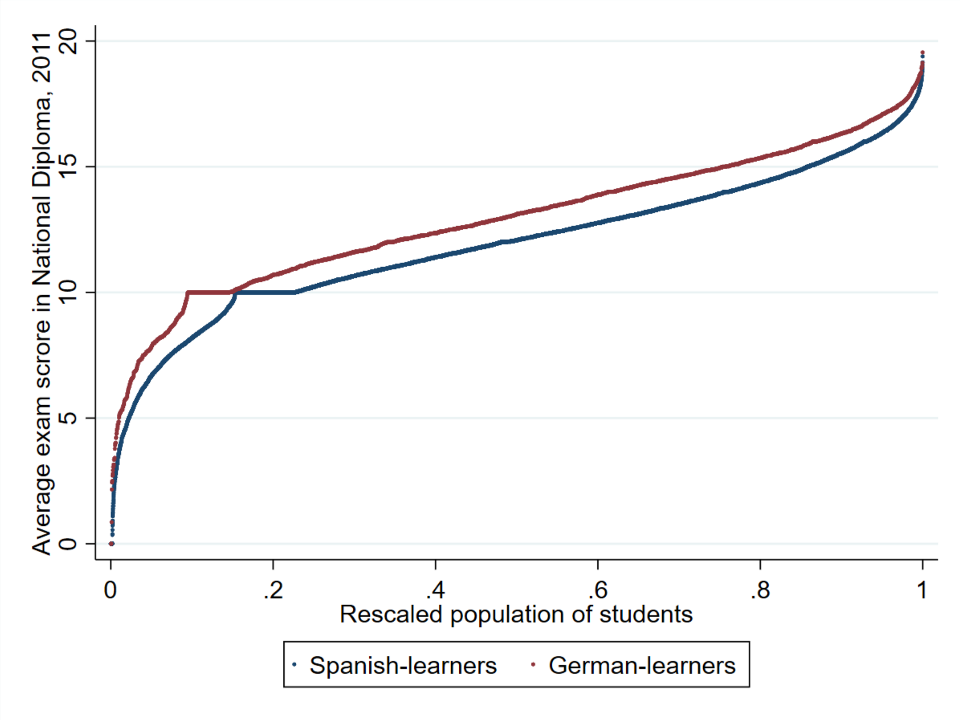

Figure 5: The graphs show the distribution of exam scores for the National Diploma. Populations of students are rescaled to be between zero and one, all scores are shown in ascending order, and each dot represents one student with Spanish-learners in blue colour and German-learners in red colour.

G Propensity score matching estimates

	Overall Average (Cont. and exam)				
German class	0.322**	0.271***	0.177†	0.306**	0.210†
	(0.002)	(< 0.001)	(0.081)	(0.007)	(0.062)
N	17950	17039	9047	7014	6980
	Overall Average (Cont.)				
German class	0.331**	0.265***	0.175	0.321**	0.235†
	(0.003)	(0.001)	(0.104)	(0.008)	(0.051)
N	17950	17039	9047	7014	6980
	Mathematics(Cont.)				
German class	0.298†	0.185	0.080	0.213	0.224
	(0.056)	(0.111)	(0.619)	(0.215)	(0.204)
N	17914	17007	9032	7007	6973
	Mathematics(Exam)				
German class	0.363*	0.324*	0.234	0.277	0.391*
	(0.032)	(0.013)	(0.183)	(0.150)	(0.042)
N	17645	16771	8921	6946	6913
	French(Cont.)				
German class	0.351**	0.317***	0.212†	0.290*	0.356**
	(0.003)	(< 0.001)	(0.079)	(0.031)	(0.007)
N	17915	17008	9032	7005	6971
	French(Exam)				
German class	0.260*	0.184†	0.248†	0.270*	0.128
	(0.042)	(0.050)	(0.058)	(0.050)	(0.350)
N	17675	16796	8929	6951	6918
	French(Dict.)				
German class	0.341†	0.276*	0.278	0.284	0.170
	(0.061)	(0.040)	(0.133)	(0.152)	(0.386)
N	17669	16793	8928	6950	6917
	French(Essay)				
German class	0.214	0.223*	0.186	0.279†	0.200
	(0.115)	(0.033)	(0.192)	(0.071)	(0.205)
N	17661	16786	8927	6950	6917
National Standardized test	✓	✓	✓	✓	✓
Specific Standardized test	✗	✓	✓	✓	✓
Parent's Monthly Revenue	✗	✗	✓	✓	✓
Other socioeconomic	✗	✗	✗	✓	✓
Parent's involvement	✗	✗	✗	✗	✓

Table 13: Difference in grade in Brevet des College exam in 2011 estimated with Propensity Score Matching. The propensity score is estimated using a logit model and matching is performed on a common support based on the odds ratio.

	History and Geography (Exam)				
German class	0.307*	0.331***	0.446***	0.272*	0.438**
	(0.011)	(< 0.001)	(< 0.001)	(0.045)	(0.002)
N	17261	16385	8683	6739	6706
	History and Geography (Cont.)				
German class	0.260*	0.184†	0.248†	0.270*	0.128
	(0.042)	(0.050)	(0.058)	(0.050)	(0.350)
N	17675	16796	8929	6951	6918
	Physics and Chemistry (Cont.)				
German class	0.429**	0.336***	0.347*	0.273†	0.404**
	(0.001)	(0.001)	(0.011)	(0.070)	(0.008)
N	17902	16996	9029	7004	6970
	Biology (Cont.)				
German class	0.339**	0.341***	0.263*	0.348*	0.376**
	(0.006)	(< 0.001)	(0.033)	(0.012)	(0.007)
N	17902	16996	9027	7001	6967
	Technology (Cont.)				
German class	0.373***	0.235**	0.221*	0.010	0.093
	(< 0.001)	(0.002)	(0.039)	(0.930)	(0.424)
N	17872	16965	9012	6992	6959
National Standardized test	✓	✓	✓	✓	✓
Specific Standardized test	✗	✓	✓	✓	✓
Parent's Monthly Revenue	✗	✗	✓	✓	✓
Other socioeconomic	✗	✗	✗	✓	✓
Parent's involvement	✗	✗	✗	✗	✓

Table 14: Difference in grade in Brevet des College exam in 2011 estimated with Propensity Score Matching. The propensity score is estimated using a logit model and matching is performed on a common support based on the odds ratio.

	Plastic Art (Cont.)				
German class	0.330**	0.217**	0.065	0.070	0.270*
	(0.003)	(0.008)	(0.555)	(0.556)	(0.029)
N	17866	16961	9012	6991	6957
	Music (Cont.)				
German class	0.430***	0.360***	0.351**	0.320**	0.269*
	(< 0.001)	(< 0.001)	(0.002)	(0.009)	(0.026)
N	17800	16901	8982	6968	6934
	Sport (Cont.)				
German class	0.008	−0.052	0.075	0.073	0.076
	(0.932)	(0.454)	(0.427)	(0.470)	(0.453)
N	17644	16759	8911	6918	6884
	Civic Education (Cont.)				
German class	0.131	0.317**	0.339*	0.314*	0.478**
	(0.317)	(0.002)	(0.012)	(0.032)	(0.002)
N	15401	14616	7763	6034	6006
	Conduct Mark (Cont.)				
German class	0.440***	0.380***	0.268*	0.366**	0.250*
	(< 0.001)	(< 0.001)	(0.022)	(0.002)	(0.047)
N	17917	17010	9035	7006	6972
National Standardized test	✓	✓	✓	✓	✓
Specific Standardized test	✗	✓	✓	✓	✓
Parent's Monthly Revenue	✗	✗	✓	✓	✓
Other socioeconomic	✗	✗	✗	✓	✓
Other socioeconomic	✗	✗	✗	✗	✓

Table 15: Difference in grade in Brevet des College exam in 2011 estimated with Propensity Score Matching. The propensity score is estimated using a logit model and matching is performed on a common support based on the odds ratio.

H Propensity score matching balancing check

Variable	Mean value		bias(%)	t-test	
	Treated	Control		t	$p > t$
French 2007	12.96	13.02	-1.80	-0.41	0.68
Mathematics 2007	14.58	14.61	-1.00	-0.24	0.81
Mathematics 2008	12.73	12.71	0.40	0.09	0.93
French: Sentences 2008	10.76	10.79	-0.60	-0.14	0.89
French: Understanding 2008	14.39	14.40	-0.30	-0.08	0.94
French: Lexicon 2008	14.92	14.83	2.30	0.56	0.58
French: Reasoning 2008	11.72	11.68	1.20	0.29	0.77
monthlyrevenue	3578.70	3629.90	-2.60	-0.55	0.58
nbrroom	5.68	5.68	-0.10	-0.03	0.98
ownroom	0.82	0.84	-3.20	-0.75	0.45
motheruniversity	0.42	0.42	-1.40	-0.31	0.76
fatheruniversity	0.36	0.36	0.80	0.18	0.86
borninfrance	0.97	0.97	3.60	0.79	0.43
motherborninfrance	0.91	0.91	0.00	0.00	1.00
fatherborninfrance	0.89	0.88	4.80	1.10	0.27
parentsrepresentative	0.14	0.14	-2.90	-0.63	0.53
parentsinassociation	0.18	0.20	-5.30	-1.16	0.25

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: Balance check with Propensity Score Matching based on observables from National test, Specific Standardized test, socioeconomic variables, and parent's involvement.

I Cost of Studying German

In this section, we perform a structural estimation of the average cost of studying the German language across different regions of France. We use the logit to model language class choices and assume, without loss of generality, that the cost of Spanish is 0, while the cost of German is $C > 0$. Simplified expressions for utilities in this case can be written as follows:

$$u_i(G) = a_i + p_i(g_G) - C \text{ and } u_i(S) = a_i + p_i(g_S).$$

We use these utility functions to estimate a structural model of choice. First, for each student, the ability a_i is computed as the average grade in two most important courses–mathematics and French language–in the 2007 National test (first two columns of Table 2). Having computed a_i for all i, we compute for each students i the peer effect he would received by joining German ($p_i(g_G)$) and Spanish ($p_i(g_S)$).

The available data does not provide information on exact schools that students attend. Instead, we utilise information on the department–geographical region–where schools are located. Hereby we aggregate information on students at the department level. There are 100 departments in our data, and we exclude four departments where we observe zero

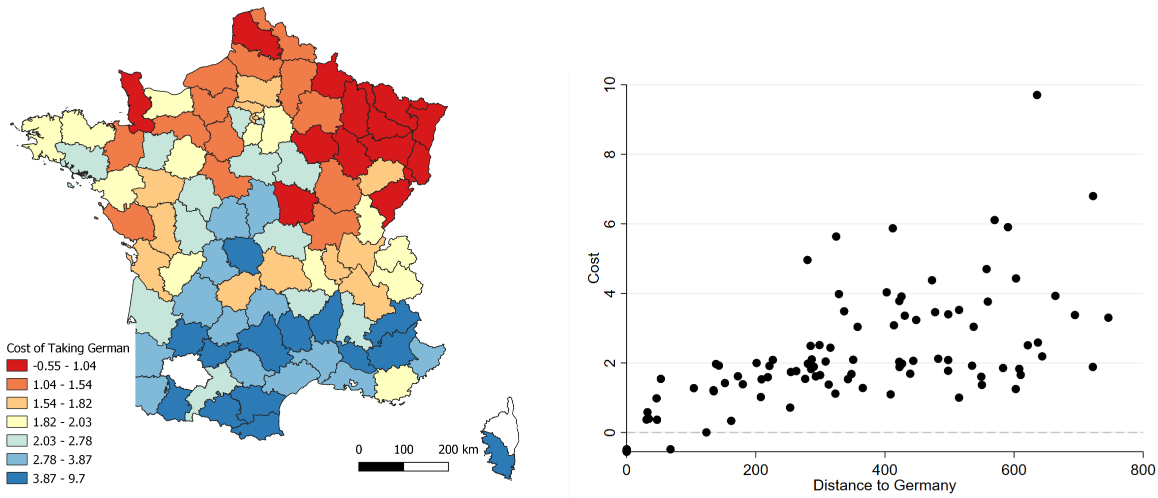

Figure 6: Left panel depicts the map where for each department different estimated costs of taking German are indicated with different colors from red (zero cost) to dark blue (large cost). Right panel depicts the scatter plot of all departments with the distance to the German border (x-axis) and cost estimates (y-axis).

students studying German.[22] Our baseline model makes deterministic predictions over second foreign language choices. Given the parameters of their utility function, students choose one language with certainty. However, in reality, there is likely to be some noise in their decision possibly due to all the factors not captured by the model. We capture this randomness we use the logit model, where the noise follows a Gumbel distribution.[23] We define the probability that student i choose German language $P(German_i)$ as follows:

$$P(German_i) = \frac{e^{u_i(G)}}{e^{u_i(G)} + e^{u_i(S)}}.$$

We estimate the only free parameter of the model C using maximum likelihood estimator. The corresponding log-likelihood function LL is defined as follows:

$$LL(German_i|C) = \sum_{i=1}^{N} ln[P(German_i) * 1_{[German_i=1]} + (1 - P(German_i)) * 1_{[German_i=0]}],$$

where $German_i = 1$ is a dummy variable equal to 1 if the student took German.

For example, if one assumes a constant cost across departments, then the estimated cost C is 5.67. It is statistically significant and positive, which is in line with the model.[24]

Next, we ease a restriction on the homogeneity of the cost and allow the relative cost of studying German to vary by department. The left panel in Figure 6 shows a map of the point estimates. The right panel in Figure 6 shows the results depending on the department's distance to the Spanish and German border. Note that these estimates are not very informative due to large confidence intervals. Standard errors, in this case, are

[22]We exclude Haute-Corse, Gers, Guadeloupe, and Mayotte as we cannot compute the peer effect of studying German in those departments.

[23]Please refer to Train (2009) for a textbook treatment of the model.

[24]The standard error is 0.43 and the number of observations N=17,775.

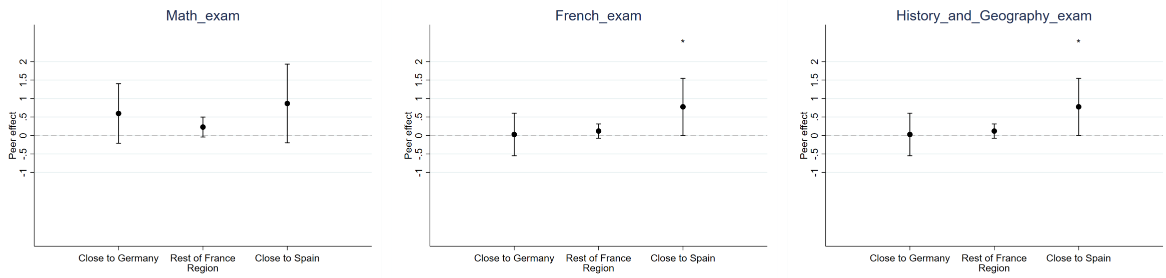

Figure 7: Peer effects for Mathematics, French Language, and History and Geography exams estimated separately for "Close to Germany", "Rest of France", and "Close to Spain" regions.

large because of a very limited number of observations per department. Nevertheless, one can observe a tendency for the cost to be smaller in those departments that are closer to the border with Germany. This result is in line with the intuition of the cost we provide to motivate our modeling choices. For example, those students who live closer to Germany have a higher probability of having German-speaking family members, which reduces the cost of learning. Such students are also more likely to have plans to move to Germany in the future that makes learning German more attractive and therefore relatively less costly. The same logic applies to the costs of learning Spanish and proximity to the border with Spain.

J More results on regional analysis

	socioeconomic status								Parent's involvement	
	Income	N room	Own room	M Uni	F Uni	Born in F	M Born in F	F Born in F	Representative	Association
German class	697.346*	0.062	0.014	0.165***	0.204***	0.012	−0.077*	−0.045	0.045	0.078*
	(0.015)	(0.635)	(0.668)	(< 0.001)	(< 0.001)	(0.385)	(0.012)	(0.173)	(0.103)	(0.013)
Close Germany	−98.273	0.111	−0.040†	−0.082**	−0.048†	−0.017†	−0.031	−0.012	−0.011	−0.027
	(0.607)	(0.201)	(0.054)	(0.004)	(0.096)	(0.053)	(0.125)	(0.575)	(0.536)	(0.185)
German class×Close Germany	−579.612	0.017	−0.036	−0.120*	−0.168**	−0.012	0.067†	0.012	−0.033	−0.067†
	(0.109)	(0.921)	(0.363)	(0.024)	(0.002)	(0.502)	(0.083)	(0.778)	(0.349)	(0.087)
Constant	3043.103***	5.301***	0.861***	0.350***	0.309***	0.980***	0.872***	0.868***	0.109***	0.150***
	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
German class+	117.734	0.079	−0.023	0.045	0.036	0.000	−0.010	−0.033	0.012	0.010
German class×Close Germany	(0.594)	(0.434)	(0.348)	(0.176)	(0.286)	(0.976)	(0.664)	(0.188)	(0.559)	(0.660)
	2095	3579	3648	3440	3132	3510	3533	3194	3635	3642

Table 17: Difference in socioeconomic status between students who choose German and Spanish and regions (close to Germany or close to Spain). Variable German is a binary variable that is equal to one if a student takes a German class and zero–if Spanish. Variable Close (to Germany) is a binary variable that is equal to one if a student lives near Germany and zero if a students lives near Spain. Estimates and confidence intervals are obtained using OLS.

	Overall		Mathematics		French Language		Physics &	Sports
	Exam & Cont.	Cont.	Exam	Cont.	Exam	Cont.	Chemistry	
GM-estimated	0.156*	0.134*	0.353**	0.076	0.147	0.239**	0.329***	−0.001
	(0.068)	(0.098)	(0.011)	(0.556)	(0.149)	(0.011)	(0.007)	(0.990)
N	2440	2440	2436	2416	2436	2416	2436	2416
PSM-estimated	0.210*	0.235*	0.351**	0.224	0.128	0.356***	0.404***	0.076
	(0.062)	(0.051)	(0.041)	(0.204)	(0.350)	(0.007)	(0.008)	(0.453)
N	6980	6980	6913	6973	6918	6971	6970	6884

Probability p in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 18: Differences in grades for courses included in the National Diploma between matched students.

K Results of Genetic Matching

Genetic matching is shown to generate accurate estimates of the treatment effect in non-experimental settings such as, for example, Lalonde data.[25] The ability of the genetic matching to analyze non-experimental data is especially important because in the environment under study RCTs are hardly implementable.

We perform genetic matching on all observables listed in Tables 2 and 3. The algorithm found a matrix of weights that ensured the balance of observables across marks, socioeconomic background, and parents' involvement. Appendix D contains results of matching models that use only subsets of observables. Hereafter, we focus on the results that are obtained based on matching with all observables, as in this case, the balance in observables is achieved, see Appendix E for the results of various balance checks. Due to missing observations, we start matching with approximately $6,000$ observations and, on average, obtain the matched sample of size $2,450$.

L Results of Testing the Signaling

To access the informativeness of taking German as a signal about academic performance and ability, we predict results of five exams included in National Diploma using observed covariates. We include in a linear regression continuous assessment marks and National test results from 2007. Table 19 reports the results, that suggest that taking German does not provide valuable information that can be used to predict the ability of the student measured with all five exams in the National Diploma.

[25]This data set combines data from a randomized job training experiment–the National Supported Work Demonstration Program (NSW)–and observational survey data. Ability of matching estimators to accurately predict treatment effects using observables from the survey was initially analyzed in LaLonde (1986), and later in Dehejia and Wahba (1999), Smith and Todd (2001), Dehejia and Wahba (2002), Dehejia (2005), and Smith and Todd (2005).

Exam	Covariate	Coef.	Std. Err.	t	$P > t$	[95% Interval]	
History and Geography							
	German class	-0.02	0.04	-0.35	0.72	-0.10	0.07
French essay							
	German class	0.04	0.07	0.62	0.54	-0.09	0.17
French dictation							
	German class	-0.06	0.06	-0.92	0.36	-0.18	0.06
French exam							
	German class	-0.02	0.04	-0.35	0.72	-0.10	0.07
Math Exam							
	German class	0.09	0.06	1.57	0.12	-0.02	0.20

Table 19: Effect of taking German on exams in National Diploma when continuous assessment marks and National test results from 2007 are included in OLS regression.

M Comparison with a Signaling Model

This section provides a signaling model that is based on the assumptions use in costly peer-seeking model form the main text. We keep the same assumptions on the model primitives such as $F(a)$ and $c_l(a_i)$, but the players' utility functions takes a different form and in this case do not include peer effects. We also introduce an employer who wants to hire the players and can use language class choices as informative signals.

There are n players. There is commonly known distribution, $F(a)$, over the set of possible abilities, $A = [\underline{a}, \bar{a}]$. Each player observes her own ability, a_i, and decide to choose G or S. Each action costs the player differently as in the main text. That is, $\Delta c(a_i) := c_G(a_i) - c_S(a_i) > 0$ and $\partial \Delta c(a_i)/\partial a_i < 0$ for all $a_i \in A$.

Suppose that there is an employer who wants to hire an employee. The employer wants to maximize the profit that he can make out of his hiring decision. Assume that the employer's profit is the difference between an employee's ability and the wage paid to the employee:

$$u^E(a_i, w) = a_i - w.$$

Assuming that the potential employees are the players above, the game proceeds as follows. First, the players simultaneously decide their actions. The employer cannot observe a potential employee's ability. Instead, the employer can observe the potential employee's action and decide the wage. Thus, the employer's strategy is

$$s_E : \{G, S\} \to \mathbb{R}_+.$$

For simplicity, we assume that the employer does not have budget constraint and can hire all players. Given the employer's strategy, each player's payoff is

$$u^P(l) = w_l - c_l(a_i) \quad \text{for} \quad l \in \{G, S\},$$

and each player's strategy is $s_i : A \to \{G, S\}$. Now note that since each player's payoff only depends on the wage offer by the employer. Hence, we have n independent games, where each game is played between the employer and each player i.

Assume that the labor market is perfectly competitive, and, thus, the employer only can make the zero profit. Then, given a player's strategy, s_i, the employer observes either G or S. Then, the employer updates his beliefs about the ability of the player and makes a wage offer that makes his profit zero at an equilibrium: i.e., $w_l = E[a_i | l, s_i]$ for $l \in \{G, S\}$.

Now suppose that each player play a threshold strategy as in the main text:

$$\begin{aligned} \hat{s}_i &= G \text{ if } a_i \geq \hat{a}, \\ &= S \text{ if } a_i < \hat{a}. \end{aligned}$$

Given $\hat{s}_i$, we have

$$\begin{aligned} E[a_i | G, \hat{s}_i] &= E[a_i | a_i \geq \hat{a}], \quad \text{and} \\ E[a_i | S, \hat{s}_i] &= E[a_i | a_i < \hat{a}]. \end{aligned}$$

Thus, each player's payoff given $\hat{a}_i$ becomes

$$\begin{aligned} u^P(l; a_i) &= E[a_i|a_i \geq \hat{a}] - c_G(a_i) \;\text{ for }\; l = G, \\ &= E[a_i|a_i < \hat{a}] - c_S(a_i) \;\text{ for }\; l = S, \end{aligned}$$

which is equivalent to the player's payoff in the model from the main text, where insted of employer there are peer effects. In other words, we have

$$\begin{aligned} \Delta u_i(a_i) &= \Delta P_i^e(\hat{a}) - \Delta c(\hat{a}) = E[a_i|a_i \geq \hat{a}] - E[a_i|a_i < \hat{a}] - (c_G(a_i) - c_S(a_i)) \\ &= w_G - w_S - (c_G(a_i) - c_S(a_i)) = \Delta w - \Delta c(a_i) = \Delta u^P(a_i). \end{aligned}$$

Then, Lemma 1 holds after we replace $\Delta P_i^e(\hat{a})$ with Δw. Furthermore, Proposition 1 provides the sufficient conditions for an equilibrium with s_i^* to exist, where $s_i^* = G$ for $a_i \geq a^*$ and $s_i^* = S$ for $a_i < a^*$.

Note that now we have n Perfect Bayesian Nash Equilibria for n independent games, where each player employs the threshold strategy with $\hat{a} = a^*$. Then, the aggregated outcome of these n signaling games is equivalent to the equilibrium we have in the language choice game with peer-seeking motive.