Endogenous Tracking: Sorting and Peer Effects*

Preliminary and incomplete–please do not cite or circulate
This version: July 7, 2022

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

[a] *Center for Behavioral Institutional Design (C-BID), New York University Abu Dhabi*
[b] *Department of Economics, Deakin University*
[c] *Division of Social Science, New York University Abu Dhabi*

Abstract

French junior high schools do not use ability tracking and traditionally group students into classes based on their choices of which second foreign language to learn. It is believed that better-performing students would endogenously sort themselves into classes that learn German as a second foreign language instead of Spanish. A choice of German over Spanish incurs higher effort costs, but the costly signaling motive has a limited presence as abilities are regularly measured with tests, including exit examination. We propose an alternative model of human capital accumulation with costly sorting, where students have a peers-seeking motive expecting positive peer effects. In equilibrium, academically better-performing students endogenously track themselves into German-learning classes by bearing an extra cost of choosing German over Spanish. We perform an observational study of a large panel collected by the French Ministry of Education that contains information on academic performance, class choices, and socioeconomic background of students. We show that, in line with our theoretical predictions, students who choose German, on average, perform better by 10% in Mathematics and the French Language exams. We use regional differences to show that, in line with the comparative statics, sorting is stronger in regions closer to Spain, where the cost of taking German is higher. Using propensity score matching, we find positive peer effects on academic performance that drive endogenous tracking into German-learning classes. Notably, we show that apart from inequality in educational experience induced by unequally distributed peers and their effects, identified endogenous tracking also leads to an inequality in socioeconomic status between the formed groups.

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

*We thank Celeste Carruthers, Hervé Crès, Raquel Fernandez, Ben Golub, Maia Guell, John Ham, Jean Imbs, Francis Kramarz, Ernesto Reuben, Sevi Rodriguez Mora, Georg Schaur, John Singleton, Tayfun Sönmez, William Thomson, Utku Ünver, Etienne Wasmer, John Wooders, Myrna Wooders, Ulf Zölitz and participants of the SOLE 2022, the PET 2022, 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

Allocating students into classes within a school based on their academic performance, also known as "tracking", is a popular yet disputable approach in educational systems worldwide.[1] Students may benefit from tracking because of positive peer effects from studying with peers of similar academic performance, while tracking may lead to inequality in educational outcomes between groups.[2] Tracking may take various forms, and in the present paper, we analyze a form of tracking, which we refer to as an *endogenous* tracking. Unlike in standard tracking, where sorting is exogenously imposed by the system via, for example, the minimum required exam scores, in endogenous tracking, students sort themselves into groups that differ in ability compositions.[3] Unlike in the signaling model, where sorting is a byproduct of students' efforts to signal their ability, in our model, sorting is the main motive for students who seek positive peer effects. The relative importance of peer-seeking and signaling motives in observed sorting lies beyond the scope of this paper.[4] To demonstrate the validity of the theoretical model, we test the predictions using the data on students from French high schools and show: i) the sorting of students into groups with different abilities, ii) positive peer effects on academic performance within sorted groups, and iii) socioeconomic inequality between the sorted groups.

Most students in France enter junior high school, also known as middle school when they are 11 years old and receive a diploma after four years of secondary education. When students enter junior high school, they choose the first foreign language to study, and by far, the majority choose English. The same or the next year, students choose a second foreign language, and the majority choose between German and Spanish. This paper is motivated by anecdotal evidence in the form of an "open secret"–excellent students prefer German over Spanish as a second foreign language.[5] German language is a costly choice–it is harder to learn for French-speaking students while the returns are limited. Such choice can not be well explained by the signaling motive because the abilities of students are closely monitored with various exams, including National exams.[6] We provide a model

[1]Slavin (1990) provides the discussion of US case and Pischke and Manning (2006) discuss England and Wales.

[2]Brunello and Checchi (2007) report ambiguous results on effects of tracking on equality of opportunities. Hanushek and Wößmann (2006) find early tracking to increase inequality and possibly decrease academic achievements. Adverse effects of academic tracking were found, e.g., in Matthewes (2021).

[3]Self-selection in education have been widely studied and oftentimes concerns the schooling choices on the basis of financing constraints and expected schooling returns, see Willis and Rosen (1979) and Keane and Wolpin (1997) among many others.

[4]In Appendix M we establish an observational equivalence between costly peer-seeking model and signaling model outcomes if peer effects in a standard fashion are modeled to be linear in means. In this way, we theoretically show that previously observed sorting effects in education might be consistent even with pure peer-seeking motives without the signaling component.

[5]Herbaut et al. (2019) provides an excellent overview on tracking in French high schools and mentions the foreign language class choice problem we study in this paper with the observation of similar endogenous tracking outcomes: "Most notably, the choice of German as a first foreign language, of 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 optional choice of Latin is very different as it does not affect class composition but rather allows to get an extra points in diploma.

[6]Our data is not suitable to completely rule out signaling as a motive to choose German over Spanish, but we provide some evidence that the second foreign language choice does not "signal". In Appendix L, we show that continuous assessment marks in Graduate Diploma and National test results are sufficient for predicting all exam scores in National Diploma, and the choice of a second foreign language is not

of skill acquisition, that may serve as an input into the Roy model of wage formation, with the peer-seeking motive that predicts observed endogenous tracking.[7] The model incorporates a critically important feature of the class formation process at the junior high school level–until the reforms of 2016, class composition for *all* school subjects was often based on the second language choice. In the model, better students are willing to bear the cost of learning German to enjoy larger positive peer effects in learning all other subjects included in the National exit exams. Note that since all classes share the curriculum, pedagogical methods, and even teachers most of the time, peers and their effects on academic progress are the only driving force of sorting within schools.[8] We demonstrate the model's validity with the observational study of several cohorts of French students–both sorting and positive peer effects are identified. Our empirical analysis, therefore, supports the motivating anecdotal evidence–better-performing students do prefer German over Spanish. While the tendency was observed for many years, this paper appears to be the first attempt to investigate the issue both theoretically and empirically.

Ability grouping is a popular form of horizontal stratification of students and is often used, for example, when instructing mathematics. According to the analysis presented in the OECD (2010) report on schools' practices and policies, approximately a third of the variation in mathematics performance can be explained by differences in the degree of horizontal stratification within the educational system. While there exists a debate in the literature regarding the overall effects of tracking on educational outcomes at the aggregate level, it is widely accepted that tracking leads to inequality in the educational process of students. For that reason, some educational systems, including the French system, avoid tracking at the junior high school level partly because of social equality concerns. We show that even in the systems with egalitarian preferences, tracking may endogenously arise within schools, for example, due to class formation rules, including one based on a second foreign language choice.[9] We identify socioeconomic inequality between endogenously formed groups, and that may be of concern for French educational system designers.

With our theoretical model of endogenous tracking, we contribute to the literature on sorting where papers concern, inter alia, tournament participation in Morgan et al. (2018) and Azmat and Möller (2018), club participation in Windsteiger (2021), and residential choices in Tiebout (1956) and Rothstein (2006) among many other.[10] In sorting models, there are externalities generated by a group and some form of a cost

informative.

[7]Skills were found to be the main source of wage inequality accounting for approximately three-quarters of variation, see Taber and Vejlin (2020).

[8]Existing evidence on the effects of the class size on academic performance is mixed. While Hoxby (2000b) and Urquiola (2006) report positive effects of smaller classes, Angrist et al. (2017) and Angrist et al. (2019) find little to no effects. In French school, 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.

[9]In this way, we contribute to the literature on sorting where the focus is usually made on sorting between schools rather than within schools. The role of school qualities on academic performance and housing locations are widely studied, see, for example, Hoxby (2000a) and Abdulkadiroğlu et al. (2014). In MacLeod and Urquiola (2009) the authors show that if schools are allowed to select students, then the competition leads to stratification by background characteristics. In our paper, we demonstrate theoretically and empirically that stratification may happen endogenously.

[10]An interesting example of experimental work on club formation 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.

for joining groups that jointly generate sorting patterns. Sorting in the form of the competition between firms in attracting talents is studied in Damiano et al. (2012). In this paper, we provide a simple theoretical model that generates endogenous tracking in education at the class level using peer-seeking motive instead of signaling–to our best knowledge is the first model of such kind. At the same time, we believe that the intuition of this model on endogenous club formation goes beyond the educational framework and may provide additional explanations for various observed stylized facts about self-sorting in a situation where peer composition matters. For example, one may use the proposed model to explain why many talented specialists join startups instead of better-paying jobs at established companies–they are willing to sacrifice a fraction of their paycheck to enjoy peer effects from those who are also willing.[11] Alternatively, consider restrictively high fees to enter gulf clubs–one may rationalize them using our model, where wealthy players prefer to play with wealthy peers, and high fees help them to isolate themselves from other not-so-wealthy golf players.

To derive our theoretical model of endogenous sorting, we make two main assumptions. First, we assume that there is a sunk cost of choosing German over Spanish. This assumption is reasonable because German imposes higher learning costs in terms of time and effort since French and Spanish belong to the Romance family of languages, while German–to the Germanic family of languages.[12] Second, we assume that peer composition of formed classes affects the academic progress of students. Peer effects in education and tracking are widely studied in the literature, and we only mention a few related papers. Card and Giuliano (2016) find positive peer effects on the performance of students representing minorities from having high achieving peers in large urban district schools. Booij et al. (2017) find peer effects analyzing results of the randomized experiment where students studying economics were tracked into groups by ability. Garlick (2018) studies the effects of tracking students into dormitories and find peer effects from living side by side on academic performance.[13] Note that tracking in these studies is exogenous, and the closest paper on peer effects that acknowledges the importance of endogenous mechanisms is Carrell et al. (2013), where the authors show that tracking may harm some students as they avoid interacting with their exogenously assigned peers.[14] Our empirical analysis demonstrates the existence of peer effects in groups where student self-select in line with the recent evidence from a framed field experiment reported in Kiessling et al. (2021).

To empirically investigate the effects induced by endogenous tracking, we perform an

[11]Sorenson et al. (2021) analyze Danish data and show that startup employees earn 17% less than corporate employees but attribute that to the smaller human capital of these employees. In a somewhat related setup, Craig and Vierø (2013) study sorting effects induced by choices between academia and the private sector.

[12]An ideal illustration would be to have data on returns on learning German and Spanish in France, but unfortunately, such analysis can hardly be preformed due to endogeneity issues, including self-sorting that is studied in this paper. Instead, some fragmented evidence from a few papers exists. Using data from European Community Household Panel Survey (1994-2001) Williams (2006) shows that Spanish has a larger positive effect on income than German. Adsera and Chiswick (2007) find that the earnings of immigrants in France are 20% higher if they speak Romance languages.

[13]Peer effects were identified in many settings apart from classrooms. Falk and Ichino (2006) find peer effects in the clean, controlled in-field laboratory setting. Mas and Moretti (2009) identify peer effects at the workplace of a large supermarket chain.

[14]This observation and also evidence from Ly and Riegert (2014) may explain why some studies such as, for example, Duflo et al. (2011), Bui et al. (2014) and Tangvatcharapong (2020), do not identify peer effects in certain cases of exogenous formation of peer groups.

observational study of the data on French students. In particular, we use the information on more than 22,500 students from the representative panel data provided by the French Ministry of Education. This panel data set follows students from 2007, when students entered junior high school, until 2014, when they graduated.[15] This cohort made a second foreign language choice in the 2009-10 academic year, with 84% of students choosing Spanish and 16%–German. The panel provides information on the academic progress of students with test results before and after the second foreign language class choice and also reports students' socioeconomic status with family survey responses. A simple comparison of Spanish-taking and German-taking students' academic performances in 2007 shows predicted sorting effects–students who choose German over Spanish have significantly better test scores by approximately 10%. We use various background characteristics, including income and parents' level of education, to show that students choosing German also have higher socioeconomic status. We also use regional variation and exploit the fact the closer to Germany, the relative cost of choosing German goes down, while the closer to Spain–up. We find significantly stronger sorting effects in regions close to Spain in line with comparative statics predictions. To validate one of the assumptions of the model and evaluate inequalities in the educational experience, we check if peer effects differ in German-learning and Spanish-learning classes. To identify possible peer effects in groups, we match similar students choosing different classes and apply a propensity score matching from Rosenbaum and Rubin (1983).[16] Using students' observable characteristics before they choose which second foreign language to study, we match students based on both academic performance and various background characteristics. Propensity score matching shows a positive effect of German-taking class composition on performance for many school subjects that are included in the National Diploma assessment. While there are no differences in observable characteristics between matched groups, there may be differences in unobservable characteristics, and, therefore, the results should be taken with caution.

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 there is a relative cost of joining a German-learning group.[17] Based on the theoretical model, we derive two conjectures that we later empirically validate with the data: German-learners and Spanish-learners differ in terms of ability composition

[15] In this paper, we focus on the latest panel as it is the most detailed available panel. The language choice defined peer composition for students in this cohort; the rule was changed with the reforms of French secondary education in 2016.

[16] As a robustness check we also do matching directly on observables using genetic matching estimator from Diamond and Sekhon (2013) and confirm the result obtained using propensity score matching. The analysis is available in Appendix K.

[17] 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 that 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.

(sorting effects) and education experience (peer effects).

2.1 Model

Consider n-player simultaneous-move incomplete information game. There is a commonly known distribution, $F(a)$, over the set of possible abilities, $A = [\underline{a}, \overline{a}]$.[18] Each player observes her own ability, a_i, drawn from $F(a)$. Then, each player chooses one of two languages, G or S. Hence, each player's strategy is a mapping from $[\underline{a}, \overline{a}]$ to $\{G, S\}$.[19]

Denote by g_l the group of players who choose language l, where $l = G, S$. Each member of a group bears the cost of learning the chosen language but also enjoys benefits from her peers' abilities, i.e., the peer effect. We assume that both the learning cost of each language and the peer effect depend on players' abilities. We denote by $c_l(a_i)$ the learning cost of language l for $l \in \{G, S\}$. In particular, we denote by $\Delta c(a_i)$ the difference between $c_G(a_i)$ and $c_S(a_i)$, $c_G(a_i) - c_S(a_i)$. We denote by $P_i^e(g_l)$ the peer effect that player i in group g_l experiences.

Assumption 1. *(Learning Costs and Peer Effects)*

(a) $c_l(a_i)$ *is continuous for* $l \in \{G, S\}$,

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

(c) $P_i^e(g_l) = \left(\sum_{j \in g_l \setminus \{i\}} a_j\right) / |g_l \setminus \{i\}|$, *where* $|\cdot|$ *is the cardinality of a set.*

Assumption 1 imposes restrictions on the learning costs and peer effects. Assumption 1.(a) is a technical assumption. Assumption 1.(b) implies that (i) it is more demanding to learn G than S given any ability of a player, and (ii) the difference between two learning costs decreases in a_i, i.e., a player with a higher ability experiences less increase in the learning cost when changing her choice from S to G. Especially, the decreasing differences allow us to derive an equilibrium in the spirit of market signaling proposed in Spence (1973).[20] Willingness to bear a positive cost $c_G - c_S$ of choosing German over Spanish allows high types to isolate themselves from low types in a separating equilibrium. Assumption 1.(c) says that the player i's peer effect in the group g_l is the average of the realized abilities of other players in that group, excluding her own ability.

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 the peer effect she experiences in g_l *but* decreases in the cost of choosing a language.[21] We assume that the educational benefit is linear in these two factors. Given realized value $a_i \in A$ and a strategy profile, $\mathbf{s} = (s_1, \cdots, s_n)$, if a player chooses G, then her payoff is

$$u_i(G|a_i, s_{-i}) = E[P_i^e(g_G)|a_i, s_{-i}] - c_G(a_i),$$

[18] An assumption of the closed interval is only meant to match the commonly used way of measuring academic abilities–marks that vary within explicitly defined bounds.

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

[20] While there is definitely a component of signaling in our model, we also introduce a coordination component, and the combination of these two components generates endogenous sorting effects.

[21] Here we do not model preferences for within-group ranking that may affect decisions to join more competitive or less competitive groups. See Villeval (2020) for a recent overview on existing evidence on effects of providing information on relative ranking.

on the other hand, if a player chooses S, then her payoff is

$$u_i(S|a_i, s_{-i}) = E[P_i^e(g_S)|a_i, s_{-i}] - c_S(a_i).$$

We are interested in a symmetric equilibrium in which every player employs the same "threshold strategy." That is, for every player i, $s_i(a_i)$ takes the following form:

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

Given that everyone follows $\hat{s}_j(a_j)$ and a realization of player i's ability a_i, consider an event in which there are k number of players whose ability realizations are greater than $\hat{a}$ and $n-k$ number of players whose ability realizations are 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 the "success" probability of $(1-F(\hat{a}))$. Given that event $\mathcal{E}_k$ happened, the expected peer effect for player i for joining g_G is

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

Then, the expected peer effect is simply

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

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

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

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

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

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 follows $\hat{s}_j$, $\Delta P_i^e(\hat{a})$ is identical across all players. More importantly, it is independent of a_i, a realized ability of player i. Lemma 1 below discuss the properties of $\Delta P_i^e(\hat{a})$.

Lemma 1. *Given $F(a)$ with $a \in [\underline{a}, \bar{a}]$, $\Delta P_i^e(\hat{a})$ is continuous at any $\hat{a} \in (\underline{a}, \bar{a})$. In addition,*

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

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

The proof of Lemma 1 is relegated to the Appendix. 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) $\Delta c(\underline{a}) > a_m - \underline{a}$ and (ii) $\Delta c(\bar{a}) < \bar{a} - a_m$, there exists a separating 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, $\hat{s}_i(a_i|\hat{a})$, 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

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

and, by choosing S, this type would get

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

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

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

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

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

Since (i) $\Delta c(\underline{a}) > a_m - \underline{a}$ and (ii) $\Delta c(\bar{a}) < \bar{a} - a_m$ by assumption, $\lim_{\hat{a} \to \underline{a}} \Delta u_i(\hat{a}) < 0$ and $\lim_{\hat{a} \to \bar{a}} \Delta u_i(\hat{a}) > 0$. As $\Delta u_i(\hat{a})$ is continuous in $(\underline{a}, \bar{a})$, there must exist $a^* \in (\underline{a}, \bar{a})$ such that $\Delta u_i(\hat{a} = a^*) = 0$. Let every player follow $\hat{s}_i(a_i|\hat{a} = a^*)$. Every type of each player has the identical $\Delta P_i^e(a^*)$, and her payoff difference is

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

For $a_i = a^*$, $\Delta u_i(a_i) = \Delta P_i^e(a^*) - c(a^*) = 0$ by the definition of a^*. For any $a_i < a^*$, $\Delta u_i(a_i) = \Delta c(a^*) - \Delta c(a_i) < 0$, and, for any $a_i > a^*$, $\Delta u_i(a_i) = \Delta c(a^*) - \Delta c(a_i) > 0$ by Assumption 1.(b). Thus, $s^* = (\hat{s}_1(a_1|a^*), \cdots, \hat{s}_n(a_n|a^*))$ is an equilibrium. □

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

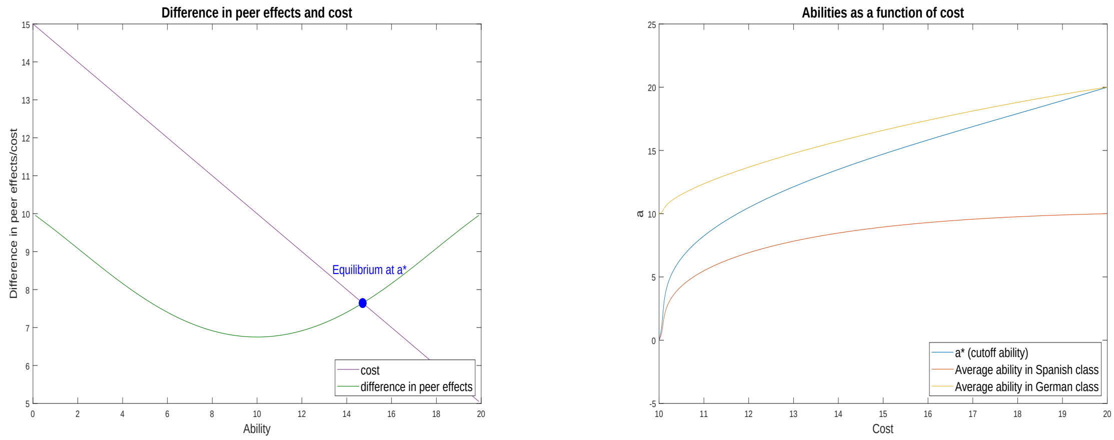

Figure 1: Left panel: an illustration of the equilibrium following the example. Right panel: an illustration of varying cost following the example.

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 when $C = 15$. 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. Right panel of Figure 1 demonstrates how equilibrium ability cutoffs a^* varies with the cost. One can see that with the increase in the cost, average abilities in *both* group increase.

2.2 Conjectures

We use the theoretical model of the language class choice to formulate three conjectures that we later test against the data.

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

Conjecture 1 naturally follows from Proposition 1 if one assumes the relative cost of languages supports sorting in equilibrium for a given ability curve. We use a somewhat weaker implication of the model and focus on using average values of abilities in the groups to account for 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 average ability of both German learners and Spanish learners.*

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

Conjecture 3. *On average, students who choose German enjoy 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 validation of the model's assumptions. The difference in peer effects between the groups is also one of the reasons why tracking, even endogenous, may be considered an unfavorable educational practice.

3 Empirical Strategy

In this section, we provide a brief overview of relevant features of the French educational system at the junior high school level, describe the panel collected by the French Ministry of Education, and outline which variables are used in the analysis. We then explain our empirical strategy to test three theoretical conjectures introduced in Section 2.

3.1 Data

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

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

The French Ministry of Education collects information on students' educational experience approximately every decade. We use the latest available panel data of $35,000$ French students following students through their schooling DEPP (2018).[22] This rich panel data has 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 presented in the panel.[23] 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 junior high school (usually 11 years old) and stop in 2014 when they finish high school and go to university or start working. Table 1 shows the data from the panel which 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

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

[23]Students of junior high school 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 this data. For schools in the *Ambition Réussite* program, the sample rate is 1/8, and it is 1/23 for the other schools.

year, students took standardized tests in French and Mathematics.[24] 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.[25] At the end of the year 2012, students received the National Diploma (*"Brevet des collèges"*), which collects marks for all subjects, including graduation exams taken by the end of junior high school. We use the standardized tests and the exam results to measure students' ability before and after studying in the groups formed based on their second language choices.

At the beginning of junior high 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 that follow the usual path make a choice of a second foreign language in 2010-2011.[26] We also exclude those observations that are not suitable for the study because marks for the National Diploma are missing.[27] Finally, we exclude those students who chose neither German nor Spanish as their second foreign language class.[28] We are left with $26,508$ students, which represents 75.8% of the original sample. Out of these $22,538$ students $3,696$ (16%) are studying German and $22,812$ (84%) are studying Spanish. Descriptive statistics for all variables used are presented in Appendix C.

3.2 Method

Conjecture 1 concerns the sorting effects stating that better-performing students are expected to prefer German over Spanish. To test such conjecture in Section 4.1 we use exam scores as a proxy for academic abilities. In this case, the analysis is straightforward–we just compare average values for various variables of interest. 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. 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–close to Germany.

Conjecture 3 relates to peer effects, and it states that peer effects are larger for those students who choose German. Ideally, to estimate treatment effect, one would need to have an exogenous variation either in the form of a Controlled Randomized Trial (CRT) or some kind of a (quasi-)natural experiment. Unfortunately, data on neither policy intervention in the form of CRT nor nature-induced variation is available.

Another 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.[29] However, based on the significant differences

[24] Those tests were administered to all french students starting junior high school this year. We denote them *National Standardized test.*

[25] Those tests were administered only to students participating in this study. We denote them *Specific Standardized test.*

[26] From 2016, students have to make this choice one academic year later in their schooling.

[27] There are many possible reasons for missing observations: students may go to pre-vocational junior high school, the exam data is missing, students retook a year, etc.

[28] In this case, we exclude 3,970 students choosing Italian, Portuguese, Arabic, etc.

[29] DID approach in the empirical analysis of educational data was previously used, for example, in Meghir and Palme (2005) to study peer effects, in Bonhomme and Sauder (2011) to study effects of selective education.

between sorted students established in Section 4.1, one can suggest 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 can not construct some synthetic analog as proposed, for example, as in Arkhangelsky et al. (2019).

Instead, to disentangle peer effects from sorting effects in Section 4.3 we use matching on the propensity score and complement the analysis with genetic matching directly on observables reported in the appendix.[30] Using students' observable characteristics before choosing which second foreign language to study, we match students who studied German to similar students who studied Spanish. The balance is achieved if the propensity score matching is used. We implement the nearest matching estimator with a replacement that gives us pairs of students where 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 with the matching with replacements have a desirable property of being invariant to the order of matching. As using the bootstrap for calculation of the standard errors is invalid, as shown in Abadie and Imbens (2008), we use consistent estimator from Abadie and Imbens (2006).

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). Validity of the common support assumption is relatively easy to show, for example, by demonstrating that the method generates a sufficient number of matched pairs. Indeed, when we perform the matching, we impose the common support restriction. 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 as we collect many relevant observables that should decrease the likelihood of missing the variable that may be responsible for the estimated treatment effect. 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, 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 defines whether they prefer to be in a more competitive environment of German-learners or a less pressing environment of Spanish-learners.

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, peer effects, and socioeconomic and educational inequality that relates to both effects. We show that sorting effects are present in the

[30] 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. Genetic matching estimator is particularly popular instrument in political science, see, for example, Hopkins (2010) and Frymer and Grumbach (2021) for applications.

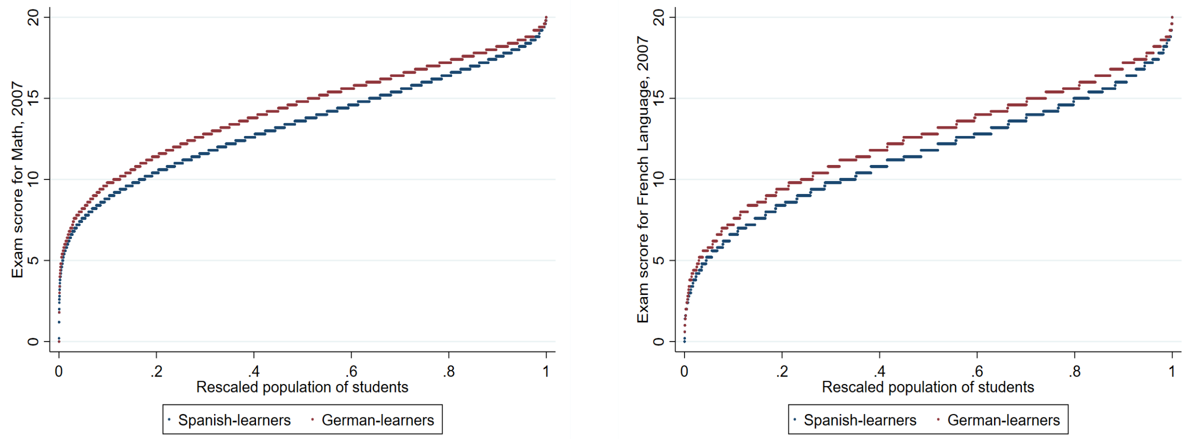

Figure 2: The graphs show the distribution of exam scores for 2007 National test. Populations of students are rescaled to be between zero and one, all scores are shown in ascending order, and each dot represents one student with Spanish-learners in blue colour and German-learners in red colour. The left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

data–students who choose German have better grades than those who choose Spanish. We also show that the strength of sorting varies as the cost of choosing German varies.[31] We then show that sorting is associated with socioeconomic inequality–students choosing different languages differ across many background characteristics. When performing matching, the balance across academic progress and socioeconomic background is achieved. Using matching, we find that students choosing German enjoy larger positive peer effects, and that creates inequality in the education experience.

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 subjects. When entering junior high school, students take National standardized tests in French and Mathematics.[32] Table 2 shows that students who would choose to study German scored 0.9 points higher (out of 20) than the ones studying Spanish in both French and Mathematics.[33] We further illustrate the sorting effects with the graphs of the distributions of the scores in Figure 5. One can clearly see that grades of German-learning students first-order stochastically dominates grades of Spanish-learning students.

As part of the data collection for the panel, students passed a special 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

[31] In Appendix, 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.

[32] All french students passed this test, whether they are included in this data or not.

[33] The grades to the different tests are on a different scale. Throughout the paper, to harmonize the grades and allow for direct comparison, we rescale some grades to be on a 0-20 scale, which is the typical grade scale in France.

	National Test		Special Test				
	Mathematics	French Language	Mathematics	French Language			
				Treatment of Incomplete Sentences	Understanding	Lexicon	Reasoning
German	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)
α	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 subjects 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.

data, geometry) and the French language (lexicon and comprehension). We estimate the following simple model of academic performance that help to highlight the differences between those who chose German and those who chose Spanish:

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

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 idd noise.

We use OLS to estimate Model (1), where the outcome variable is test scores. Table 2 reports the results–students learning 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%.

We further explore the differences between students choosing different second foreign languages and show that students sorting into German have 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 housing, whether the child has his own room, whether the parents studied after high school, whether the child was born in France, and whether they were born in France.

We use Model (1) and employ various socioeconomic indicators as O_i. Table 3 reports the results of OLS estimation. We can clearly see that students from more wealthy and educated families choose German over Spanish– the corresponding variable is positive and highly statistically significant for related outcome variables. In particular, parents of students studying German earn 236 euros more per month, their house is larger by a quarter of the room, and the child is 2.3% more likely to have his 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 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 enjoy 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

	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	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)
α	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 its 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 is born in France (*Born in F*), whether the mother is born in France (*M Born in F*) and whether the father is born in France (*F Born in F*). Parent's involvement indicators include: the probability that one of the parents is in a class representative (*Representative*) and is part of the Parents' Association (*Association*).

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 students' schooling. The last two columns in Table 3 shows that, in line with the intuition presented, students choosing German are about 3% more likely to have parents that are actively involved in the scholarly of their children.[34]

4.2 Varying Cost and Strength 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– average abilities of students in both classes also increase. 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.[35] This assumption is based on the fact that in a border area, students are more likely to have family members speaking the corresponding foreign language and 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.

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

[35]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.

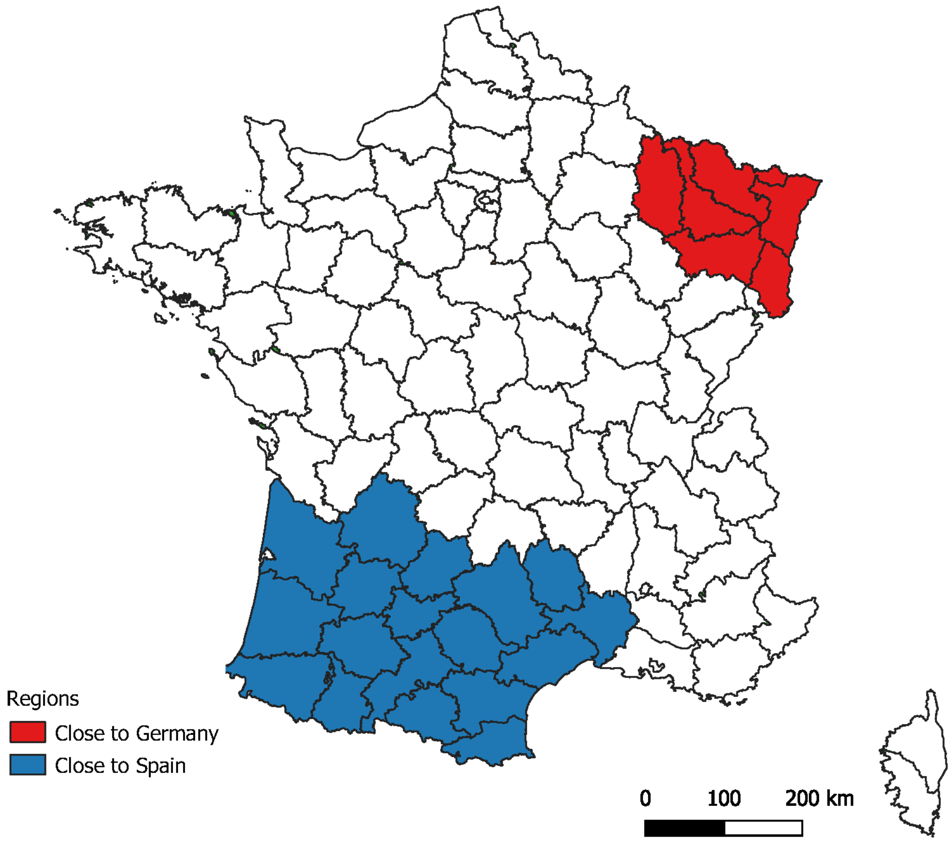

Figure 3: Map of France with "Close to Germany" and "Close to Spain" regions indicated.

The available data allows us to identify the department–unit of the administrative divisions of France–where the observed student attends a school.[36] 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.[37] We follow the identification strategy from Section 4.1 and estimate sorting effects separately for each region using the data on socioeconomic background and academic progress measured using 2007 and 2008 tests. We focus on the difference between the strength of sorting between Close to Germany and Close to Spain regions as those differ most in terms of the cost. We run the following regression:

$$O_i = \alpha + German \times \mathbb{1}_{i\ chose\ German} + Close \times \mathbb{1}_{i\ close\ to\ German} + \\ GermanAndClose \times \mathbb{1}_{i\ chose\ German} \times \mathbb{1}_{i\ close\ to\ German} + \epsilon_i, \quad (2)$$

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, $\mathbb{1}_{i\ close\ to\ German}$ 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 idd noise.

Table 4 provides the results of the OLS estimation of Model (2). The coefficient for

[36]We identify in which department students study using the department in which they lived in 2007 at the start of the panel.

[37]The Close to Germany region includes Lorraine and Alsace, the Close to Spain region includes Aquitane, Midi Pyrenees, and Lanquedoc.

	National Test		Special Test				
	Mathematics	French Language	Mathematics	French Language			
				Treatment of Incomplete Sentences	Understanding	Lexicon	Reasoning
German	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	−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)
GermanAndClose	−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+ GermanAndClose	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 subjects 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. 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 the OLS.

Close is negative and statistically significant for the majority of subjects. 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 *GermanAndClose* 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–$German + GermanAndClose$ 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 showed that students studying German have better grades before making the choice of a second foreign language, but it is also true if one considers exams at the end of junior high school. In this section, we show that the performance of German-learning students compared to Spanish-learning students improved more from 2009 to 2011, and that may suggest that (sorted) 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 seven thousand of 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 only one year together before taking the test, and it is possibly too little time for peer effects to reveal themselves.

	Final Grade	Math	French Language	History & Geography	Physic & Chemistry	Sports
	Exam & Cont.	Exam	Exam	Exam	Cont.	Cont.
German	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 selected subjects results in the National Diploma between matched students who choose German and Spanish after studying two years attributed to peer effects.

Second, the specific test was not incentivized, unlike the National Diploma, which plays an important role in students' life paths.[38]

Table 5 contains the results of the PSM-based estimation of peer effects for a number of subjects from National Diploma.[39] 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: continuous assessment and exam. 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 in diploma.[40] The Final Grade is a weighted average of all subjects in the National Diploma, where scores for Mathematics and French Language are the most impactful, and the effect on grade is positive and significant at 10%. Effects for subjects' exams are positive, but is significant at 5% only for Mathematics.[41] Further, the effect is large and statistically significant at 1% for History and Geography exam. We also consider two subjects with continuous assessment marks. First, we include Psychics and Chemistry, where one can expect an education production function that strongly depends on peers. Indeed, we find a positive

[38]As an evidence of the importance of the National Diploma we provide an illustrative graph with average exam scores in the National Diploma 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 analysis of exam scores bunching in high-stakes tests.

[39]PSM-based and GM-based estimates are very similar across subjects included in National Diploma with GM-based effects being more conservative with smaller sizes of estimated effects, see Table 18 in Appendix K for illustrative comparison.

[40]We investigate the implications of the variation in the strength of sorting effects, established in Section 4.2, on peer effects. For that, 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 allowed for one difference in the estimation process and did not include revenue among socioeconomic characteristics as there are many missing observations for this variable. The results are under-powered due to sample splitting indicate 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 are peer effects as sorting of the peers is more intense, see Figure 7 in Appendix J

[41]Interestingly, the estimated effect on 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.

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. Placebo test is passed with no differences in grades for Sports.[42]

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.[43] In this way, the results suggest that students choosing German perform better in many important subjects because of the class compositions that is a product of endogenous tracking.[44] 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.

5 Discussion

Many countries around the world shaped their educational systems to harmonize students' junior high school experience. Over the last half of a century, many European countries, including France, abandoned early tracking. This paper shows that educational system features, such as fixing class composition for all subjects based on the second foreign language choice in France, may lead to endogenous tracking with sorting effects. Such endogenous sorting within schools is less evident than, for example, sorting between schools and corresponding residential areas, but is nevertheless associated with sorting not only by ability but also by socioeconomic status. We would like to point out that any policy recommendation that can be developed based on the reported results should be treated with care due to probable general equilibrium effects that would follow possible interventions. 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.[45]

In the absence of experiment-like data with random allocation to classes, we use observational data in our analysis. To use non-experimental data, we make some assumptions and simplifications in the paper that require explicit discussion as they may naturally impose some limits on the accuracy of the presented results. Hereafter we discuss some of the limitations while commenting on why those issues may not be critical for the validity of the analysis.

First, we assume that the sunk cost of choosing German over Spanish is the same across agents in a model, and this is, of course, a simplification. The model can be extended to include heterogeneous costs and accommodate different costs within each school cohort. Preliminary theoretical analysis suggests that results on sorting would hold in this case, but we leave the formal treatment of the model for future research alongside other extensions.

[42]In fact, oftentimes, students' progress in sports is evaluated against their own previous records leaving small room for peer effects to have an impact on grades in sports.

[43]We do not distinguish between different channels through which peers affect performance, and those may include learning from peers, the overall speed of class progress, teachers tailoring lectures, etc.

[44]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.

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

Second, in the model, we assume common knowledge of the coordination game that is played. This may not be a very realistic assumption as some students may not even be aware of the strategic implications of the language choice they make.[46] In particular, it is often the case that students receive advice from their involved parents who know about the "game" implying there are some inter-generations information spillover effects that we briefly discuss in the corresponding Section 4.1.

Third, to identify peer effects, we use matching on observables, and results are sensitive to the validity of the assumption that all relevant factors are included in the list of characteristics we match on.[47] Omitted variable bias may be presented in our estimates. One may assume, for example, that there are some extra-motivated students who join German classes and improve over time more than their matched-on-observable Spanish-learning counterparts. In this case, since we do not observe motivation and don't match based on it, we may wrongly attribute the academic progress of such students to peer effects, not to extra motivation. While this is possible, we believe that employed observables from the panel specifically designed by the French Ministry of Education to collect all relevant factors of academic progress suffice to avoid bias.

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.

Adsera, A., Chiswick, B.R., 2007. Are there gender and country of origin differences in immigrant labor market outcomes across european destinations? Journal of Population Economics 20, 495–526.

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., Elder, T.E., Taber, C.R., 2005. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. Journal of Political Economy 113, 151–184.

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.

[46] A natural extension of our theoretical model is to include a possibility of partial knowledge about the game in our model. At the same time, theoretical modeling of games with unawareness is a non-trivial problem, and the related field is still under development, see Schipper (2014) for an overview of the current stance of the literature.

[47] To identify the omitted variable bias one can assume that the relationship between treatment effect and unobservables is recoverable from the relationship between treatment effect and observables, as in Altonji et al. (2005).

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., 2019. Synthetic difference in differences. Technical Report. National Bureau of Economic Research.

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

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

Bonhomme, S., Sauder, U., 2011. Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling. Review of Economics and Statistics 93, 479–494.

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.

Brunello, G., Checchi, D., 2007. Does school tracking affect equality of opportunity? New international evidence. Economic Policy 22, 782–861.

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.

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.

Craig, A., Vierø, M.L., 2013. Academia or the private sector? Sorting of agents into institutions and an outside sector. The BE Journal of Theoretical Economics 13, 303–345.

Damiano, E., Li, H., Suen, W., 2012. Competing for talents. Journal of Economic Theory 147, 2190–2219.

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.

Falk, A., Ichino, A., 2006. Clean evidence on peer effects. Journal of Labor Economics 24, 39–57.

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

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

Garlick, R., 2018. Academic peer effects with different group assignment policies: Residential tracking versus random assignment. American Economic Journal: Applied Economics 10, 345–69.

Hanushek, E.A., Wößmann, L., 2006. Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. The Economic Journal 116, C63–C76.

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. Does competition among public schools benefit students and taxpayers? American Economic Review 90, 1209–1238.

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

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

Keane, M.P., Wolpin, K.I., 1997. The career decisions of young men. Journal of Political Economy 105, 473–522.

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.

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

MacLeod, W.B., Urquiola, M., 2009. Anti-lemons: School reputation and educational quality. Technical Report. National Bureau of Economic Research.

Mas, A., Moretti, E., 2009. Peers at work. American Economic Review 99, 112–45.

Matthewes, S.H., 2021. Better together? Heterogeneous effects of tracking on student achievement. The Economic Journal 131, 1269–1307.

Meghir, C., Palme, M., 2005. Educational reform, ability, and family background. American Economic Review 95, 414–424.

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

OECD, 2010. Pisa 2009 results: What makes a school successful? – Resources, policies and practices. (vol. iv) URL: http://dx.doi.org/10.1787/9789264091559-en.

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.

Pischke, J.S., Manning, A., 2006. Comprehensive versus selective schooling in England in Wales: What do we know?

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

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.

Schipper, B.C., 2014. Unawareness—a gentle introduction to both the literature and the special issue. Mathematical Social Sciences 70, 1–9.

Slavin, R.E., 1990. Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. Review of Educational Research 60, 471–499.

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.

Sorenson, O., Dahl, M.S., Canales, R., Burton, M.D., 2021. Do startup employees earn more in the long run? Organization Science .

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.

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

Willis, R.J., Rosen, S., 1979. Education and self-selection. Journal of Political Economy 87, S7–S36.

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

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 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.[48] 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*

[48] 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.[49] 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* $\dfrac{a_k + .. + a_N}{N - k + 1} - \overline{a}_S^k \leq C \leq \overline{a}_G^k - \dfrac{a_1 + .. + a_k + a_N}{k + 1}$.

[49] Here we do not model preferences for ranking that may affect decisions to join more competitive or less competitive groups. See Villeval (2020) for a recent overview on existing evidence.

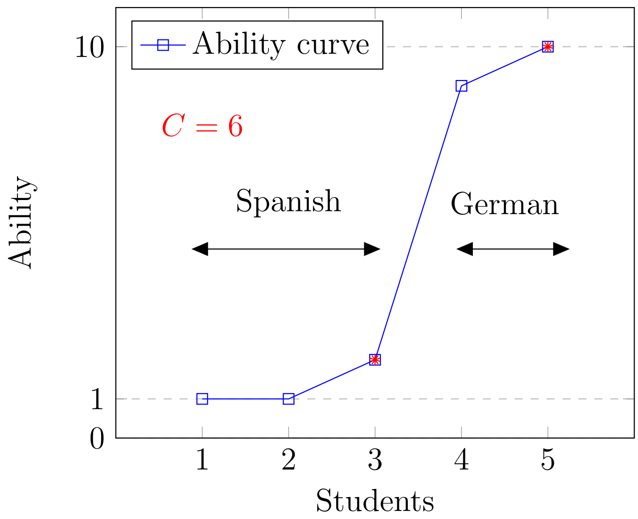

Figure 4: An example of the sorting equilibrium in a cohort of students with abilities $\{1, 1, 2, 9, 10\}$ and relative cost of choosing German language is $C = 6$. Students with highest ability in each group are marked with red starts–their incentives not to deviate are ensured by the relative cost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

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 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:

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

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

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

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

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

$\square$

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	0.306^{***}	$0.111^{\dagger}$	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	0.380^{***}	0.148^{*}	$0.169^{\dagger}$	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	0.287^{***}	0.053	0.190	0.113	$0.241^{\dagger}$
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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	−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	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	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†)	(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†)	(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†)
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

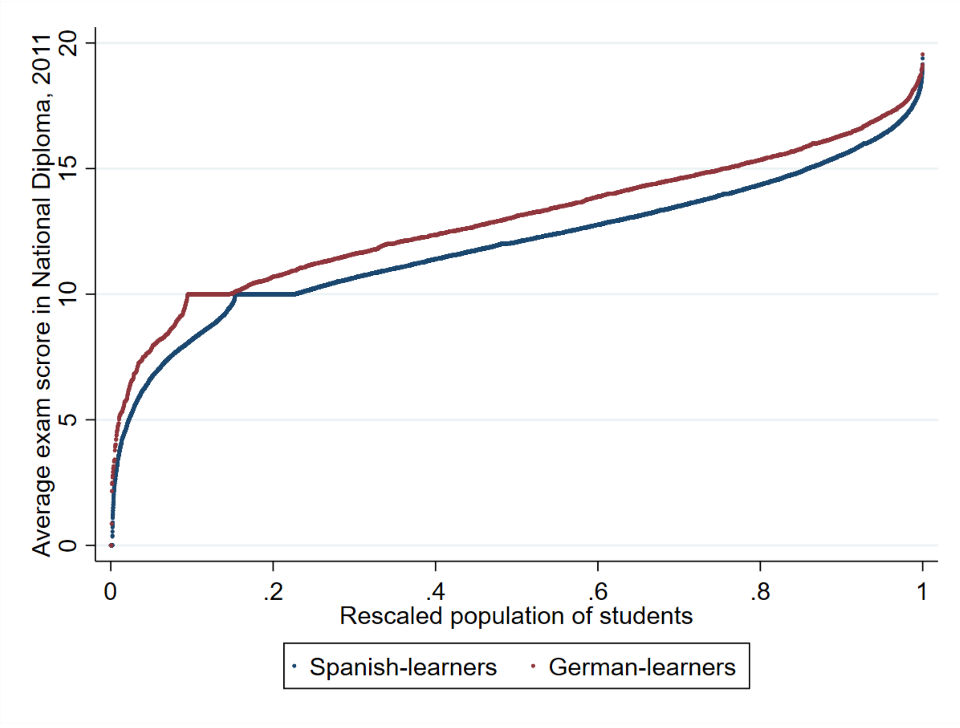

Figure 5: The graphs show the distribution of exam scores for the National Diploma. Populations of students are rescaled to be between zero and one, all scores are shown in ascending order, and each dot represents one student with Spanish-learners in blue colour and German-learners in red colour.

G Propensity score matching estimates

	Overall Average (Cont. and exam)				
German	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	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	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	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	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	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	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	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	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	0.260^{*}	$0.184^{\dagger}$	$0.248^{\dagger}$	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	0.429^{**}	0.336^{***}	0.347^{*}	$0.273^{\dagger}$	0.404^{**}
	(0.001)	(0.001)	(0.011)	(0.070)	(0.008)
N	17902	16996	9029	7004	6970
	Biology (Cont.)				
German	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	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	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	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	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	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	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

	Mean value			t-test	
Variable	Treated	Control	bias(%)	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 subjects–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

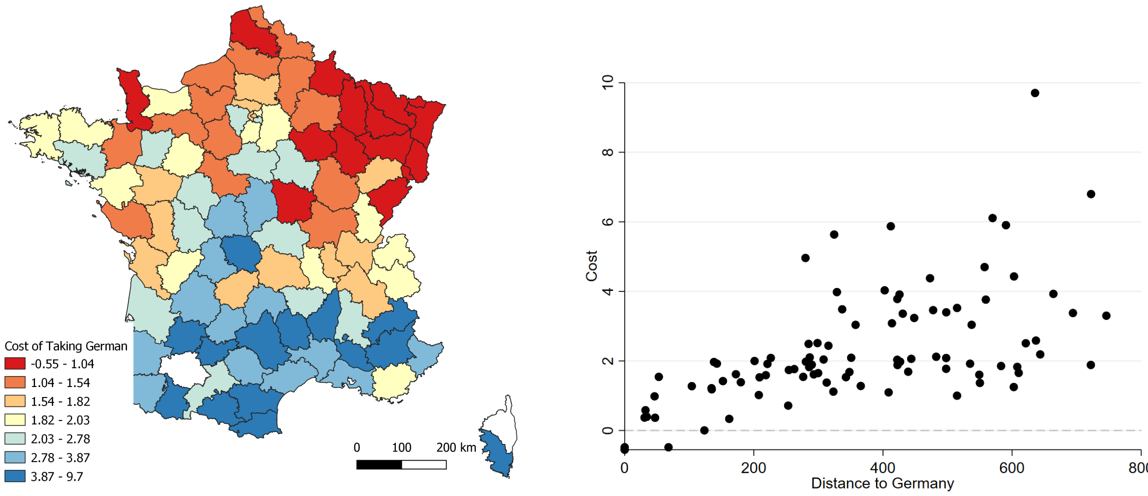

Figure 6: Left panel depicts the map where for each department different estimated costs of taking German are indicated with different colors from red (zero cost) to dark blue (large cost). Right panel depicts the scatter plot of all departments with the distance to the German border (x-axis) and cost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

students studying German.[50] 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.[51] 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.[52]

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

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

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

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

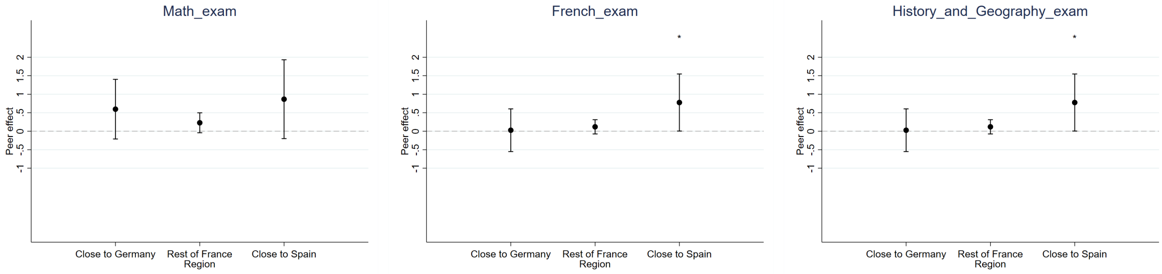

Figure 7: Peer effects for Mathematics, French Language, and History and Geography exams estimated separately for "Close to Germany", "Rest of France", and "Close to Spain" regions.

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	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	$-0.040^{\dagger}$	-0.082^{**}	$-0.048^{\dagger}$	$-0.017^{\dagger}$	-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*Close Germany	-579.612	0.017	-0.036	-0.120^{*}	-0.168^{**}	-0.012	$0.067^{\dagger}$	0.012	-0.033	$-0.067^{\dagger}$
	(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+German*Close Germany	117.734	0.079	-0.023	0.045	0.036	0.000	-0.010	-0.033	0.012	0.010
	(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 selected subjects results in the National Diploma between matched students who choose German and Spanish after studying two years in classes based on second foreign language choice attributed to peer effects.

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.[53] 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.

[53] 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	-0.02	0.04	-0.35	0.72	-0.10	0.07
French essay							
	German	0.04	0.07	0.62	0.54	-0.09	0.17
French dictation							
	German	-0.06	0.06	-0.92	0.36	-0.18	0.06
French exam							
	German	-0.02	0.04	-0.35	0.72	-0.10	0.07
Math Exam							
	German	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

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

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

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

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

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

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.