ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DEPARTMENT OF ECONOMICS

SECOND CYCLE DEGREE

ECONOMICS AND ECONOMETRICS

DO SCHOOL BUREAUCRATS DISCRIMINATE AGAINST FOREIGNERS? EXPERIMENTAL EVIDENCE FROM ITALY

Supervisor

Prof. Enrico Cantoni

Defended by

Alessio Giorgetta

Graduation Session July 2025

Academic Year 2024/2025

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DEPARTMENT OF ECONOMICS

SECOND CYCLE DEGREE

ECONOMICS AND ECONOMETRICS

DO SCHOOL BUREAUCRATS DISCRIMINATE AGAINST FOREIGNERS? EXPERIMENTAL EVIDENCE FROM ITALY

Supervisor

Prof. Enrico Cantoni

Defended by

Alessio Giorgetta

Graduation Session July 2025

Academic Year 2024/2025

DO SCHOOL BUREAUCRATS DISCRIMINATE AGAINST FOREIGNERS? EXPERIMENTAL EVIDENCE FROM ITALY

ALESSIO GIORGETTA*

ABSTRACT

I study whether public Italian schools discriminate against foreigners. Within province, I contact each school inquiring about enrollment procedures and welfare with either an Italian, French, or Arabic alias. I show that clerks are 5.32% less likely to answer if the sender is a foreigner. Moreover, lower income per capita increases bias: 10% of schools show biases $\geq$ 11%. When bureaucrats do respond, they are 17.6% less likely to send informative answers and 4.77% less likely to be polite. These results contradict the widely recognized theories of statistical discrimination and contact theory, at least in this setting. Economic theories of taste-based discrimination, jointly with psychological theories that postulate bias against outer groups depends on the feelings that group causes, are able to explain the observed discrimination.

I. INTRODUCTION

Education is a pillar of modern democracies. In a world where information is faster and faster, fake news are on the rise and social networks are replacing old media, education creates citizens able to face these challenges.

On the contrary, when a state discriminates a social group, preventing it from accessing education, it is essentially creating a threat for itself. Besides the economic disadvantages an uneducated population poses, when people lack the cognitive instruments that school provides, misinformation, conspiracies and, even, terrorism stand against the objectives that a modern democracy has.

Since discriminatory practices in education may be the worse way in which a State hurts itself, the Italian law recognizes education as a universal right. In other words, everybody has the right and the duty to enroll in the school system. Moreover, laws regulating immigration

*I would like to thank Enrico Cantoni for the careful, precise supervision. A thanking goes to Federica, Sofia and Giulio for the useful comments on my work and their loving support. I would like to thank Gabriele, Anna, Giuseppe, Andrea, and Maurizia for everything: Bologna wouldn't have been the same without you. The final thaning goes to my parents, Fabrizio and Mirna, and my brother Manuele.

specify that foreigners should be treated by schools as citizens and, furthermore, are assigned some rights natives are not recognized.[1] If these law provisions were respected, each foreigner, regardless of his right to stay in Italy, would access the Italian educational system and collect the profits education guarantees.

While the law mandates absolute equality in education between citizens and foreigners, the State is actually embodied trough a chain-of-power in its lower-level bureaucrats. Indeed, clerks are asked day by day to apply the law within a discretionary space. Marginal bias in the behavior of clerks may result in strong, widespread, and hurtful discrimination against foreigners. Brodkin (1997), a seminal work in the analysis of bureaucracy, argues that clerks show racial, political, and social bias when they are given flexibility in how to perform their duties and no explicit procedures are in place to prevent discrimination. Bias can arise for two reasons (White, Nathan, and Faller (2015)): statistical or taste-based discrimination. In the case of statistical discrimination, clerks may be inferring people's characteristics from their name, skin color or race (see f.e. Agan and Starr (2018)). Regarding education, if foreigners are more likely to be from disadvantaged socioeconomic backgrounds, when students they may require more attention and resources than Italian students. Clerks may infer this and discriminate against them to reduce the costs, both in terms of resources and effort, the school must face. On the contrary, in the case of taste based discrimination, people are averse to cross-group contact (Becker (1971)). Clerks may get a disutility from the interaction with foreigners and thus maximize their utility by discriminating against them (Charles and Guryan (2009)). This idea corresponds to the theory of social psychology developed by Lee and Fiske (2006): relations with outer groups depend on stereotypes, specifically a combination of how warm, pleasant, and enjoyable and how capable, rich, and successful a social group is.

The theories above explain *why* discrimination may exist but do not explain *how* it behaves. Again, social psychology has developed the *contact hypothesis*. This theory states that prejudice is reduced by positive and meaningful inter-group contacts. Thus, as long as, inter-group interactions are positive more interactions translate into less bias (Pettigrew and Tropp (2006)).

This article tests whether clerks in the Italian school system engage in discrimination against foreigners and discusses why. I do so by following the method proposed by Bertrand and Mullainathan (2004): each public school in Italy is contacted with an email by a fake parent, inquiring on the procedures to enroll their son and possible benefits for people from low socioeconomic backgrounds. All email are identical in the text but differ in the name of the parent. Aliases may be of either Italian, French or Arabic people. Randomization is performed at the province level[2] to ensure balancedness across unobservables. I then measure whether the response rate within a week is differential by treatment group and whether, conditional on the school responding, the responses differ in how informative and polite they are.

I find that foreigners are 4.3pp less likely to receive an answer from schools, against

[1] This is a recognized principle in the Italian Constitution: discriminated groups shall be brought by the State to equality with undiscriminated groups. See Article 3 of the Constitution for reference.

[2] The second level administrative division in Italy.

a baseline outcome for Italians of 81%. This effect is concentrated among French rather than Arabic names (6pp against 2.5pp respectively). The results are robust across all model specifications. Moreover, I find no effect of either the number of students in the school, the population of the municipality, or the percentage of foreign residents on the size of the bias. On the contrary, below-the-mean per capita income increases bias. Finally, I find that emails in response to foreign aliases are 17.6% less likely to be informative and 4.77% less likely to be polite (against a mean for Italians of 35% and 82% respectively). Discrimination is thus twofold and follows an additive pattern.

Given this results, I argue statistical discrimination is unlikely to be at play. French foreigners are are on average of higher economic status than both Arabic and Italian people but, nonetheless, they are the least likely to receive an answer from schools. If statistical discrimination was driving the response bias, French foreigners should show higher response rates than Italians, as on average their advantaged background would make them better students. The results are better interpretable under the light of taste-based discrimination and, specifically, adopting the framework of Lee and Fiske (2006): if French are perceived as a low likability-high capability group, whereas Arabs are seen as a low likability-low capability group it is possible that the feelings of enviousness against the first outsize those of anger against the second (Esses (2021)). This could explain why both groups of foreigners are discriminated but French are far more likely to remain unanswered. Moreover, I find no evidence of contact theory. While results in the psychology literature (Pettigrew and Tropp (2006)) show that on average contacts reduce bias, neither the population size of a municipality nor the fraction of resident foreigners on the overall population has any effect on the size of bias. Partly consistent with the theory of Brodkin (1997), municipalities with higher per capita income show less bias towards foreigners. When resources are scarcer, they cost more and the marginal cost of assigning them to a stranger may play a higher role than when resources are widely available. In other words, the disutility of giving resources to outer groups could be increasing in the cost of the resources themselves.

This article contributes to three, very related, strands of literature. The first is concerned with documenting discrimination against social groups. For example, the seminal work of Bertrand and Mullainathan (2004) showed that, in the U.S., Blacks are discriminated in the labor market even when their qualifications match those of Whites. This result has spurred a vast literature documenting both discrimination by privates (see for example Agan and Starr (2018) and Doleac and Benjamin Hansen (2020)) and public figures (as in the case of White, Nathan, and Faller (2015)). The second strand of literature is policy related and concerned with the educational gap that has been documented between foreigners and natives (Fryer Jr and Levitt (2004)) and their long term effects (Guyon, Maurin, and McNally (2012), Brunello and Checchi (2007)). Policies against such *educational segregation* has been recently proposed by Carlana, La Ferrara, and Pinotti (2022), in the case of foreigners, and Biroli et al. (2024), in the case of people from low socioeconomic status. Finally, this article interacts with the theories postulating *why* and *how* discrimination happens. The foundational ideas of such literature come from the analysis of bureaucracy (Brodkin 1997) and the social psychology literature (Pettigrew and Tropp (2006), Lee and Fiske (2006)).

The article proceeds as follows: Section II provides a general overview on the Italian school system and the data gathered (Sections II.A and II.B), explains the experiment (Section II.C) and how it has been implemented (Section II.D); Section III presents the results, with Section III.A concentrating on the differential response rates, Section III.B on heterogeneity in the results and Section III.C on the text analysis of the responses; Section IV addresses limitations on the general framework of the experiment and its validity (Section IV.A) and its specific implementation (Section IV.B); Section V concludes.

II. EMPIRICAL FRAMEWORK

The Italian school system is predominantly public and free up to the 5^{th} year of high-school. Education is recognized by the Constitution as a universal right and immigration laws prohibit discrimination by schools against foreigners, regardless of their legal status. I exploit administrative data that comprise all public Italian schools under the supervision of the Ministry of Education to test whether institutes contravene these legal provisions. Specifically, each school received one email by a fictional parent inquiring about enrolling procedures. Emails are either signed by an Italian, French or Arabic name. I then test whether the response rate after one week differs across the three groups (Section III.A) and if discrimination is also present in the text of the response emails (Section III.C).

II.A HOW DO ITALIAN SCHOOLS OPERATE?

Education in Italy is almost universally public and free (Biroli et al. (2024)). The responsibility of organizing, controlling and supervising the schools is attributed to the Ministry of Eduction (MIM). Up to 19 years old, every person has the right to education, although mandatory schooling ends with the 16^{th} birthday of students. Education is divided into three consecutive path: primary education (6 to 11 years-old), lower secondary education (11 to 14 years-old) and upper secondary education (14 to 19 years-old).[3]

Article 34 of the Constitution specifies that education is a universal right, i.e. everybody has the right to enroll into the Italian school system and, moreover, everybody has the duty to do so up to at least 8 eight years of education. Along the same lines, Article 38 of the ”Testo Unico sull'Immigrazione”, the main body of law regulating immigration, adds that: (i) non-citizens are mandated to enroll into the educational systems, under the same criteria of citizens; (ii) proposition (i) must be satisfied regardless of any legal title (i.e., even in the case of clandestines); (iii) consequently, foreigners are not required to present any document to prove their right to be in Italy.

Law dispositions thus prohibit any form of discrimination and, in fact, foster the enrollment of non-citizens. Consequently, discrimination may not only per hurtful *per se*, for example perpetuating disadvantage, but also is a direct violation of the law. Previous research has already shown that, once enrolled, 1^{st} and 2^{nd} generation foreigners undergo a process

[3]For reference see the website of the MIM.

of educational segregation: given the same scores in national standardise tests, they choose less demanding, and thus rewarding, schools (Carlana, La Ferrara, and Pinotti (2022)). This paper tests whether discrimination is present *during enrollment* rather than after it.

II.B DATA

I exploit a comprehensive, administrative dataset publicly downloadable from the Italian Ministry of Education. The data contains information on government-controlled institutes, in 18 regions out of 20.[4] This is actually an advantage for the design of the experiment, since the missing regions have a high presence of German-speaking and French-speaking minorities, which could have possibly invalidated my results.[5] The data comprises information on the schools' official emails, school characteristics such as the number of students or the state of the buildings, self-evaluations of results and internal procedures, and geographic information.

I have a total of 50,511 schools of which I discard: (i) schools that are not classified as standard, f.e. schools within penitentiaries; (ii) schools which lack contact information or their geographic location; (iii) schools that lack an administrative office. Operation (iii) is necessary since most schools share the administrative branch with larger schools and are not assigned a unique official email. Filtering for institutes that house the administrative office guarantees that one bureaucrat does not receive more than one email. I end up with 7,192 schools, each of which is associated with a unique official email address.

I augment these data with socioeconomic information on municipalities where schools are located: population data comes from the Italian Statistics Bureau (ISTAT) while economic data comes from the Ministry of Economics and Finance (MEF). I thus have information on: municipalities population; number of taxpayers; breakdown of income levels and types. These data allows for heterogeneity analysis and model selection (in combination f.e. with a double-post LASSO procedure). Notice these data are at municipality level. Schools in large cities, which host more than one institute, thus share identical municipal covariates.

II.C RANDOMIZATION & TREATMENTS: DIFFERENT SIGNATURES

Each school receives an email from a fictional parent inquiring about enrollment procedures for their son. Specifically, the email asks for: (i) when the administrative office is open, whether an appointment is necessary and if the office is open exclusively on weekdays; (ii) whether schools offer afternoon activities for the students and, in the case these are not free, whether reductions of their cost are possible handing in income declarations. The email is purposefully long and full of questions. This implies that answering is costly for the bureaucrat receiving it. A clerk is more likely to show bias when he is required effort (Brodkin (1997)). Such structure also serves the purpose of favoring text analysis: many questions imply many possible answers, which lend themselves to analysis through Natural Language Processing

[4] Schools of Trentino-Alto Adige and Valle d'Aosta, two autonomous regions, are missing from the dataset since they are under the supervision of local authorities.

[5] See the MIM's website for data and other information.

models (see Section III.C). The text of the email received by administrative offices is the following:

> *Good morning,*
> *my name is #NAME#. I will move to #MUNICIPALITY# next year with my family. I would like to ask when the administrative office is open because I would like to ask some questions about how to enroll my son.*
> *Do I need an appointment? Are you only open during weekdays?*
> *By chance, do you have afternoon activities? Are there income requirements to join them or to have reductions of the cost?*
> *Thank you,*
> *#NAME#*

where #NAME# is an Italian, French or Arabic name and #MUNICIPALITY# is the location of the school. The personalization of the email based on the institutes' location is a useful precaution to reduce the risk an email is considered spam by the bureaucrat receiving it. However, this only constitutes a safety check: emails are only sent to official addresses, contain no attached link and only ask school-related question. Thus, they are hardly classifiable as spam, phishing or other similar hacking practices. For a screenshot of a representative email, see Figure A1 in the Appendix.

The experiment consists in replacing the #NAME# with an Italian (control), French (T1) or Arabic (T2) name. Given the large number of observations, I randomize at the province level. The dataset contains 103 provinces, guaranteeing a balanced randomization, spread across all the national territory. Within each province, 50% of schools are allocated to the control group, 25% to French signatures and 25% to Arabic signatures. Since GMAIL sending limits prevented a bulk-sending of all emails in one day, these were then randomized across nine working days, within the span of two weeks. This randomization was independent of the province level one.[6]

The main outcome of interest is whether a receiving school responded within a week. I estimate the treatment effect with a Linear Probability Model, but Appendix reports estimates with LOGIT (Table A2) and double-post LASSO (Table A3, following Belloni, Chernozhukov, and C. Hansen (2014a) and Belloni, Chernozhukov, and C. Hansen (2014b)) to address concerns about the distribution of answers and the optimal model selection. Conditional on a school answering, text analysis was performed on the responses received. A NLP model was used to classify if each response was: (i) useful or not; (ii) polite or not. The criteria used for these evaluations of answers can be found in the following section on the implementation of the experiment, Section II.D.

[6]A necessary procedure since not all provinces had nine schools per treatment arm.

II.D IMPLEMENTATION

Table 1 checks the balancing of covariates across experiment arms, by reporting the results of a Seemingly Unrelated Regression performed on a system of equations where each row is a regression of the covariate on treatment dummies. Although some coefficients are statistically different from zero, this is to be expected with a high number of covariates. Moreover, the results of an F-Test on the joint significance of treatment across equation are reassuring that the randomization worked properly.

Table 1:
Randomization Check: Balance Table
School & Municipality Level Covariates

	Control Group -	French Treatmen	Arabic Treatmen	Any Treatmen	Data Level
	(1)	(2)	(3)	(4)	(5)
# Students	812	-2	-9	-5	S
	(312)	(8)	(8)	(7)	
INVALSI	4.23	-0.01	0.02	0.01	S
	(1.22)	(0.04)	(0.04)	(0.03)	
Inclusivity	5.24	0.03	0.00	0.01	S
	(0.92)	(0.03)	(0.03)	(0.02)	
Area in Decline	0.012	0.003	0.001	0.002	S
	(0.107)	(0.003)	(0.003)	(0.003)	
Train Station	0.151	0.011	-0.009	0.001	S
	(0.358)	(0.013)	(0.013)	(0.011)	
Bus Station	0.740	0.039 *	0.004	0.023 ~	S
	(0.439)	(0.015)	(0.016)	(0.013)	
Cantine	0.182	-0.015	-0.015	-0.015	S
	(0.386)	(0.011)	(0.011)	(0.009)	
Gym	0.641	-0.026 ~	-0.031 *	-0.028 *	S
	(0.480)	(0.014)	(0.014)	(0.012)	
Taxable Earnings	22382	49	65	57	M
(per capita)	(4522)	(82)	(84)	(68)	
Population	215	12	6	9	M
(by 1000)	(574)	(11)	(12)	(9)	
Foreigners (%)	0.087	0.001	0.001	0.001 ~	M
	(0.050)	(0.001)	(0.001)	(0.001)	
T1 = T2 = 0 (F-Test)		1.037	1.001	1.112	
T1 = T2 = 0 (p-value)		0.398	0.472	0.262	
Sample Size	3581	5376	5326	7120	

Notes: Treatment-control differences are from a Seemingly Unrelated Regression (SUR) performed on the system of equations. Column (5) presents the level at which data are gathered. It presents an "S" are data is at school level, an "M" if data are at municiaplity level. Standard errors, in parenthesis, are heteroskedastic robust since the randomization is at school level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.10$

To implement the experiment I created eight GMAIL accounts of which four had an Italian name. The remaining were equally split between French and Arabic names. This procedure was necessary to ensure each account sent around one-hundred emails a day, preventing exceeding sending limits and spam flagging. The experiment started on May 12^{th} 2025 and lasted nine days across the span of two weeks.[7] Responses were downloaded on June 4^{th} 2025 and matched to the units in the main dataset based on their official email address.[8] The main outcome variable is a dummy that measures whether schools answered within a week (see Table 2). I chose one week since most the schools answered within a day (70% in the control group, see Table A4). A longer time-span with response rates approaching 90% – 100% for the Italian names would invalidate a Linear Probability Model estimation (see Table A2 for the comparison with LOGIT estimates).

I then ran a text analysis to determine whether (i) usefulness or (ii) politeness varied significantly across treatment groups, conditional on receiving a response. To do so, I randomly drew 500 emails, roughly 10% of the response sample, which I manually flagged as (not) useful and (not) polite. For usefulness I employed the following criteria:

1. A mail is useful if it contains any information other than the opening hours of the administrative office. Any information means not only answers to the questions in the email (f.e. a telephone number to call for information was enough to classify the email as useful).

2. A mail is not useful if it doesn't include the administrative offices opening hours.

Similarly, the following criteria were used for politeness:

1. To be classified as polite a mail had to include openings and greetings.

2. Every email using informal personalization of verbs was classified as not polite, i.e. not using the *Lei* or *Voi* form (which in Italian distinguish formal, polite writing from direct, personal discourse).

3. Emails in caps-lock, regardless of the conditions above, were classified as not polite, lacking a necessary condition of *netiquette*.

After manually flagging these emails, I used them to train BERT in its Italian version (GOOGLE's NLP model).[9] Once trained, my NLP model automatically flagged responses as (not) useful and (not) polite. Before the analysis, the manually flagged answers were dropped from the sample, effectively being only used in model training and validation. Validating metrics of the model are found in Table A1 of the Appendix.

[7] Emails were only sent on the evening before weekdays, since administrative offices are usually closed during weekends. For the same reason emails were sent each evening between 7pm and 8pm.

[8] Unfortunately, this implies some responses were lost since administrative offices sometimes forwarded emails to personal accounts of bureaucrats, making it impossible to trace them the original email I sent.

[9] Training parameters were: 4 (2) emails per batch, 1e-5 (5e-6) learning rate, 4 epochs for usefulness (politeness). The learning rate is standard but quite low: since my criteria are quite simple, this rate is necessary to prevent overfitting. Higher learning rates lead the model to internalize complicated patterns which did not actually reflect the characteristics of interest. The best epoch was chosen based on its F1 score.

III. RESULTS

I find that 68.8% of schools respond within one day (Table A4 in the Appendix). Most institutes respond within one week (81%, Table 2). Similarly, Figure 1 shows that answers are condensed in the first three working hours of the day after treatment. Figure 2 shows average response rates by region, differential by treatment branch. Eastern regions seems to show both higher response rates and less discriminatory practices.[10]

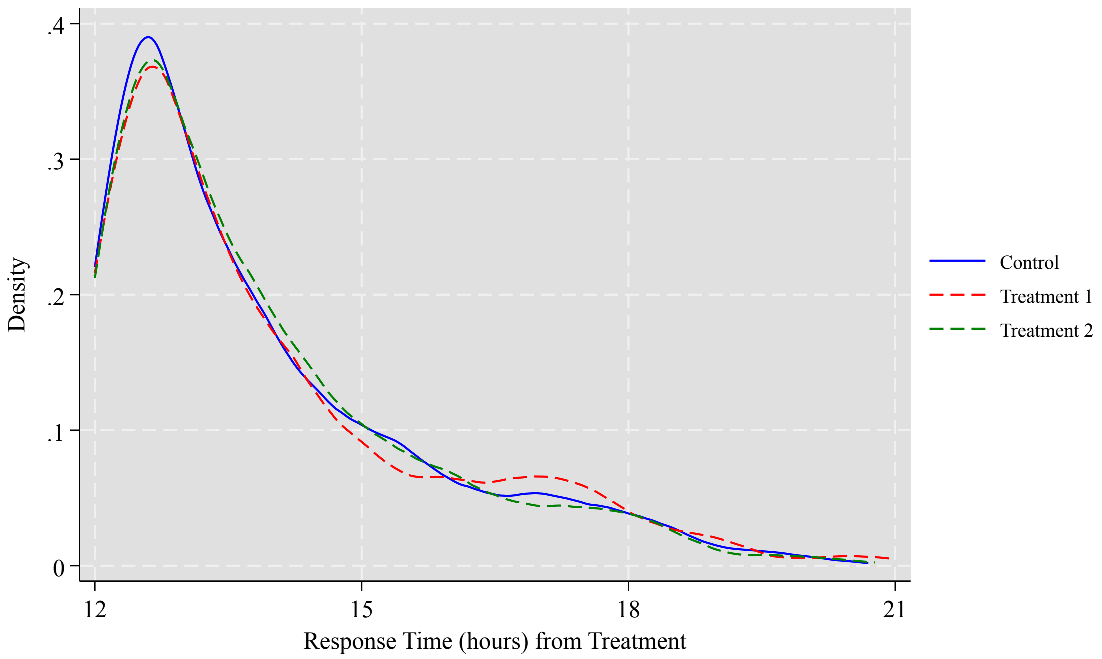


Figure 1: Response Densities Within 24 hours

I find that schools are 4pp less likely to answer within a week if the signature is not Italian, a 5.32% effect at the mean of the control group. This effect seems to be stronger among French signatures rather than only in Arabic signatures (6pp and 2.5pp, respectively). See Table 2. These effects are independent of the schools' number of student, the municipalities' population or fraction of foreign residents. They are instead mitigated by higher income levels (Table 3).

Conditional on receiving a response, I find that discrimination is present in the tone of responses. Emails from treated schools are 5.9pp less likely to be useful and 3.9pp less likely to be polite with respect to the control group mean. These effects are concentrated among Arabic signatures (see Table 4).

Italian schools seem to discriminate against foreigners, especially those from European, high-income countries. Conditional on responding, schools send less informative and polite emails to foreigners, with larger effects for Arabic names. Thus, discrimination displays an additive pattern: a foreigner is less likely to receive an answer and more likely to receive a uninformative, unpolite answer.

[10]The Appendix contains the map for the whole sample, Figure A1.

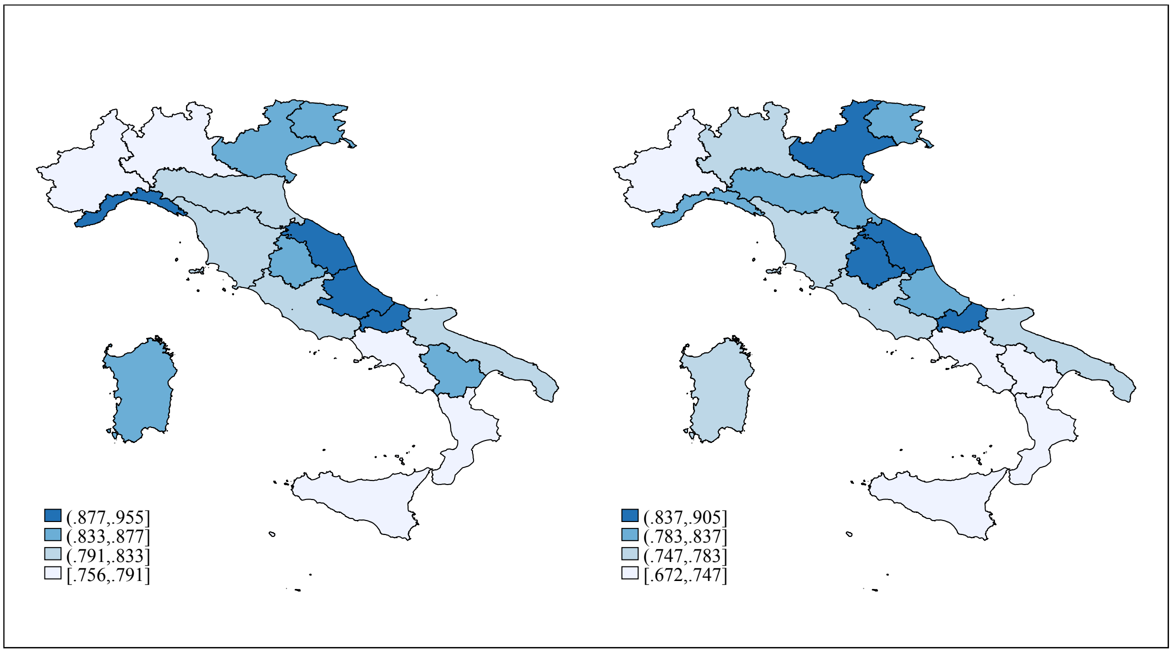


Figure 2: Average Weekly Response Rates by Region: Control (left); Treatment (right).

III.A DO SCHOOLS RESPOND LESS TO FOREIGNERS? MAIN RESULTS.

Table 2 reports LPM estimates of β_i for $i \in \{1,2\}$, coefficients of Equation (1) for $t = 7$:

$$y(t)_{s,p,d} = \alpha + \sum_{i\in\{1,2\}} \beta_i T_i + \sum_{p=1}^{P} \gamma_p + \sum_{d=1}^{9} \kappa_d + X'\eta + \varepsilon(t)_{s,p,d} \tag{1}$$

where $y(t)_{s,p,d}$ is a dummy that equals one if school s, in province p, treated at day d answered within t days. The coefficients of interest are $\beta_{1,2}$ which measure difference between control and treatment groups in the average response rate at t days. γ_p and κ_d are respectively province and day-of-sending fixed effects. X is a set of control variables that includes: school number of students; average taxable earnings; fraction of foreigners; population. Appendix includes estimates of Equation (1) with LOGIT (Table A2), double-LASSO (Table A3) and for $t = 1$ (i.e., one-day response rates, Table A4).

Point estimates in columns (1)-(2) of Table 2 show that a foreign name is responsible for a reduction in the 1-week response rate of 4.1-4.3 percentage points, depending on the specification. Given response rates in the control group of 81%, this implies a reduction of 5.07-5.32%.[11] These results are highly significant and robust across specifications (LOGIT and LASSO estimates in Tables A2-A3 suggest even higher estimates). Columns (3)-(4) break down the effect by treatment: French names are 7.29% less likely to receive an answer; Arabic names instead only 2.97% (with respect to the control group mean). Thus the response discrimination against French signatures is more than double that on Arabic signatures. This result is confirmed by a joint F-test of significance against the null the two coefficients are

[11] A results in line with what found by White, Nathan, and Faller (2015) in the context of U.S. bureaucrats and Latino names.

the same in magnitude.

Table 2:
Treatment Effects on School Responses Rate (after one week)
Main Results

	(1)		(2)		(3)		(4)	
Any Treatment	-0.043	**	-0.041	**				
	(0.010)		(0.010)					
French Treatment (T1)					-0.060	**	-0.059	**
					(0.012)		(0.012)	
Arabic Treatment (T2)					-0.025	*	-0.024	*
					(0.012)		(0.012)	
T1 = T2 = 0 (p-value)					0.000		0.000	
T1 = T2 (p-value)					0.012		0.012	
Outcome Mean (Control)	0.809		0.809		0.809		0.809	
Sample Size	7120		7120		7120		7120	
Strata FE	Y		Y		Y		Y	
Controls	N		Y		N		Y	

Notes: The table reports the estimated coefficients and standard errors for the main equation.The dependent variable is a dummy equalling 1 if the school answered the email within one week. Columns (2) and (4) add as controls: school number of students; average taxable earnings (municipality level); fraction of foreigners (municipality level); population (municiplaity level). All regressions include FE at the province level (stratum of randomization) and day-of-sending level. Standard errors, in parenthesis, are heteroskedastic robust since the randomization is at school level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.10$

Two stylized facts have to be highlighted:

1. First, the results of the experiment are partly consistent with the contact hypothesis (i.e., the idea that stereotypes and discrimination are reduced when groups come into contact and have meaningful interactions, see Pettigrew and Tropp (2006)). Since a large number of immigrants with Arabic names resides in Italy while very few French citizens are present on the national territory, if bureaucrats had positive interactions with Arabic immigrants this would reduce their bottom-line bias. This could explain why French names get significantly more discriminated, as clerks would statistically get less meaningful interactions with French people, given their sparse presence in Italy. However, a more credible hypothesis has been posed in the social psychology literature by Lee and Fiske (2006): perception towards immigrants is determined by the combination of the agreeableness and capability[12] of groups of immigrants. French foreigners may be perceived as low agreeableness and high capability immigrants, implying feelings of enviousness and jealousy in the schools' clerks (Fiske and Lee (2012)).[13]

[12]Defined as the immigrants group's status and wealth.
[13]This results are confirmed by Table 3 and contradict the *contact hypothesis*.

2. Second, the bias is slightly decreasing over time (Table A4) but given the very high response rate, already after one week, it is unlikely to wear out. The decision of a bureaucrat to answer seems to be partly independent from the decision of *when* to do so.

III.B DO SCHOOLS RESPOND LESS TO FOREIGNERS? HETEROGENEITY ANALYSIS.

Table 3 performs some heterogeneity analysis on the results presented in Section III.A. The coefficient of interest η come from the LPM estimation of Equation (2):

$$y(t)_{s,p,d} = \alpha + \beta T + \eta T * \left(X_s - \sum_{s=1}^{S} X_n \right) + \sum_{p=1}^{P} \gamma_p + \sum_{d=1}^{9} \kappa_d + \varepsilon(t)_{s,p,d} \qquad (2)$$

The analysis is performed on the following continuous variables X: number of students of the school; population; percentage of foreign residents; taxable income. For the first three variables I find no statistically signficant effect.

By contrast, taxable income (at the municipality level) positively reduces bias: that is, locations with above-the-mean income act on average less discriminatory. Specifically, a one-standard deviation increase in average municipal income, approximately 4'500€, reduces the bias in one-week response rates against foreigners to 1.6pp (1.96% of the control group mean). In other words 7000€, a 1.5 SD, are enough to completely eliminate ethnic bias. However, while the dataset comprises 2'807 municipalities, only 40 of these have incomes more than 1.5 SD above the mean. Estimates at the tail of the distribution are likely to be unreliable. Moreover, the distribution of income is left-skewed. If anything, these estimates show that most Italian municipalities should show higher bias than the mean one. F.e., while only 40 municipalities have an income 1.5SD above the median, there exists 268 locations that are more than 1.5SD below the median. Given the model estimates these should show a bias of more than 10% with respect to the control group mean.

These results seem incompatible with *contact hypothesis* as mentioned in Section III (Pettigrew and Tropp (2006)). This theory would imply that population and the share of foreigners residing in a municipality would be the main drivers of bias. If this effect is present it is not detectable at any standard level of significance. Moreover, the effect of income is unlikely to be reflecting local political preferences. Although I do not test it directly, left-leaning is known to be correlated with population: if this is the case column (2) of Table 3 should at least show a positive coefficient while the point estimate is basically zero. Furthermore, population estimates come from the Italian Bureau of Statistics which is in charge of the Census and are, thus, unlikely to be noisy.

These schools are controlled by the central government and thus should differ only slightly in the amount of resources they receive, at least within provinces. A plausible explanation is that a substitution effect is in place: as I argued in the Introduction, the more tasks a bureaucrat is required to do the likelier he is to show bias, in the case his distaste for foreigners is increasing in the cost of resources. Rich municipalities may shift some tasks on the local authorities, reducing resources' costs (f.e., the time and effort required to bureaucrats) and making clerks less likely to enact discrimination (f.e., richest municipalities are more likely to have well-functioning social services and thus reduce the burden the school bears in the education of students from disadvantaged backgrounds). This result is consistent both with bureaucracy theory developed by Brodkin (1997) and taste-based discrimination as proposed by Becker (1971).

Table 3:
Treatment Effects on School Responses Rate (after one week)
Heterogeneity Analysis

		(1)	(2)	(3)	(4)
Any Treatment		-0.043 **	-0.042 **	-0.041 **	-0.042 **
		(0.010)	(0.010)	(0.010)	(0.010)
Interaction with:	# Students (by 100)	0.001			
		(0.003)			
	Population (by 10'000)		0.000		
			(0.000)		
	% Foreigners			0.312	
				(0.206)	
	Taxable Income (by 1000€)				0.006 *
					(0.002)
Covariate Mean		8.094	22.02	0.088	22.41
Outcome Mean (Control)		0.809	0.809	0.809	0.809
Sample Size		7120	7120	7120	7120
Strata FE		Y	Y	Y	Y

Notes: The table presents estimates for heterogenity on number of students, population, fraction of foreign residents and average taxable income. Population, foreign residents and average taxable income are measured at the municipality level. Interaction variables have been demeaned. The dependent variable is a dummy equalling 1 if the school answered the email within one week. All regressions include FE at the province level (stratum of randomization) and day-of-sending level. Standard errors, in parenthesis, are heteroskedastic robust since the randomization is at school level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.10$

III.C DO SCHOOL RESPOND WORSE TO FOREIGNERS? A TEXT ANALYSIS.

Figures 3 and 4[14] display the geographic distribution of usefulness and politeness, as defined in Section II.D. This preliminary evidence shows two stylized facts:

1. In Figure 3 the distribution of usefulness between control and treatment is almost opposite: regions in the South-East have the most useful answers for mails signed by Italian names while regions in the North-West display the highest usefulness for foreign names.

[14]Figures A4 and A5 show the regional averages for the overall sample.

2. In Figure 4 the distribution of politeness is very similar across control and treatment groups.

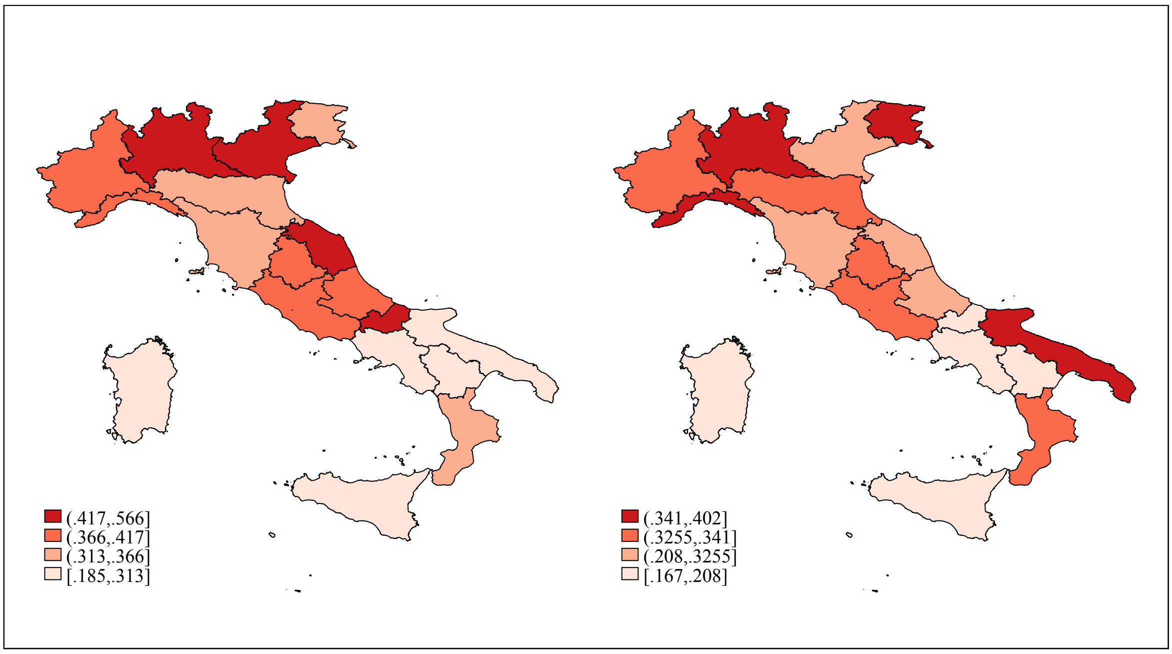


Figure 3: Average Usefulness by Region: Control (left); Treatment (right).

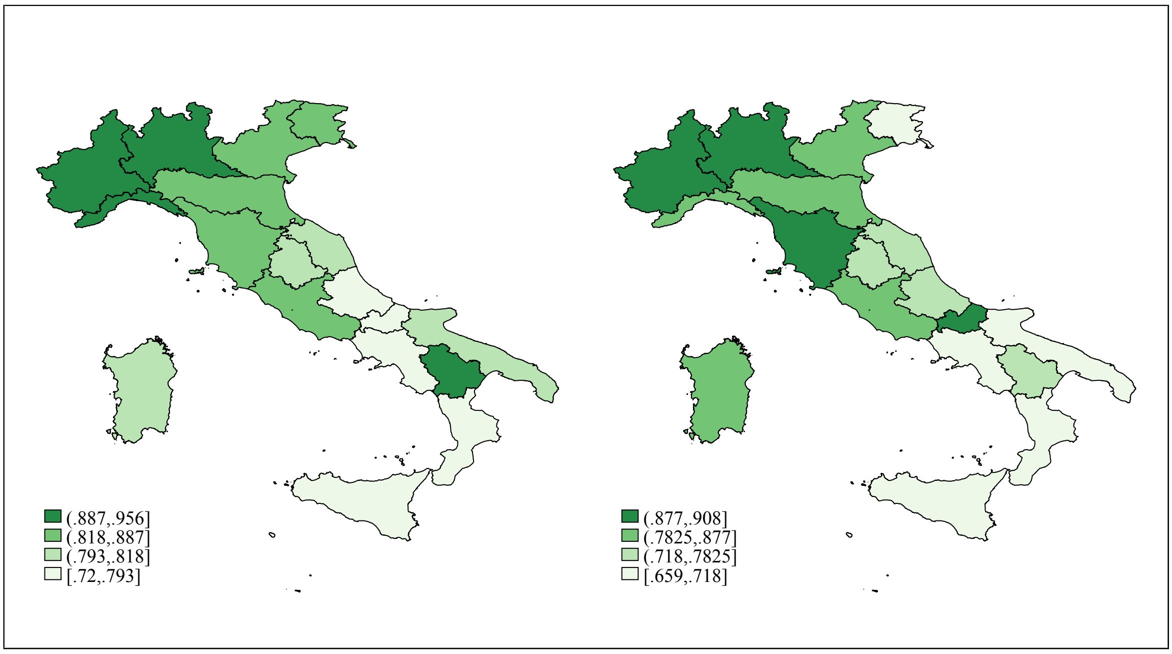


Figure 4: Average Politeness by Region: Control (left); Treatment (right).

Thus, usefulness seems to follow complex geographical patterns: some regions may show higher discrimination than what I find in Table 4.[15] On the opposite, the distribution of politeness across regions seems to indicate a different comprehension of bureaucratic

[15]Moving from an ATE to a LATE framework may highlight this geographic differences.

politeness across the Italian territory. It seems to be the case that for politeness the ATE of Figure 4 are good approximations of LATE across Italy.[16]

Table 4 shows the estimates of $\beta_{1,2}$ in the following versions of Equation 1:

$$y(t)_{s,p,d} = \alpha + \sum_{i\in\{1,2\}} \beta_i T_i + \sum_{p=1}^{P} \gamma_p + \sum_{d=1}^{9} \kappa_d + X'\eta + \varepsilon(t)_{s,p,d} \tag{3}$$

where $y(t)_{s,p,d}$ is a dummy that equals one if the response email is classified by the NLP model (see Section II.D) as useful (columns (1)-(2)) or polite (columns (3)-(4)).

Table 4:
Treatment Effects on BERT-assigned Usefulness/Kindness Labels
Text Analysis

	Usefulness (0/1)		Politeness (0/1)	
	(1)	(2)	(3)	(4)
Any Treatment	-0.060 **		-0.039 **	
	(0.012)		(0.011)	
French Treatment (T1)		-0.037 *		-0.026 *
		(0.015)		(0.013)
Arabic Treatment (T2)		-0.083 **		-0.053 **
		(0.015)		(0.013)
T1 = T2 = 0 (p-value)		0.000		0.000
T1 = T2 (p-value)		0.007		0.087
Outcome Mean (Control)	0.272	0.272	0.818	0.818
Sample Size	5107	5107	5107	5107
Strata FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y

Notes: Columns (1) and (2) present treatment effects on a dummy equalling 1 if the response email was flagged as useful by the NLP model BERT. Columns (3) and (4) repeat the experiment but classifying emails as polite or not polite. Supervised learning was performed through Google's NLP model BERT in its Italian version. 500 emails, roughly 10% of the sample, were manually flagged as useful/kind or not. The model was then trained on those and used to classifiy the remaning 5107 emails. All regressions include FE at the province level (stratum of randomization) and day-of-sending level. Standard errors, in parenthesis, are heteroskedastic robust since the randomization is at school level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.10$

The mean usefulness is 34% in the control group indicating that most of the response emails are not answering the enrollment questions. I find effect of discrimination: response emails to foreigners are 5.9pp less likely to be classified as useful (column (1) of Table 4). The discriminatory effect translates to 17.6% of the control mean: a sizable and highly significant effect. By treatment, this corresponds to 9.85% for French names and 25.1% for Arabic ones.

Column (3) of Table 4 shows that the ATE for both treatments jointly is 0.039pp, 4.77% of

[16] Also recall from Equation 1 that every specification is *within province*.

the control group mean. Although the estimates are significant and sizable for both treatments, Arabic names show approximately double the effect than French names (6.48% of the control group mean against 3.18%, respectively). However, a joint test of significance is able to reject the null that the treatment effect is the same for Arabic and French names only at the 10% level.

The coefficients in Table 4 suggest that discrimination follows a complex pattern: bureaucrats are less likely to answer to foreigners and, additionally, when they do they send less informative and polite emails with very large effects for Arabic names. This documents an important discriminatory practice. The treatment email in Section II.C was specific in asking enrollment questions: as long as (i) Arabic names are 25.1% less likely to receive information other than the opening hours of the administrative office and (ii) immigrants from Arabic-speaking countries face higher costs in gathering information (f.e. because they are on average poorer or less educated) this discriminatory practice may have even a higher effect than the differential response rate (see Table 2). Furthermore, although discrimination in politeness may be perceived as less important, it shows these practices are deep-rooted in the bureaucrats' behavior. While retrieving and condensing information to answer the questions posed in the treatment email is a costly process, politeness is not. Bureaucrats should be indifferent to (non-)politeness. The fact they are not shows they either perceive a benefit in being non-polite to foreigners or a cost in being polite. This effect, once again, hints at the fact that clerks dislike inter-group contact (Becker (1971)).

IV. ADDRESSING LIMITATIONS

This section addresses possible concerns to either the general framework of the study (Section IV.A) or its specific implementation (Section IV.B).

IV.A CRITICISM: GENERAL FRAMEWORK

As regards the general framework, the study treated every Italian public school that had an administrative office. Consequently, the internal validity is very high: I basically had access to the universe of Italian public schools. However, these results are not easily generalizable to other schooling systems for various reasons:

1. A different balance between public and private schools.

2. A different legal framework on the enrollment of foreigners.

3. A different cultural perception of how socially acceptable non-politeness is.

However, I believe this paper can document a wide and generalized phenomenon. Given the recent insurgence of far-right movements all over the world, discrimination and xenophobia are not confined to the bureaucratic arm of the Italian schools. Whereas the magnitude of the effects I find is specific to my setting, the presence of such effects is widely generalizable. For example, such magnitude resembles closely what White, Nathan, and Faller (2015) find in the context of information on ID-Laws provided by U.S. bureaucrats to (non-) Latino citizens.

IV.B CRITICISM: SPECIFICS

As regards the specific implementation of the experiment and the subsequent analysis, three main criticisms may be drawn:

1. Differential response rates may not reflect actual discrimination but rather bureaucrats recognizing emails as spam or other phishing practices.
2. The choice of a Linear Probability Model may be suboptimal or even bias the results.
3. The results of the text analysis do not reflect actual discrimination since they are endogenous: I manually flagged the training sample and, in any case, schools auto-select in answering emails.

As regards (1.), there is no way to prove the reasons why school clerks did answer differentially between groups. However, both *a priori* and *a posteriori* reasoning suggest that the perception of the email as fraudulent is not a reasonable cause. *A priori*, the emails were sent to schools' official email addresses, contained no external links or questions unrelated to schools, and the treatment emails were personalized in that each email specified the municipality in which the school is. *A posteriori*, no results suggest that this is the case: larger municipalities should have more contact with both French- and Arabic-named people, and thus heterogeneity analysis should show a positive effect of population on the discriminatory effect. If clerks perceived the emails as phishing because they had never received an email from a French account, for example, they should answer more in larger municipalities where there are more French-named people. This is, instead, not the case. Moreover, there is a consistent French-speaking population in Italy, which makes the assumption that they could move around Italy likely.

As regards (2.), a Linear Probability Model, given the binary outcomes, may obviously be a bad approximation of the true underlying Conditional Expectation Function (CEF). However, under standard assumptions (Bruce Hansen (2022)), an LPM provides the best linear approximation to the CEF, with all the known benefits of linearity. For robustness, Table A2 replicates the estimates of Table 2 by employing a LOGIT model. The estimated Average Marginal Effects in Table A2 are essentially the same as the Average Treatment Effects measured by the LPM in Table 2, suggesting that linearity is a good approximation of the underlying CEF and mitigating concerns about the choice of an LPM.

As regards (3.), I addressed the potential endogeneity of the manual flagging by imposing strict criteria that emails had to follow to be classified as (non-)useful or (non-)polite. These criteria are openly stated in Section II.D. Obviously, schools auto-select into answering, and thus the assignment of treatment is not exogenous with respect to the outcome variable. As a consequence, coefficients in Table 4 do not reflect the causal effect of treatment on the usefulness or politeness of response emails. In spite of this, I argue that if anything they are biased towards zero and thus underestimate what would be the actual, in-population, discriminatory practice. Indeed, for the coefficients to be biased away from zero, the schools auto-selecting out of the sample (i.e., schools not answering emails by foreign names) should have sent more useful and polite emails than the rest of the sample if they had answered. This circumstance appears unlikely.

V. CONCLUSIONS

Theories of bureaucracy argue that when clerks are endowed with discretion and resources they enact discrimination, absent strict procedures to prevent them (Brodkin (1997)). Two reasons are possible for this behavior (White, Nathan, and Faller (2015)): statistical discrimination or taste-based discrimination. Moreover, theorists of social psychology have argued that: (i) discrimination depends of how agreeable and capable an outer group is, depending on the stereotypes that have developed against it (Esses (2021)); (ii) discrimination and prejudice are reduced by positive interactions with members of other groups and these interactions are more relevant than negative ones, the *contact hypothesis* (Pettigrew and Tropp (2006)).

Recent evidence has shown that foreigners undergo a process of *educational segregation* in the Italian school system (Carlana, La Ferrara, and Pinotti (2022)) but has not identified whether this may be the results of bureaucratic discrimination.

I employ a Randomized Controlled Trial to test whether bureaucratic discrimination is present during the enrollment procedures to public schools, something which is strictly prohibited by the Italian law, and reconcile the results with the aforementioned theories of discrimination and social contact. Practically, each public school in Italy receives an email from a fictional parent inquiring about enrollment procedure and welfare for poorer families. Within province, each email is signed by either an Italian, French or Arabic sounding name.

I find that schools discriminate foreigners: non-Italian aliases are 5.32% less likely to receive an answer from schools, against a control group mean of 81%. This discriminatory practices are independent of how big schools are, population size of the schools' municipalities or percentage of foreign residents. Ceteris paribus, poorer municipalities show larger biases with a relevant fraction, almost 10%, of municipalities that are more than 1.5SD left of the mean and thus exhibit more than double the average bias. This effect is not counterbalanced by richer municipalities since the income distribution within the sample is skewed to the left. Additionally, when schools respond they are 17.6% less likely to provide useful information to the inquire and 4.77% less likely to be polite, when the sender has a foreign alias (with respect to the control mean of 35% and 82% respectively). While French aliases are more likely to not receive an answer, Arabic aliases are more likely to receive uninformative and unpolite answers.

These results argue in favor of taste based discrimination. Politeness has little to no cost for a clerk, whereas un-politeness has no immediate practical return. This implies that bureaucrats may gain utility from rudeness towards foreigners, consistent with the argument posed by theories of taste-based discrimination. Moreover, statistical discrimination is not likely to be at play: French citizens are on average from better socioeconomic backgrounds than Italian ones and thus, as long as this background is correlated with how good a student may be, schools should actually discriminate *against* Italians to get French students.

Moreover, these results contradict the contact hypothesis: neither the number of students, the population of the municipality where the school is located, or its percentage of foreign residents has any effect on the size of the bias. On the contrary, the theory developed by Lee and Fiske (2006) is able to reconcile all the results, in conjunction with taste-based discrimination. If French are perceived as a group of foreigners with low likability but high socioeconomic status while Arabs are perceived as low likability and, also, low socioeconomic status the difference in the feelings these group cause may explain differences in bias (f.e. enviousness against anger, see Esses (2021)). In addition, taste-based discrimination can explain why only income per capita affects bias: I argue that poorer municipalities are able to shift less tasks onto local authorities (f.e. social services) and thus require clerks to perform more duties. Effort and time are scarce resources and thus their cost opportunity is increasing in the number of chores a clerk has to perform. Under simple reasoning, a clerk allocates one unit of time to a foreigner when the marginal gain of doing so is the same as allocating one unit to an Italian citizen and as the price of its time and effort. If, given the scarcity of resources, its effort's cost-opportunity is higher he is less likely to allocate time to anyone but also less likely to allocate time to a foreigner with respect to an Italian.

REFERENCES

Agan, Amanda and Sonja Starr (2018). "Ban the box, criminal records, and racial discrimination: A field experiment". In: *The Quarterly Journal of Economics* 133.1, pp. 191–235.

Becker, GS (1971). *The Economics of Discrimination 2nd ed (University Of Chicago Press, Chicago)*.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014a). "High-dimensional methods and inference on structural and treatment effects". In: *Journal of Economic Perspectives* 28.2, pp. 29–50.

— (2014b). "Inference on treatment effects after selection among high-dimensional controls". In: *Review of Economic Studies* 81.2, pp. 608–650.

Bertrand, Marianne and Sendhil Mullainathan (2004). "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination". In: *American economic review* 94.4, pp. 991–1013.

Biroli, Pietro et al. (2024). "Talent is Everywhere, Opportunity is Not: Online Role Model Mentoring and Students' Aspirations". In.

Brodkin, Evelyn Z (1997). "Inside the welfare contract: Discretion and accountability in state welfare administration". In: *Social Service Review* 71.1, pp. 1–33.

Brunello, Giorgio and Daniele Checchi (2007). "Does school tracking affect equality of opportunity? New international evidence". In: *Economic policy* 22.52, pp. 782–861.

Carlana, Michela, Eliana La Ferrara, and Paolo Pinotti (2022). "Goals and gaps: Educational careers of immigrant children". In: *Econometrica* 90.1, pp. 1–29.

Charles, Kerwin Kofi and Jonathan Guryan (2009). "Taste-based discrimination". In: *The New Palgrave Dictionary of Economics*. Springer, pp. 1–8.

Doleac, Jennifer L and Benjamin Hansen (2020). "The unintended consequences of "ban the box": Statistical discrimination and employment outcomes when criminal histories are hidden". In: *Journal of Labor Economics* 38.2, pp. 321–374.

Esses, Victoria M (2021). "Prejudice and discrimination toward immigrants". In: *Annual review of psychology* 72.1, pp. 503–531.

Fiske, Susan T and Tiane L Lee (2012). "Xenophobia and how to fight it: Immigrants as the quintessential "other"." In: *Social Categories in Everyday Experience*.

Fryer Jr, Roland G and Steven D Levitt (2004). "Understanding the black-white test score gap in the first two years of school". In: *Review of economics and statistics* 86.2, pp. 447–464.

Guyon, Nina, Eric Maurin, and Sandra McNally (2012). "The effect of tracking students by ability into different schools: A natural experiment". In: *Journal of Human resources* 47.3, pp. 684–721.

Hansen, Bruce (2022). *Econometrics*. Princeton University Press.

Lee, Tiane L and Susan T Fiske (2006). "Not an outgroup, not yet an ingroup: Immigrants in the stereotype content model". In: *International Journal of Intercultural Relations* 30.6, pp. 751–768.

Pettigrew, Thomas F and Linda R Tropp (2006). "A meta-analytic test of intergroup contact theory." In: *Journal of personality and social psychology* 90.5, p. 751.

White, Ariel R, Noah L Nathan, and Julie K Faller (2015). "What do I need to vote? Bureaucratic discretion and discrimination by local election officials". In: *American Political Science Review* 109.1, pp. 129–142.

VI. APPENDIX - ADDITIONAL FIGURES & TABLES

VI.A ADDITIONAL TABLES

Table A1:
Evaluation Metrics of BERT Training
Appendix

	Accuracy	F1	Precision	Recall
	(1)	(2)	(3)	(4)
Usefulness:	0.680	0.675	0.684	0.680
Politeness:	0.88	0.874	0.883	0.88

Notes: The table reports evaluation metrics for the training of the NLP model BERT. Metrics are only report for epoch 4, the last epoch of training.

Table A2:
Treatment Effects on School Responses Rate (after one week)
Appendix: LOGIT

	(1)		(2)		(3)		(4)	
Any Treatment	-0.043	**	-0.042	**				
	(0.009)		(0.009)					
French Treatment (T1)					-0.060	**	-0.060	**
					(0.012)		(0.012)	
Arabic Treatment (T2)					-0.025	*	-0.024	*
					(0.012)		(0.012)	
T1 = T2 = 0 (p-value)					0.000		0.000	
T1 = T2 (p-value)					0.011		0.010	
Outcome Mean (Control)	0.809		0.809		0.809		0.809	
Sample Size	7120		7120		7120		7120	
Strata FE	Y		Y		Y		Y	
Controls	N		Y		N		Y	

Notes: The table replicates the main results of Table 2 by employing a logit model. Coefficients have been trasformed to Average Marginal Effects (AME) to guarantee interpretability.The dependent variable is a dummy equalling 1 if the school answered the email within one week. Columns (2) and (4) add as controls: school number of students; average taxable earnings (municipality level); fraction of foreigners (municipality level); population (municiplaity level). All regressions include FE at the province level (stratum of randomization) and day-of-sending level. Standard errors, in parenthesis, are heteroskedastic robust since the randomization is at school level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.10$

Table A3:
Treatment Effects on School Responses Rate (after one week)
Appendix: Post-Double LASSO

	(1)		(2)	
Any Treatment	-0.046	**		
	(0.013)			
French Treatment (T1)			-0.059	**
			(0.017)	
Arabic Treatment (T2)			-0.033	*
			(0.016)	
T1 = T2 = 0 (p-value)			0.001	
T1 = T2 (p-value)			0.183	
Outcome Mean (Control)	0.809		0.809	
Sample Size	7120		7120	
Strata FE	Y		Y	

Notes: The table replicates the main results of Table 2 by employing a post-double LASSO model, following Belloni, Chernozhukov and Hansen (2014b). The dependent variable is a dummy equalling 1 if the school answered the email within one week. All regressions include FE at the province level (stratum of randomization) and day-of-sending level. Standard errors, in parenthesis, are heteroskedastic robust since the randomization is at school level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.10$

Table A4:
Treatment Effects on School Responses Rate (after one day)
Appendix: 1-day response rates

	(1)		(2)		(3)		(4)	
Any Treatment	-0.055	**	-0.054	**				
	(0.011)		(0.011)					
French Treatment (T1)					-0.078	**	-0.077	**
					(0.014)		(0.014)	
Arabic Treatment (T2)					-0.032	*	-0.031	*
					(0.014)		(0.014)	
T1 = T2 = 0 (p-value)					0.000		0.000	
T1 = T2 (p-value)					0.011		0.010	
Outcome Mean (Control)	0.688		0.688		0.688		0.688	
Sample Size	7120		7120		7120		7120	
Strata FE	Y		Y		Y		Y	
Controls	N		Y		N		Y	

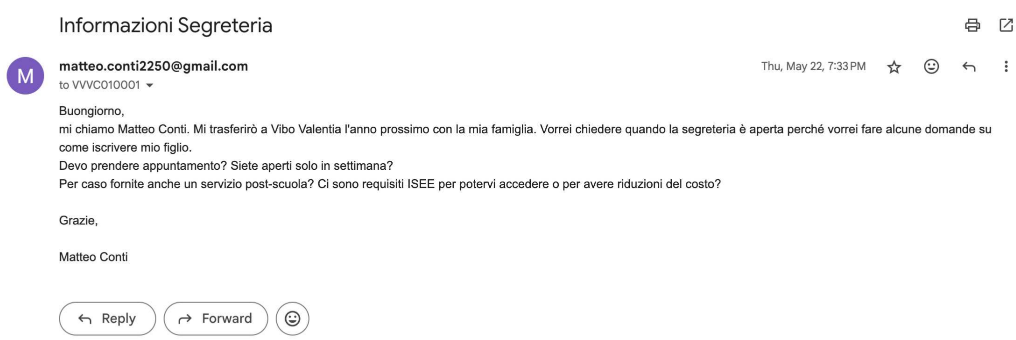
Notes: The table reports the estimated coefficients and standard errors for the main equation.The dependent variable is a dummy equalling 1 if the school answered the email within one day. Columns (2) and (4) add as controls: school number of students; average taxable earnings (municipality level); fraction of foreigners (municipality level); population (municiplaity level). All regressions include FE at the province level (stratum of randomization) and day-of-sending level. Standard errors, in parenthesis, are heteroskedastic robust since the randomization is at school level. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.10$

Table A5:
Treatment Effects on BERT-assigned Usefulness/Kindness Labels
Appendix: LOGIT

	Usefulness (0/1)		Politeness (0/1)	
	(1)	(2)	(3)	(4)
Any Treatment	-0.061 **		-0.040 **	
	(0.013)		(0.011)	
French Treatment (T1)		-0.033 *		-0.026 *
		(0.016)		(0.013)
Arabic Treatment (T2)		-0.089 **		-0.054 **
		(0.016)		(0.013)
T1 = T2 = 0 (p-value)		0.000		0.000
T1 = T2 (p-value)		0.002		0.079
Outcome Mean (Contrc	0.347	0.347	0.818	0.818
Sample Size	5107	5107	5107	5107
Strata FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y

Notes: The table replicates the text analysis results of Table 4, using a LOGIT model. Coefficients have been trasformed to Average Marginal Effects (AME) to guarantee interpretability. Columns (1) and (2) present treatment effects on a dummy equalling 1 if the response email was flagged as useful by the NLP model BERT. Columns (3) and (4) repeat the experiment but classifying emails as polite or not polite. Supervised learning was performed through Google's NLP model BERT in its Italian version. 500 emails, roughly 10% of the sample, were manually flagged as useful/kind or not. The model was then trained on those and used to classifiy the remaning 5107 emails. The dependent variable is a dummy equalling 1 if the school answered the email within one week. All regressions include FE at the province level (stratum of randomization) and day-of-sending level. Standard errors, in parenthesis, are heteroskedastic robust since the randomization is at school level. ** p < 0.01, * p < 0.05, ~ p < 0.10

VI.B ADDITIONAL FIGURES


Informazioni Segreteria

matteo.conti2250@gmail.com
to VVVC010001
Thu, May 22, 7:33 PM

Buongiorno,
mi chiamo Matteo Conti. Mi trasferirò a Vibo Valentia l'anno prossimo con la mia famiglia. Vorrei chiedere quando la segreteria è aperta perché vorrei fare alcune domande su come iscrivere mio figlio.
Devo prendere appuntamento? Siete aperti solo in settimana?
Per caso fornite anche un servizio post-scuola? Ci sono requisiti ISEE per potervi accedere o per avere riduzioni del costo?

Grazie,

Matteo Conti

Reply Forward

Figure A1: Example of Email from Italian Account.

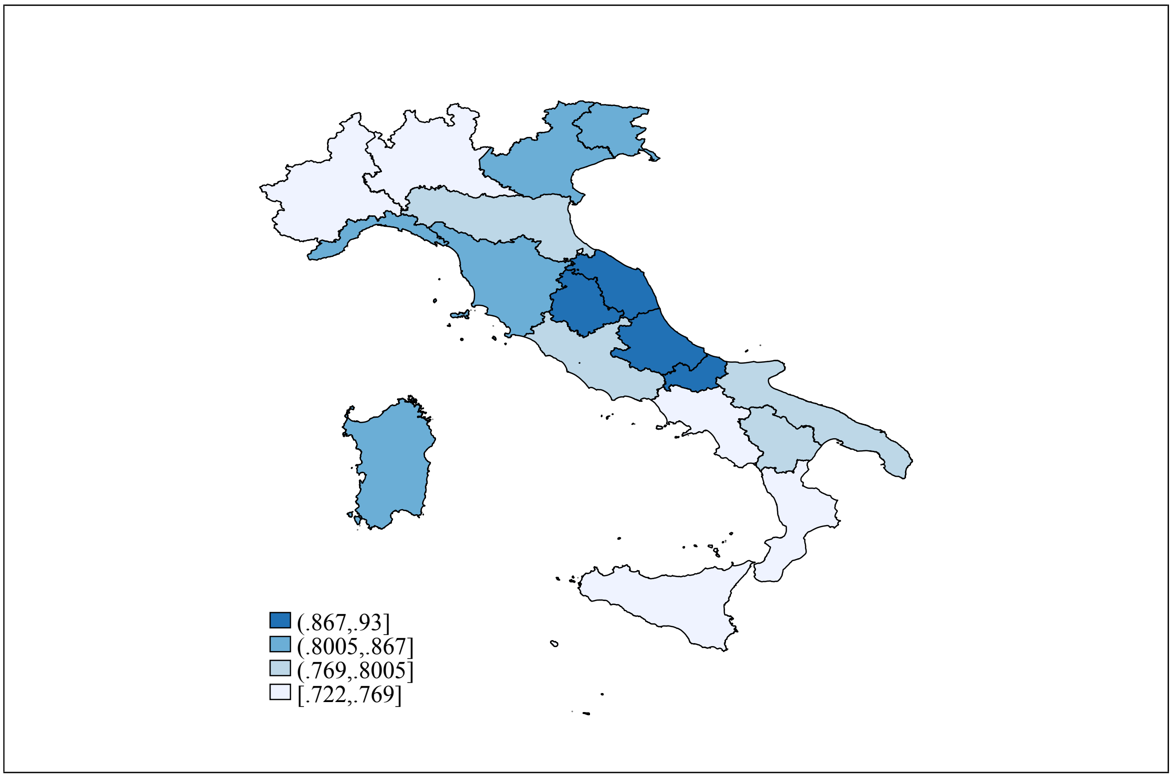


Figure A2: Average Response Rate by Region: Whole Sample.

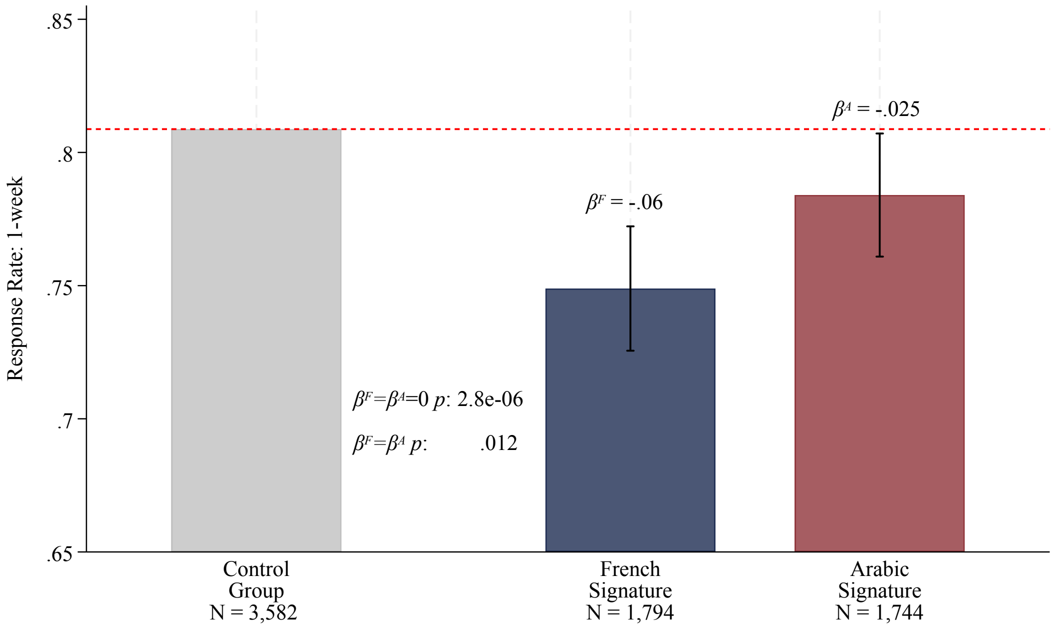


Figure A3: Treatment Effects, 1 Week Timespan.

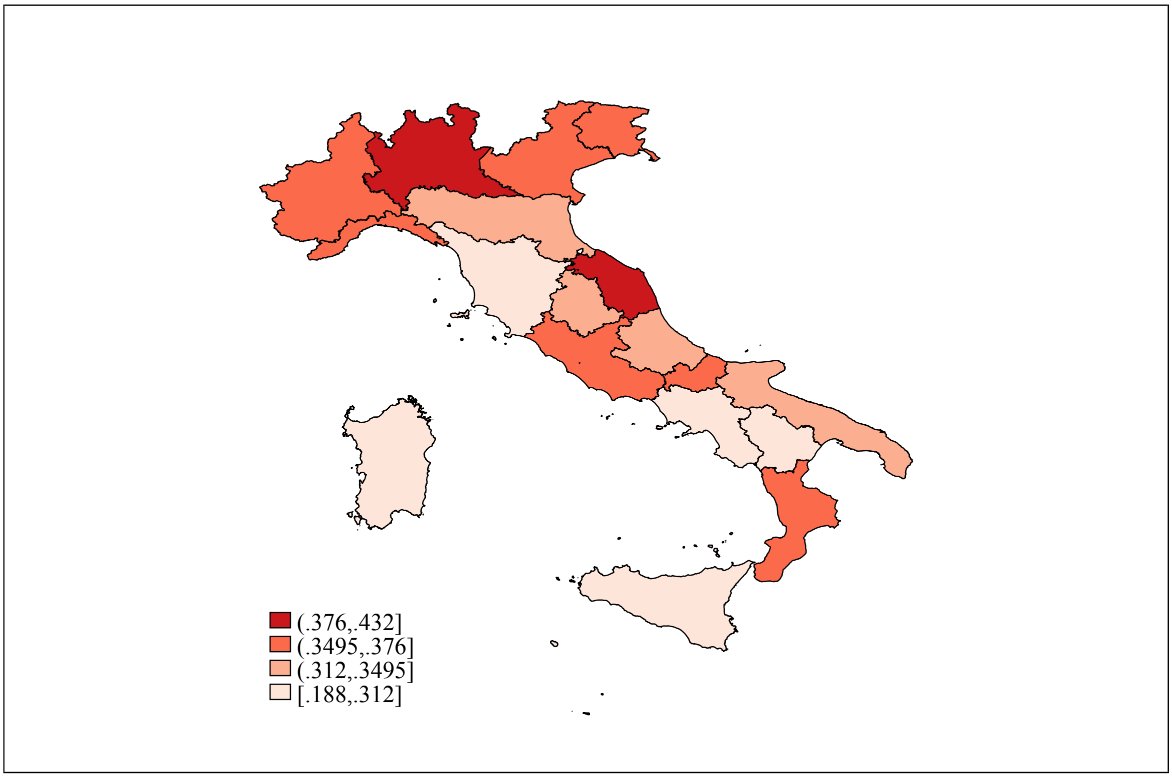


Figure A4: Average Usefulness by Region: Whole Sample.

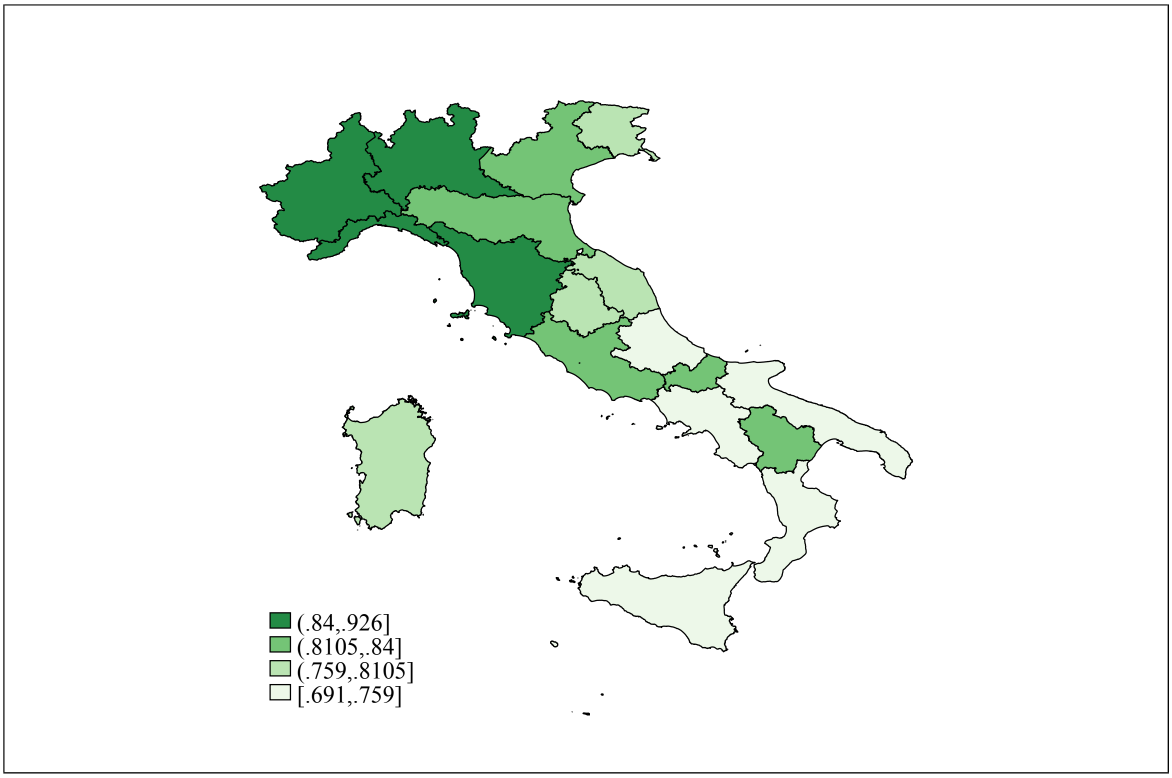


Figure A5: Average Politeness by Region: Whole Sample.

VI.C LIST OF EMAIL NAMES

Italian names:

1. Matteo Conti (matteo.conti2250@gmail.com)
2. Giovanni Ferrari (giovanni.ferrari873@gmail.com)
3. Luca Moretti (luca.moretti873@gmail.com)
4. Marco Russo (marco.russo2250@gmail.com)

French names:

1. Antoine Lambert (antoine.lambert2250@gmail.com)
2. Jeanclaude Martin (jeanclaude.martin873@gmail.com)

Arabic names:

1. Omar El Haddad (omar.elhaddad873@gmail.com)
2. Youssef Nasser (youssef.nasser2250@gmail.com)