Do School Bureaucrats Discriminate Against Foreigners? Experimental Evidence from Italy

Alessio Giorgetta
alessio.giorgetta@studio.unibo.it

Alma Mater Studiorum - University of Bologna

July, 2025

Introduction: An Annoying Email

Imagine:

Introduction: An Annoying Email

Imagine:

- You are a clerk in a high-school of Bologna

Introduction: An Annoying Email

Imagine:

- You are a clerk in a high-school of Bologna
- You receive the following email

Introduction: An Annoying Email

Information

school.bologna@istruzione.it

Information

Good morning,
my name is #NAME#. I will move to Bologna next year with my family. I would like to ask when the administrative office is open because I'd 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 offer afternoon activities? Are there income requirements to join them or to have reductions of the cost?
Thank you,
#NAME#

Notes: Example of a treatment email (translated).

Introduction: An Annoying Email

Would you be more likely to answer if the sender was:

Introduction: An Annoying Email

Would you be more likely to answer if the sender was:

- **Matteo Conti**

Introduction: An Annoying Email

Would you be more likely to answer if the sender was:

- **Matteo Conti**
- **Antoine Lambert**

Introduction: An Annoying Email

Would you be more likely to answer if the sender was:

- **Matteo Conti**
- **Antoine Lambert**
- **Omar El Haddad**

Introduction: An Annoying Email

Would you be more likely to answer if the sender was:

- **Matteo Conti**
- **Antoine Lambert**
- **Omar El Haddad**

→*And would you be informative and polite in the same way?*

Outline of the Talk: 5W

We will follow the 5Ws of journalists:

Outline of the Talk: 5W

We will follow the 5Ws of journalists:

- **WHO & HOW:** an email RCT with Italian school bureaucrats

Outline of the Talk: 5W

We will follow the 5Ws of journalists:

- **WHO & HOW:** an email RCT with Italian school bureaucrats
- **WHAT:** they discriminate against foreigners
  1. They respond less
  2. They respond less informatively/politely

Outline of the Talk: 5W

We will follow the 5Ws of journalists:

- **WHO & HOW:** an email RCT with Italian school bureaucrats
- **WHAT:** they discriminate against foreigners
  1. They respond less
  2. They respond less informatively/politely
- **WHEN:** heterogenity by covariates

Outline of the Talk: 5W

We will follow the 5Ws of journalists:

- **WHO & HOW:** an email RCT with Italian school bureaucrats
- **WHAT:** they discriminate against foreigners
  1. They respond less
  2. They respond less informatively/politely
- **WHEN:** heterogenity by covariates
- **WHERE:** heterogenity by region

Outline of the Talk: 5W

We will follow the 5Ws of journalists:

- **WHO & HOW:** an email RCT with Italian school bureaucrats
- **WHAT:** they discriminate against foreigners
  1. They respond less
  2. They respond less informatively/politely
- **WHEN:** heterogenity by covariates
- **WHERE:** heterogenity by region
- **WHY:** reconciling with theory

How to Measure Discrimination? RCTs

- Data Collection:
  1. Collect all emails of Italian public schools ($\approx 50,000$)
     (but no Trentino-Alto Adige & Valle d'Aosta)
  2. + school characteritics and locations (MIUR)
  3. collect population data (ISTAT) & economic data (MEF)

How to Measure Discrimination? RCTs

- Data Collection:
  1. Collect all emails of Italian public schools ($\approx 50,000$) (but no Trentino-Alto Adige & Valle d'Aosta)
  2. + school characteritics and locations (MIUR)
  3. collect population data (ISTAT) & economic data (MEF)
- Data Cleaning:
  1. drop unusual schools (within prisons, Slovenian speaking ect)
  2. drop schools with no email address or no location
  3. aggregate schools that share an administrative office
     → final dataset comprises 7120 schools

How to Measure Discrimination? RCTs

- Data Collection:
  1. Collect all emails of Italian public schools ($\approx 50,000$)
     (but no Trentino-Alto Adige & Valle d'Aosta)
  2. + school characteritics and locations (MIUR)
  3. collect population data (ISTAT) & economic data (MEF)
- Data Cleaning:
  1. drop unusual schools (within prisons, Slovenian speaking ect)
  2. drop schools with no email address or no location
  3. aggregate schools that share an administrative office
     → final dataset comprises 7120 schools
- Randomize the email **within province** (Table 1):
  1. 50 % Italian names
  2. 25% French names
  3. 25% Arabic names

How to Measure Discrimination? RCTs

- Data Collection:
  1. Collect all emails of Italian public schools ($\approx 50,000$) (but no Trentino-Alto Adige & Valle d'Aosta)
  2. + school characteritics and locations (MIUR)
  3. collect population data (ISTAT) & economic data (MEF)
- Data Cleaning:
  1. drop unusual schools (within prisons, Slovenian speaking ect)
  2. drop schools with no email address or no location
  3. aggregate schools that share an administrative office
     → final dataset comprises 7120 schools
- Randomize the email **within province** (Table 1):
  1. 50 % Italian names
  2. 25% French names
  3. 25% Arabic names ⇒ **Randomization is succesful**

Balance Table → Randomization is Successful

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		

How to Measure Discrimination? RCTs

- Analyze the following outcomes:

How to Measure Discrimination? RCTs

- Analyze the following outcomes:
  1. **One Week Response Rate** (Figure 1, Table 3)
     → whether school answered the email within a week

How to Measure Discrimination? RCTs

- Analyze the following outcomes:
  1. **One Week Response Rate** (Figure 1, Table 3)
     → whether school answered the email within a week
  2. **Usefulness** (Table 4, Figure 3)
     → at least info on opening hours & anything more

How to Measure Discrimination? RCTs

- Analyze the following outcomes:

  1. **One Week Response Rate** (Figure 1, Table 3)
     → whether school answered the email within a week
  2. **Usefulness** (Table 4, Figure 3)
     → at least info on opening hours & anything more
  3. **Politeness** (Table 4, Figure 4)
     → if greetings, *Lei* or *Voi* form, no capslock

Results: Clerks Respond Less To Foreigners

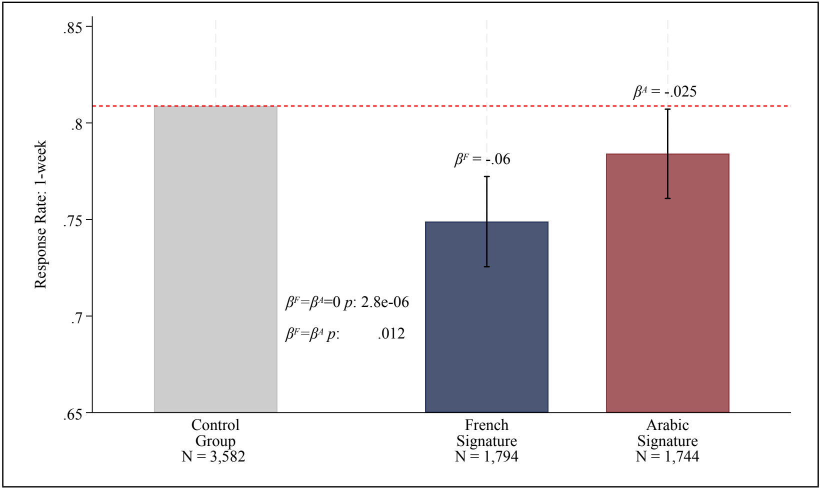


Figure 1: Treatment Effects by Treatment Arm.

Results: Poorest Discriminate More?

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	0.001			
	(by 100)	(0.003)			
	Population		0.000		
	(by 10'000)		(0.000)		
	% Foreigners			0.312	
				(0.206)	
	Taxable Income				0.006 *
	(by 1000€)				(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

Results: Clerks Are Less Useful & Polite

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

	Usefulness (0/1)		Politeness (0/1)	
	(1)	(2)	(3)	(4)
Any Treatment	-0.059 **		-0.039 **	
	(0.013)		(0.011)	
French Treatment (T1)		-0.033 *		-0.026 *
		(0.016)		(0.013)
Arabic Treatment (T2)		-0.084 **		-0.053 **
		(0.016)		(0.013)
T1 = T2 = 0 (p-value)		0.000		0.000
T1 = T2 (p-value)		0.005		0.078
Outcome Mean (Control)	0.335	0.335	0.818	0.818
Sample Size	5107	5107	5107	5107
Strata FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y

Results: Large Heterogeneity in Bias

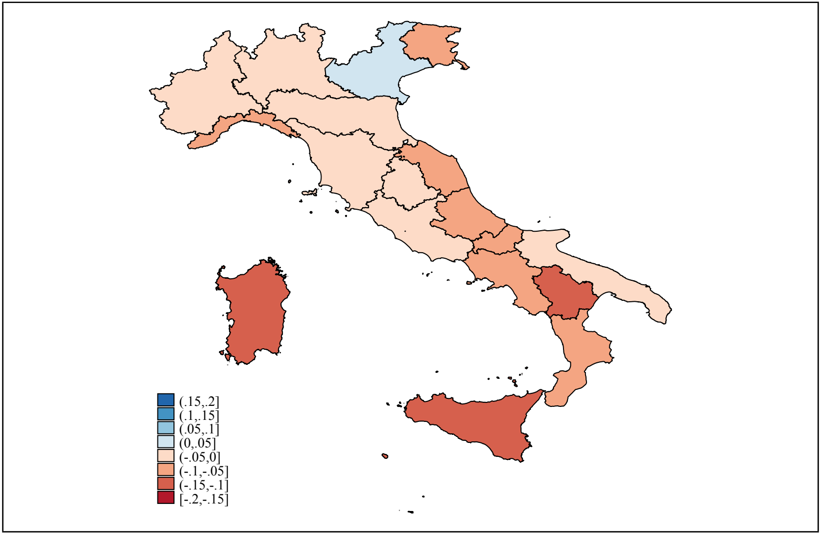


Figure 2: CATE by Regions, 1-week Response Rates.

Large Heterogeneity in Bias → from -12pp to +2pp

Table 5:
Treatment Effects by Region
Text Analysis: Heterogeneity Analysis

	Response Rates (One Week)								
	Any Treatment		French (T1)		Arabic (T2)		SE of (1), (2) & (3)		
	(1)		(2)		(3)		(4)	(5)	(6)
Abruzzo	-0.071		-0.039		-0.094		(0.051)	(0.063)	(0.068)
Basilicata	-0.133		-0.157		-0.111		(0.094)	(0.120)	(0.119)
Calabria	-0.055		-0.070		-0.048		(0.054)	(0.067)	(0.068)
Campania	-0.055	~	-0.086	*	-0.019		(0.032)	(0.040)	(0.038)
ER	-0.029		-0.050		-0.002		(0.036)	(0.046)	(0.045)
FVG	-0.061		0.021		-0.150	~	(0.064)	(0.067)	(0.090)
Lazio	-0.035		-0.019		-0.044		(0.031)	(0.038)	(0.040)
Liguria	-0.094	~	-0.099		-0.103		(0.056)	(0.072)	(0.072)
Lombardia	-0.018		-0.046		0.011		(0.026)	(0.033)	(0.031)
Marche	-0.052		-0.067		-0.033		(0.046)	(0.061)	(0.058)
Molise	-0.064		0.044		-0.163		(0.086)	(0.081)	(0.131)
Piemonte	-0.024		-0.038		-0.009		(0.038)	(0.047)	(0.047)
Puglia	-0.041		-0.047		-0.039		(0.039)	(0.048)	(0.049)
Sardegna	-0.118	*	-0.121	~	-0.107		(0.054)	(0.068)	(0.071)
Sicilia	-0.109	**	-0.169	**	-0.060		(0.034)	(0.045)	(0.043)
Toscana	0.041		0.035		0.035		(0.038)	(0.047)	(0.048)
Umbria	-0.001		0.016		-0.014		(0.063)	(0.075)	(0.080)
Veneto	0.021		-0.033		0.078	*	(0.030)	(0.039)	(0.033)
ATE	-0.044	**	-0.061	**	-0.026	*	(0.010)	(0.012)	(0.012)
Mean (Control)	0.809	**	0.809	**	0.808	**	(0.007)	(0.007)	(0.007)
CV N.	10		10		10				
Sample Size	5107		3869		3866				

Large Heterogeneity in Bias → from -36pp to +3pp

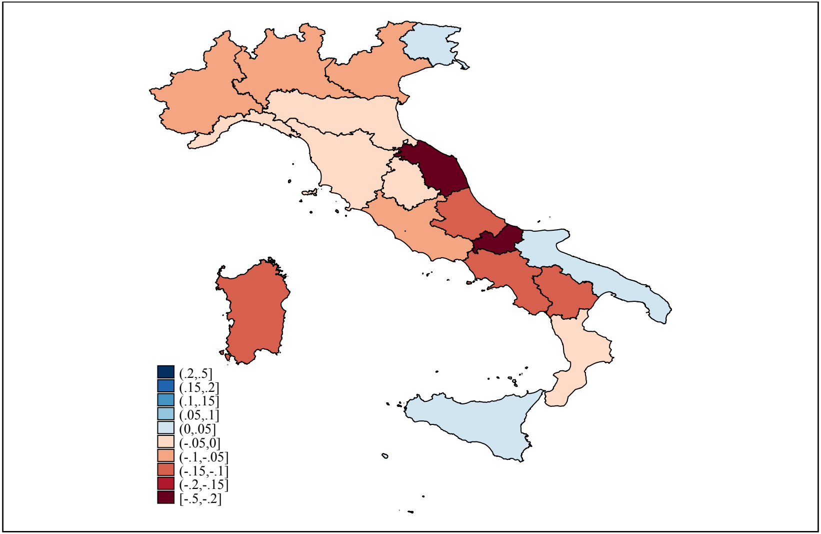


Figure 3: CATE by Regions, Usefulness (0/1).

Large Heterogeneity in Bias → from -20pp to +13pp

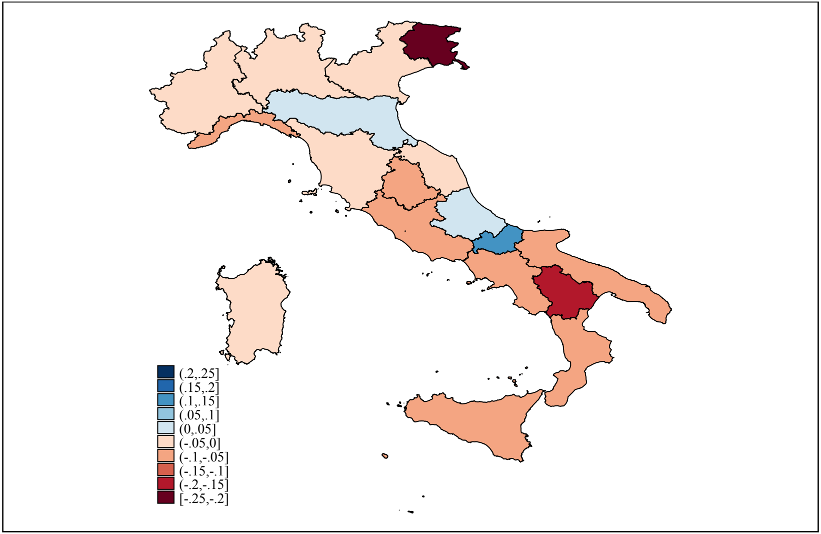


Figure 4: CATE by Regions, Politeness (0/1).

Reconciling with Theory: What Have We Found?

Brodkin (1997) has shown that **discrimination happens** when:

1. Bureaucrats have **resources**
2. They have **discretion**

Reconciling with Theory: What Have We Found?

Brodkin (1997) has shown that **discrimination happens** when:

1. Bureaucrats have **resources**
2. They have **discretion**

⇒ *This is what we know for sure.*

Reconciling with Theory: What Have We Found?

Brodkin (1997) has shown that **discrimination happens** when:

1. Bureaucrats have **resources**
2. They have **discretion**

⇒ *This is what we know for sure.*

Theories of Discrimination:

Reconciling with Theory: What Have We Found?

Brodkin (1997) has shown that **discrimination happens** when:

1. Bureaucrats have **resources**
2. They have **discretion**

⇒ *This is what we know for sure.*

Theories of Discrimination:

1. *Statistical Discrimination* vs *Taste-Based Discrimination*
   Agan and Starr (2018) vs Becker (1971)

Reconciling with Theory: What Have We Found?

Brodkin (1997) has shown that **discrimination happens** when:

1. Bureaucrats have **resources**
2. They have **discretion**

⇒ *This is what we know for sure.*

Theories of Discrimination:

1. *Statistical Discrimination* vs *Taste-Based Discrimination*
   Agan and Starr (2018) vs Becker (1971)
2. *Contact Hypothesis* vs *Likeability-Capability Approach*
   Pettigrew and Tropp (2006) vs Lee and Fiske (2006)

Theory: Results Favor of Taste-Based Discrimination

No evidence of *Statistical Discrimination* (Agan and Starr (2018)):

- French as discriminated as Arabic aliases (but differently)

Theory: Results Favor of Taste-Based Discrimination

No evidence of *Statistical Discrimination* (Agan and Starr (2018)):

- French as discriminated as Arabic aliases (but differently)

Results are in favor of *taste-based discrimination* (Becker (1971)):

- politeness costs $\mathbb{0} \rightarrow \mathbb{U}(.) \Uparrow$ from unpoliteness toward strangers
- **if** resources get scarcer, the bias increase
  $\rightarrow$ the price of allocating reources to strangers $\Uparrow$

Theory: Results for Likeability-Capability Approach

Contact Hypothesis (Pettigrew and Tropp (2006)) is not supported:

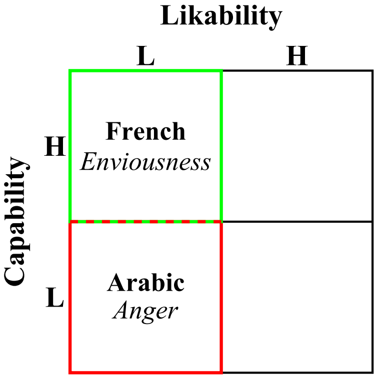
- No Effect of #Students, Population or % Foreigners (Table 3)

Theory: Results for Likeability-Capability Approach

Contact Hypothesis (Pettigrew and Tropp (2006)) is not supported:

- No Effect of #Students, Population or % Foreigners (Table 3)

Results consistent with *Likeability/Capability* (Lee and Fiske (2006)):



Bibliography I

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

Appendix Tables - CATE for Usefulness

Table 6:
Treatment Effects by Region
Text Analysis: Heterogeneity Analysis

	Usefulness (0/1)								
	Any Treatment		French (T1)		Arabic (T2)		SE of (1), (2) & (3)		
	(1)		(2)		(3)		(4)	(5)	(6)
Abruzzo	-0.126		-0.141		-0.120		(0.080)	(0.097)	(0.099)
Basilicata	-0.100		-0.044		-0.137		(0.125)	(0.153)	(0.144)
Calabria	-0.009		0.085		-0.118		(0.072)	(0.093)	(0.083)
Campania	-0.101	**	-0.106	*	-0.090	*	(0.036)	(0.045)	(0.044)
ER	-0.032		-0.007		-0.056		(0.050)	(0.062)	(0.061)
FVG	0.030		0.004		0.048		(0.100)	(0.120)	(0.127)
Lazio	-0.084	~	-0.056		-0.122	*	(0.043)	(0.054)	(0.053)
Liguria	-0.001		0.029		-0.013		(0.087)	(0.107)	(0.104)
Lombardia	-0.056		-0.062		-0.051		(0.036)	(0.045)	(0.046)
Marche	-0.277	**	-0.200	*	-0.352	**	(0.075)	(0.094)	(0.089)
Molise	-0.358	*	-0.213		-0.488	**	(0.142)	(0.184)	(0.127)
Piemonte	-0.054		0.000		-0.084		(0.051)	(0.067)	(0.061)
Puglia	0.031		0.000		0.064		(0.053)	(0.064)	(0.067)
Sardegna	-0.100		-0.067		-0.126		(0.068)	(0.085)	(0.078)
Sicilia	0.011		0.050		-0.028		(0.037)	(0.049)	(0.043)
Toscana	-0.018		0.046		-0.096		(0.052)	(0.066)	(0.062)
Umbria	-0.025		0.042		-0.119		(0.096)	(0.121)	(0.113)
Veneto	-0.093	*	-0.092		-0.105	~	(0.047)	(0.059)	(0.057)
ATE	-0.059	**	-0.036	*	-0.082	**	(0.013)	(0.016)	(0.016)
Mean (Control)	0.362	**	0.363	**	0.361	**	(0.009)	(0.009)	(0.009)
CV N.	10		10		10				
Sample Size	5107		3869		3866				

Notes: The table presents Conditional Average Treatment Effects (CATE) on the Usefulness (0/1) of the response emails. Column (1) presents the results for any treatment, Column (2) the results for French aliases, and Column (3) the results for Arabic aliases. Columns (4), (5), and (6) specifiy the SEs for the estimates in Columns (1), (2), and (3) respectively. The estimates come from a linear AIPW model (Chernozhukov et al. (2018)). The nuisance parameters are estimated through a Random Forest with the Honesty approach (Athey and Imbens (2016)). Outcome and treatment are modeled conditional on province and day-of-sending, the randomization levels. 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. ** $p < 0.01$, * $p < 0.05$, ~ p <

Appendix Tables - CATE for Politeness

Table 7:
Treatment Effects by Region
Text Analysis: Heterogeneity Analysis

	Politeness (0/1)								
	Any Treatment		French (T1)		Arabic (T2)		SE of (1), (2) & (3)		
	(1)		(2)		(3)		(4)	(5)	(6)
Abruzzo	0.034		0.065		-0.004		(0.074)	(0.087)	(0.093)
Basilicata	-0.152		-0.229		-0.076		(0.103)	(0.142)	(0.124)
Calabria	-0.081		-0.020		-0.144	~	(0.066)	(0.077)	(0.087)
Campania	-0.076	*	-0.057		-0.102	*	(0.039)	(0.049)	(0.049)
ER	0.027		0.060		0.003		(0.039)	(0.045)	(0.050)
FVG	-0.203	*	-0.173		-0.231	*	(0.086)	(0.108)	(0.115)
Lazio	-0.052		-0.028		-0.079	~	(0.034)	(0.041)	(0.046)
Liguria	-0.098	~	-0.039		-0.152	*	(0.053)	(0.059)	(0.076)
Lombardia	-0.011		-0.012		-0.014		(0.021)	(0.027)	(0.027)
Marche	-0.045		0.042		-0.129		(0.064)	(0.073)	(0.090)
Molise	0.127		0.044		0.219	*	(0.120)	(0.160)	(0.110)
Piemonte	-0.033		-0.050		-0.018		(0.029)	(0.041)	(0.036)
Puglia	-0.087	~	-0.119	~	-0.050		(0.048)	(0.062)	(0.060)
Sardegna	-0.028		-0.071		0.025		(0.063)	(0.083)	(0.073)
Sicilia	-0.068		-0.009		-0.119	*	(0.044)	(0.053)	(0.057)
Toscana	-0.007		-0.006		-0.014		(0.037)	(0.045)	(0.047)
Umbria	-0.058		-0.042		-0.089		(0.084)	(0.105)	(0.114)
Veneto	-0.015		-0.024		-0.019		(0.036)	(0.046)	(0.045)
ATE	-0.040	**	-0.027	*	-0.055	**	(0.011)	(0.013)	(0.014)
Mean (Control)	0.838	**	0.839	**	0.837	**	(0.007)	(0.007)	(0.007)
CV N.	10		10		10				
Sample Size	5107		3869		3866				

Notes: The table presents Conditional Average Treatment Effects (CATE) on the Politeness (0/1) of the response emails. Column (1) presents the results for any treatment, Column (2) the results for French aliases, and Column (3) the results for Arabic aliases. Columns (4), (5), and (6) specifiy the SEs for the estimates in Columns (1), (2), and (3) respectively. The estimates come from a linear AIPW model (Chernozhukov et al. (2018)). The nuisance parameters are estimated through a Random Forest with the Honesty approach (Athey and Imbens (2016)). Outcome and treatment are modeled conditional on province and day-of-sending, the randomization levels. 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. ** $p < 0.01$, * $p < 0.05$, ~ $p <$

Appendix Tables - NLP Evaluation Metrics

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.

Appendix Tables - Responses: Logit

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

Appendix Tables - Responses: LASSO

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. ** $p < 0.01$, * $p < 0.05$, ~ $p < 0.10$

Appendix Tables - Responses: 1Day Rate

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. ** $p < 0.01$, * $p < 0.05$, + $p < 0.10$

Appendix Tables - NLP: Logit

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 (Contro	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

Appendix Figures - CATE by Treatment: Response Rates

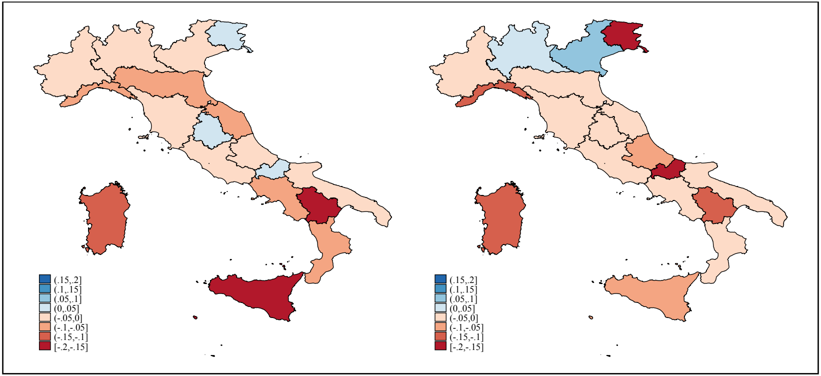


Figure A1: CATE by Treatment. French (left), Arabic (right).

Appendix Figures - CATE by Treatment: Usefulness

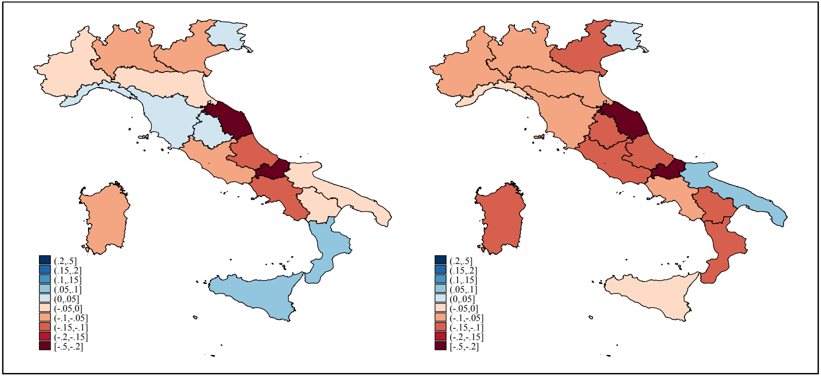


Figure A2: CATE by Treatment. French (left), Arabic (right).

Appendix Figures - CATE by Treatment: Politeness

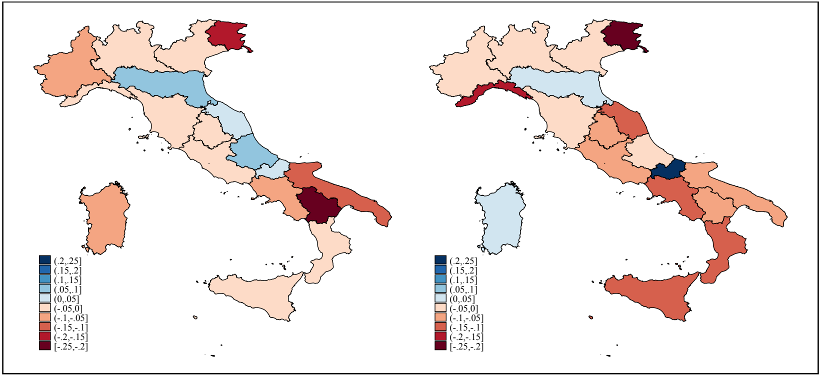


Figure A3: CATE by Treatment. French (left), Arabic (right).