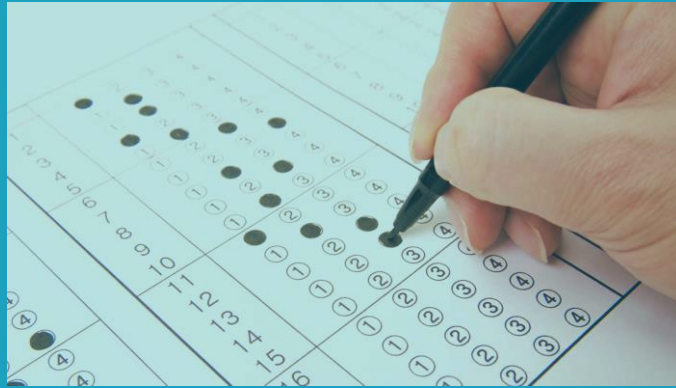


Gendered Language in Standardized Tests



**Alma Cohen, Tzur Karelitz, Tamar Kricheli-Katz,
Sepi Pumpian, and Tali Regev**

RESEARCH QUESTION

What is the effect of using a more gendered-neutral language on the performance of women and men in high-stakes exams?

THE POLICY DEBATE

Recent policy debate centers on mandating more gender-neutral language:

- 2021: U.S. House of Representatives implements gender-neutral language in all documents
- 2022: U.S. Educational Testing Service (SAT) discusses, then opts against mandating gender-neutral language

THE BROADER QUESTION: THE EFFECT OF LANGUAGE ON BEHAVIOR

Do languages affect or merely reflect the attitudes, preferences and behaviors of the people who speak them?

- ▶ Studies have documented correlations between linguistic features and grammatical structures of languages and between the attitudes, preferences and behaviors of the people who speak them.
- ▶ Speakers of languages with different structures and features vary in their processing of colors, future oriented economic behaviors, and **gendered attitudes**.

THE GENDERING OF LANGUAGES:

Gendered Languages vary parts of speech according to grammatical gender.

- ▶ In German, French, Spanish, and Hebrew, pronouns, nouns, adjectives, and verbs have feminine and masculine forms.

The good(f) wife - buena - 'tova'

The good(m) husband - bueno - 'tov'

THE GENDERING OF LANGUAGES:

Languages vary by the usage of a gender-neutral form v. a gendered form:

- ▶ A gender-neutral form does not differentiate between women and men (e.g., person).
- ▶ A gendered form refers to one gender only (e.g., waiter and waitress).

THE GENDERING OF LANGUAGES

Masculine Generics: the common usage of the masculine form as generic for both women and man (“policemen” v. “policepersons”).

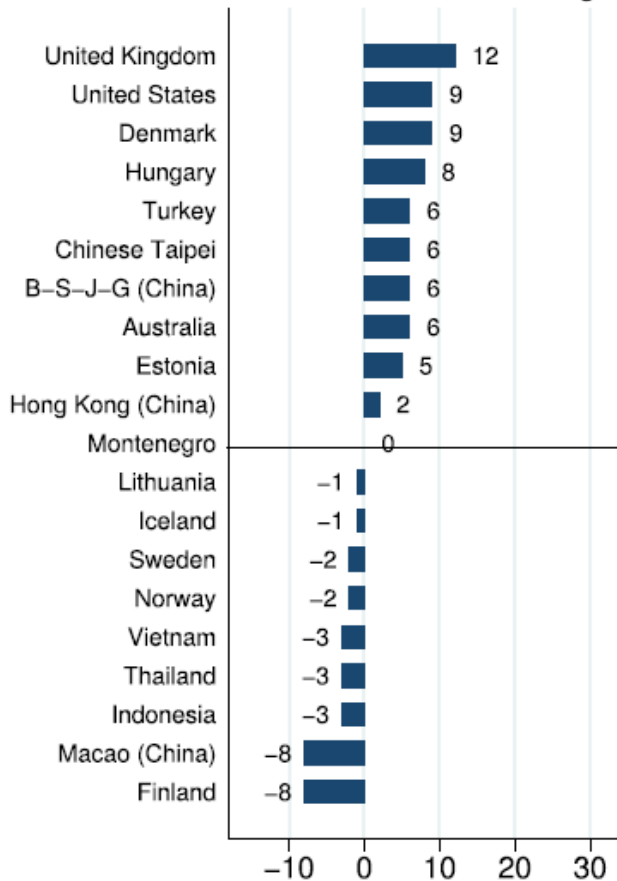
- ▶ Masculine generics are more prominent in gendered languages than in neutral gender languages.

THE GENDERING OF LANGUAGES AND GENDER INEQUALITY:

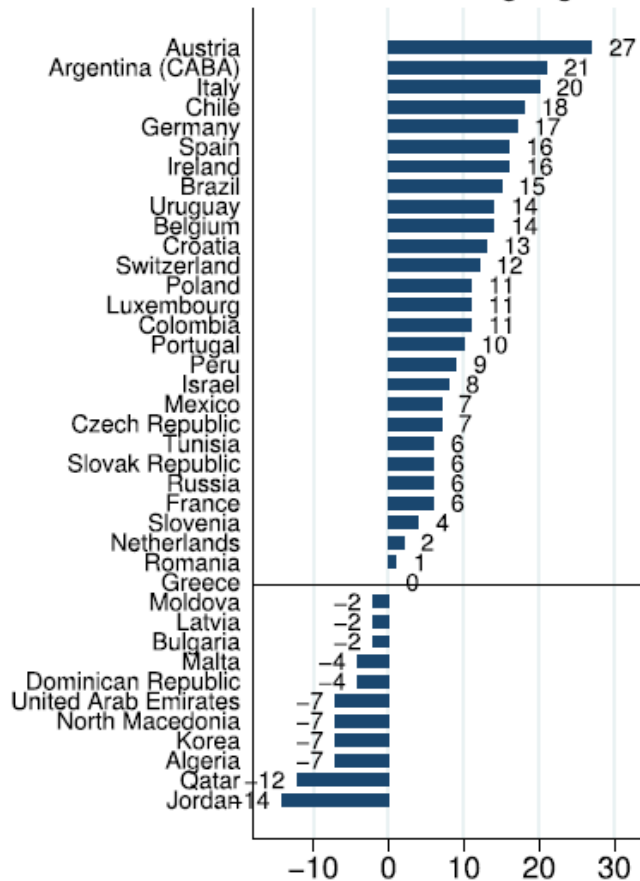
- ▶ Gender inequality is greater in countries where gendered languages are spoken compared to countries with other grammatical systems ([Prewitt-Freilino et al, 2012](#)).

THE GENDER GAP IN MATH (PISA 2015)

Genderless and Neutral Languages



Gendered Languages



WHY WOULD GENDERED LANGUAGES BE ASSOCIATED WITH A GENDER GAP?

1. **Stereotype threat:** gender-salience activates prevailing gender stereotypes about women's inferiority and affects women's performance ([Spencer, Steele and Quinn, 1999](#)).
2. **Exclusion of women:** the generic masculine form ignores and excludes women.
3. **Cultural / historical association:** the gendering of language is rooted in historical, long term disadvantages for women.

RECENT LITERATURE

Correlations:

- ▶ Prewitt-Freilino et al. 2012,
- ▶ Gay et al. 2013,
- ▶ Davis and Reynolds 2018: The education gender gap
- ▶ Jakiela and Ozier 2018:
- ▶ Galor, Ozak, and Sarid 2020: Female education attainment

Experimental evidence:

- ▶ Wasserman & Weseley, 2009: Bilingual participants exhibited more sexist attitudes in French and Spanish relative to English.
- ▶ Vainapel et al, 2015: Lower task goals for women in masculine-generics
- ▶ Kricheli-Katz and Regev 2021: Lower performance in math in masculine generics in the lab.

OUR CONTRIBUTION: EVIDENCE FROM A NATURAL EXPERIMENT

We exploit a policy change in the gendered form of address in **real high-stakes** university entrance exams.

In December 2009 the Israeli National Institute of Testing and Evaluations changed the form of address on the exams.

The singular-masculine → The more inclusive plural-masculine

ענה → ענו
responde(m) responden(m)

RESEARCH HYPOTHESES

Addressing women in the plural-masculine will improve women's grades in quantitative questions relative to the singular-masculine.

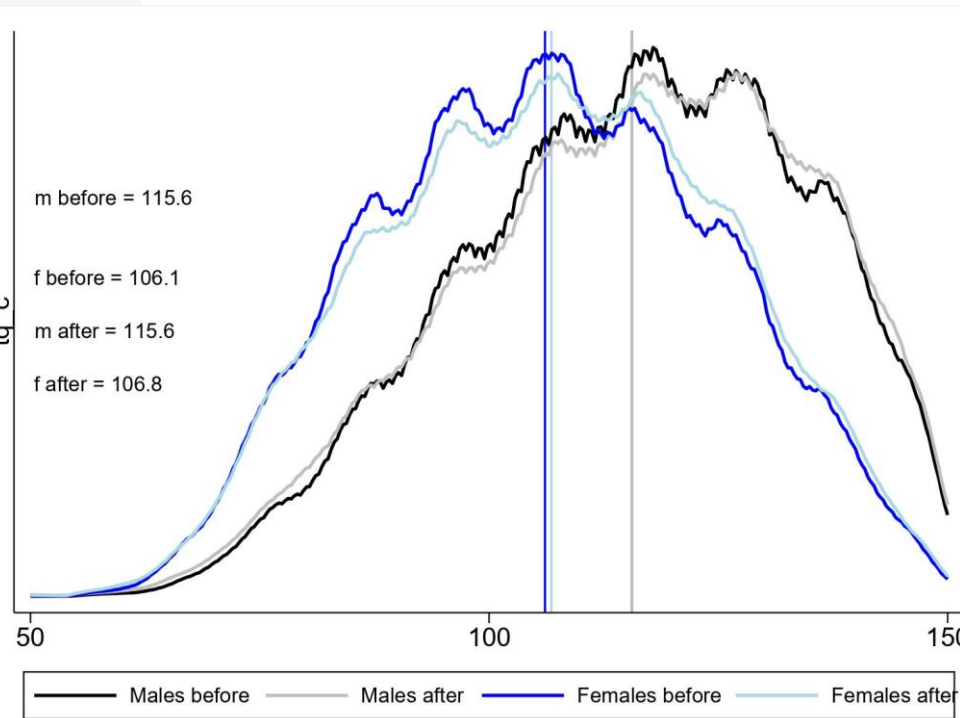
- ▶ We expect a lower or no effect in verbal questions (because stereotype threat is stronger for math).
- ▶ We expect no improvement for men in verbal or quantitative questions.

GRADE DISTRIBUTION BEFORE AND AFTER POLICY CHANGE

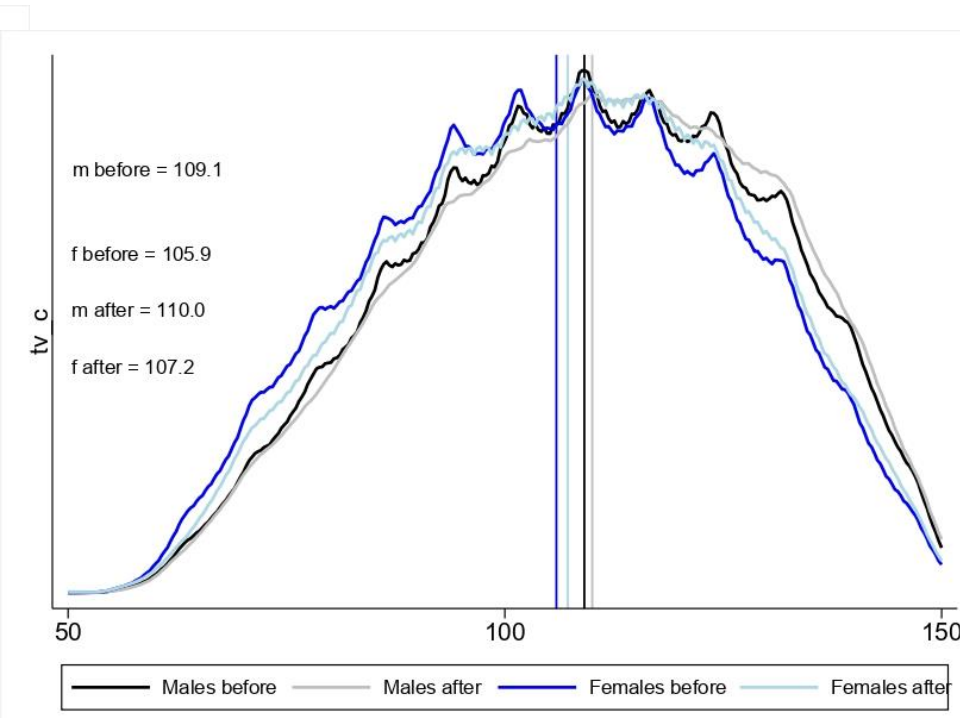
(all first time Hebrew psychometric test-takers 2000-2012)

14

Math



Verbal



PRELIMINARY

Preliminary: OLS Regression Models Predicting Exam Grades, 2000-2012

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Math	Math	Math	Verbal	Verbal	Verbal	English	English
female	-9.512*** (0.068)	-9.943*** (0.066)	-6.316*** (0.057)	-3.445*** (0.074)	-2.840*** (0.071)	2.004*** (0.055)	-9.616*** (0.085)	-8.798*** (0.082)
After	-1.795*** (0.283)	-1.646*** (0.264)	-2.036*** (0.225)	-0.462 (0.298)	-0.035 (0.276)	-0.556*** (0.210)	0.838** (0.348)	0.947*** (0.324)
Female X After	0.579*** (0.136)	0.403*** (0.126)	1.171*** (0.109)	0.434*** (0.145)	0.272** (0.134)	1.298*** (0.104)	-1.719*** (0.170)	-1.863*** (0.159)
Demographics		Y	Y		Y	Y		Y
English score			0.412*** (0.001)			0.551*** (0.001)		
Constant	116.015*** (0.133)	103.943*** (4.135)	82.752*** (3.529)	111.363*** (0.146)	45.415*** (4.625)	17.123*** (4.401)	116.822*** (0.164)	51.390*** (4.221)
Observations	395870	395870	395870	395870	395870	395870	395870	395870
Adjusted R-squared	0.067	0.201	0.414	0.013	0.176	0.516	0.048	0.191
Standard errors in parentheses		=** p<0.1	** p<0.05	*** p<0.01"				

All regressions include month and year fixed effects. Demographics include age, age squared, and fixed effects for income bracket, district, education level, mother's education level, father's education level, place of birth and years in Israel.

15

Math improved by 0.4% relative to the mean outcome

Verbal improved by 0.3%

RESEARCH METHOD

We compare women's grades before and after the policy change:

Identification issues:

1. What if the questions changed?
→ we use only **repeated chapters**
2. What if the population of test-takers changed?
→ we look **within person** at success in gendered vs non-gendered questions

REPEATED CHAPTERS

We have 8 Quantitative Chapters and 13 Verbal Chapters:

Chapter No.	04-2002	04-2005	12-2006	02-2007	04-2007	02-2008	12-2009	02-2010	04-2010	07-2010	10-2010
2845	X										X
2850				X							X
2911						X	X				
2957		X							X		
5069		X								X	
5070		X								X	
5073			X					X			
5083					X			X			

“Gendered Address Questions”, with a gendered form of address, constitute about 12% of questions in Quantitative chapters and 18% in Verbal chapters.

QUESTIONS WITH A GENDERED FORM OF ADDRESS

Before

שים לב

מקרא:	מואב	כנען	גושן	בבל	אשור	מדינת מקור	
						יעד	מדינת מקור
16:00-08:00 בוקר	2.0	2.0	3.2	5.5			אשור
24:00-16:00 ערב	1.2	2.0	2.0	2.5			
08:00-24:00 לילה	0.8	1.0	1.5	1.8			
	2.5	5.5	2.6		4.2		בבל
	2.0	3.2	2.2		1.5		
	1.5	2.1	0.5		1.5		
	2.2	4.0		4.5	3.2		גושן
	2.0	2.0		1.0	1.7		
	1.7	1.0		1.0	1.3		
	1.3		2.2	3.0	1.8		כנען
	1.1		2.2	3.0	1.8		
	1.0		2.2	3.0	1.8		
		1.3	2.2	2.5	2.0		מואב
		1.1	2.0	2.0	1.2		
		1.0	1.7	1.5	0.8		
	0	-2	+5	+8	0		הפרשי שעות מאשור

שים לב: בתשובתך לכל שאלה התעלם מנתונים המופיעים בשאלות האחרות.

After

שימו לב

מקרא:	מואב	כנען	גושן	בבל	אשור	מדינת יעד	
						מדינת יעד	מדינת מקור
16:00-08:00 בוקר	2.0	2.0	3.2	5.5			אשור
24:00-16:00 ערב	1.2	2.0	2.0	2.5			
08:00-24:00 לילה	0.8	1.0	1.5	1.8			
	2.5	5.5	2.6		4.2		בבל
	2.0	3.2	2.2		1.5		
	1.5	2.1	0.5		1.5		
	2.2	4.0		4.5	3.2		גושן
	2.0	2.0		1.0	1.7		
	1.7	1.0		1.0	1.3		
	1.3		2.2	3.0	1.8		כנען
	1.1		2.2	3.0	1.8		
	1.0		2.2	3.0	1.8		
		1.3	2.2	2.5	2.0		מואב
		1.1	2.0	2.0	1.2		
		1.0	1.7	1.5	0.8		
	0	-2	+5	+8	0		הפרשי שעות מאשור

QUESTIONS WITH A GENDERED FORM OF ADDRESS

Before the policy change:

עיינן היטב בתרשים שלפניך, וענה על ארבע השאלות שאחריו.
שים לב: בתשובתך לכל שאלה התעלם מנתונים המופיעים בשאלות האחרות.

After the policy change

עיינו היטב בתרשים שלפניכם, וענו על ארבע השאלות שאחריו.
שימו לב: בתשובתכם לכל שאלה התעלמו מנתונים המופיעים בשאלות האחרות.

QUESTIONS WITH A GENDER-AMBIGUOUS FORM OF ADDRESS

20

In some cases, the singular-masculine and singular-feminine are spelled the same way:

בסרטוט שלפניך ABCD הוא מלבן

Before the policy change, women could read these in the feminine form.

We expect no policy effect for women in these gender-ambiguous questions.

RESEARCH METHOD

We estimate the following OLS regression model:
(person i , question q , chapter c , at time t)

$$y_{iqct} = \beta_1 \text{Gendered}_q + \beta_2 \text{Gendered}_q X \text{After}_t + \beta_3 \text{Gendered}_q X \text{After}_t X \text{Female}_i \\ + \beta_4 X_{iq} + \gamma_c + \mu_q + \delta_i + \varepsilon_{iqct}$$

- ▶ y_{iqct} is a binary indicator of the answer being correct.
- ▶ All models control for:
 - chapter, question-number, and person fixed effects,
 - an interaction *Question No. X Female*,
 - the type of question (graph, geometric),
 - for “ambiguous” questions.

SAMPLE

- ▶ All examinees who took the psychometric test in Hebrew between 2000-2012
- ▶ Answered one of the chapters which were repeated pre and post the policy change (as a pilot or for calibration).
- ▶ Took the test for the first time.

STUDENTS' DESCRIPTIVE STATISTICS

Table 1: Descriptive Statistics – Means

	Quantitative			Verbal		
	Before	After	Diff in %	Before	After	Diff in %
Female	0.562	0.555	-1.403	0.547	0.548	0.087
Age	21.653	21.46	-0.892	21.457	21.608	0.702
Income	3.355	3.291	-1.899	3.312	3.283	-0.871
Education	3.513	3.762	7.093	3.439	3.533	2.749
Mother's Education	4.816	5.195	7.882	4.885	5.203	6.495
Father's Education	4.761	5.077	6.636	4.811	5.082	5.624
Born in Israel	0.761	0.677	-11.041	0.757	0.681	-10.037
Success Rate	0.625	0.643	2.821	0.644	0.647	0.436
Number of Obs.	18909	26173		35919	73264	

The range of Quantitative and Verbal grades is between 0 to 150. The age of test-takers ranges between 16 and 33.

QUESTIONS' SUCCESS RATES

Table 2: Descriptive Statistics - Question Success Rates

		Before	N	After	N	Diff in %
Quantitative	Women	0.595	179242	0.632	242676	6.27
	Men	0.687	139517	0.707	195667	2.803
Verbal	Women	0.656	319661	0.662	650077	0.894
	Men	0.678	263621	0.686	533873	1.143

MAIN RESULTS: QUANTITATIVE

OLS regression models predicting correctness of answers in quantitative questions

	All	Women	Men
Gendered	0.000 (0.004)	-0.015*** (0.004)	0.007* (0.004)
Gendered x After	-0.000 (0.004)	0.015*** (0.004)	-0.001 (0.004)
Gendered x Female	-0.010** (0.004)		
Gendered x After x Female	0.015** (0.006)		
Geometric Question	0.014*** (0.001)	0.028*** (0.002)	-0.002 (0.002)
Graph Question	0.105*** (0.003)	0.113*** (0.004)	0.094*** (0.004)
Question No.	-0.020*** (0.000)	-0.022*** (0.000)	-0.020*** (0.000)
Female x Question No.	-0.002*** (0.000)		
Constant	0.897*** (0.001)	0.866*** (0.001)	0.936*** (0.001)
N	757102	421918	335184
Adjusted R-Squared	0.27	0.27	0.27
Answer Mean	0.65	0.62	0.70
Standard errors in parentheses; * p<0.1 ** p<0.05 *** p<0.01			

VERBAL

26

OLS regression models predicting correctness of answers in verbal questions

	All	Women	Men
Gendered	0.022*** (0.002)	0.013*** (0.002)	0.022*** (0.002)
Gendered x After	-0.000 (0.002)	-0.000 (0.002)	-0.000 (0.002)
Gendered x Female	-0.009*** (0.003)		
Gendered x After x Female	-0.000 (0.003)		
Question No.	-0.010*** (0.000)	-0.011*** (0.000)	-0.010*** (0.000)
Female x Question No.	-0.001*** (0.000)		
Constant	0.850*** (0.001)	0.850*** (0.001)	0.850*** (0.002)
N	1767232	969738	797494
Adjusted R-Squared	0.19	0.18	0.18
Answer Mean	0.67	0.66	0.68

Standard errors in parentheses; * p<0.1 ** p<0.05 *** p<0.01

GENDER-AMBIGUOUS QUESTIONS

OLS regression models predicting correctness of answers in quantitative questions

Controls include: geometry, graph, question no., Female X question no.

	All	Women	Men
Gendered	-0.012*** (0.004)	-0.026*** (0.004)	-0.004 (0.004)
Gendered x After	-0.001 (0.004)	0.015*** (0.004)	-0.001 (0.004)
Gendered x Female	-0.008* (0.004)		
Gendered x After x Female	0.015*** (0.006)		
Ambiguous	-0.045*** (0.004)	-0.036*** (0.004)	-0.036*** (0.004)
Ambiguous x After	-0.009* (0.005)	-0.005 (0.004)	-0.007 (0.005)
Ambiguous x Female	0.017*** (0.005)		
Ambiguous x After x Female	0.005 (0.006)		

ROBUSTNESS

OLS regression models predicting correctness of answers in quantitative questions

28

	(1) ChapterXQuestion FE	(2) Excluding ambiguous	(3) Connected as non-gendered	(4) Connected as gendered	(5) Israeli Born
Gendered		-0.014*** (0.004)	0.037*** (0.003)	0.005** (0.002)	0.007* (0.004)
Gendered x After	0.001 (0.004)	-0.000 (0.004)	0.001 (0.004)	-0.003 (0.003)	-0.003 (0.005)
Gendered x Female	-0.009** (0.004)	-0.008* (0.004)	-0.007* (0.004)	-0.008*** (0.003)	-0.008* (0.005)
Gendered x After x Fema	0.013** (0.006)	0.015*** (0.006)	0.012** (0.005)	0.009** (0.004)	0.015** (0.007)
N	757102	674415	1117758	1117758	539416
Adjusted R-Squared	0.29	0.28	0.26	0.26	0.27
Answer Mean	0.65	0.65	0.64	0.64	0.66

Standard errors in parentheses; * p<0.1 ** p<0.05 *** p<0.01

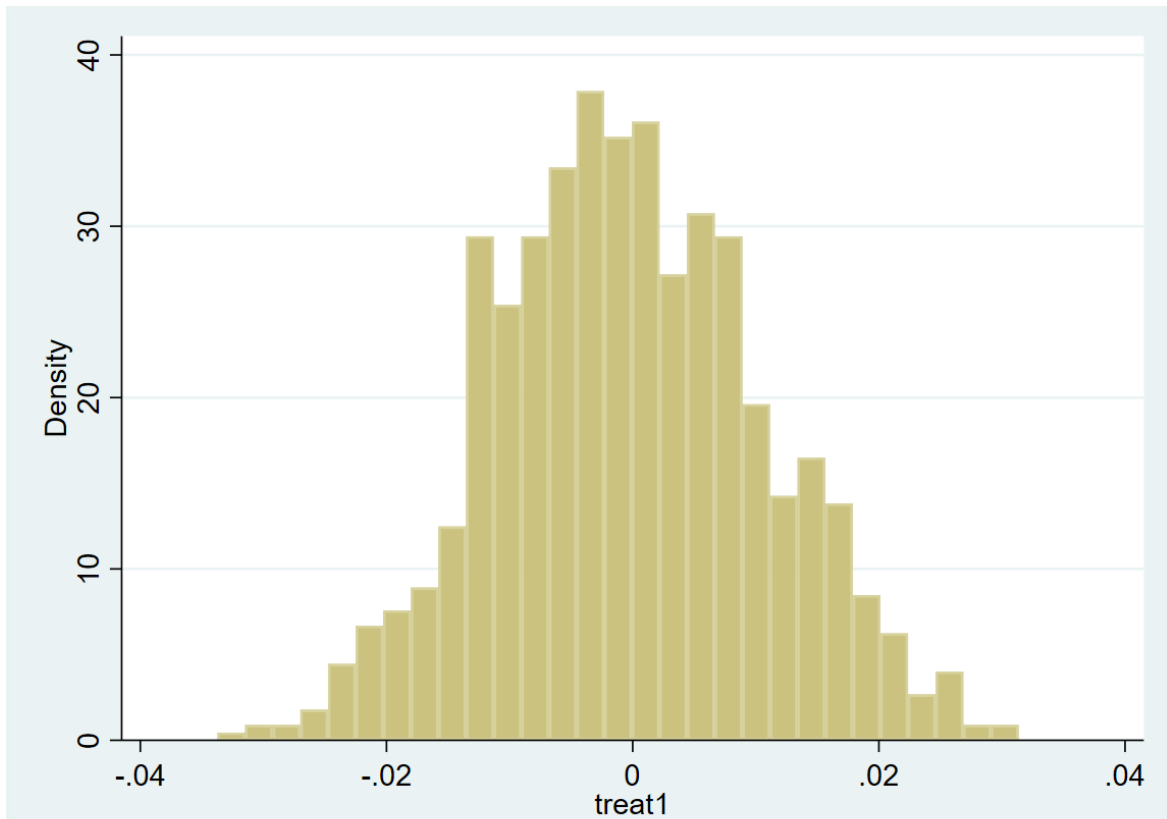
PLACEBO TEST

29

Test results of randomly assigning a 'placebo gendered address'

Coefficient on
Gendered X Female X After
when randomly choosing
which questions are
gendered,

1000 draws.
(95th pct = 0.018165)



SUMMARY

After the policy change:

- ▶ The gap between women's success in gendered vs. non-gendered questions increased by 1.5 percentage points in quantitative questions.
(An increase of 2.4% relative to women's average success of 62%)
- ▶ No change in performance was observed for men, or for women in verbal questions

LIMITATIONS

- ▶ Gendered address probably affects test-takers beyond the question in which it appears.
- ▶ Our research design is not able to capture this change.
- ▶ Therefore, our findings likely underestimate the true impact of the policy.

IMPLICATIONS

- ▶ Our findings support the hypothesis that the use of the singular masculine form of address can lead to "stereotype threat" for women in math exams.
- ▶ As such, these findings have important implications for policy on addressing individuals with gendered languages in tests and educational settings.
- ▶ Furthermore, they highlight the need for further research on the effects of gendered forms of address on women's performance in less gendered languages, such as English.

IMPLICATIONS

- ▶ English does not assign gender to nouns and pronouns
- ▶ However, it does have some gendered aspects:
 - ex: using "he" as a default pronoun
 - ex: using gendered terms for certain professions
- ▶ Our findings suggest that the transition toward a more inclusive language may have additional effects on the performance of women in tests and educational systems, that are greater than previously understood.

THE LONG-STANDING LINGUISTIC DEBATE

34

- ▶ Linguistic relativism - The Sapir-Whorf Hypothesis (1956): The languages we speak influence and possibly determine the way we think and view the real world.
(Carroll et a. 2004, Levinson 2012, Lucy 1996)
- ▶ Universal nature of human language and cognition
(attributed to Chomsky; Li and Gleitman 2002)

Our take: language as an institution

- ▶ Languages routinely and actively participate in enacting and maintaining schemas about gender, time, status...;
- ▶ Whenever a language is spoken, the schemas are activated, preferences are reinforced, and behaviors follow accordingly.

Thank You