

Cruz Cortés, Efrén; Lee, John; Lehoucq, Emilio; Suschevskiy, Vsevolod

**Working Paper**

## A Comment on "Se Habla Español: Spanish-Language Appeals and Candidate Evaluations in the United States"

I4R Discussion Paper Series, No. 165

**Provided in Cooperation with:**

The Institute for Replication (I4R)

*Suggested Citation:* Cruz Cortés, Efrén; Lee, John; Lehoucq, Emilio; Suschevskiy, Vsevolod (2024) : A Comment on "Se Habla Español: Spanish-Language Appeals and Candidate Evaluations in the United States", I4R Discussion Paper Series, No. 165, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/303903>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



No. 165

I4R DISCUSSION PAPER SERIES

# **A Comment on “Se Habla Español: Spanish-Language Appeals and Candidate Evaluations in the United States”**

efrén cruz cortés

John Lee

Emilio Lehoucq

Vsevolod Suschevskiy

October 2024

## I4R DISCUSSION PAPER SERIES

I4R DP No. 165

### **A Comment on “Se Habla Español: Spanish-Language Appeals and Candidate Evaluations in the United States”**

**efrén cruz cortés<sup>1</sup>, John Lee<sup>1</sup>, Emilio Lehoucq<sup>1</sup>, Vsevolod Suschevskiy<sup>1</sup>**

*<sup>1</sup>Northwestern University, Evanston/USA*

OCTOBER 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](https://www.zbw.eu/), and [RWI – Leibniz Institute for Economic Research](https://www.rwi-essen.de/), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

#### **Editors**

**Abel Brodeur**  
*University of Ottawa*

**Anna Dreber**  
*Stockholm School of Economics*

**Jörg Ankel-Peters**  
*RWI – Leibniz Institute for Economic Research*

## A comment on “Se Habla Español: Spanish-Language Appeals and Candidate Evaluations in the United States”

efrén cruz cortés (Northwestern University, [efren.cruzcortes@northwestern.edu](mailto:efren.cruzcortes@northwestern.edu))  
John Lee (Northwestern University, [jshlee@northwestern.edu](mailto:jshlee@northwestern.edu)) (corresponding author)  
Emilio Lehoucq (Northwestern University, [emilio@northwestern.edu](mailto:emilio@northwestern.edu))  
Vsevolod Suschevskiy (Northwestern University, [vvseva@u.northwestern.edu](mailto:vvseva@u.northwestern.edu))<sup>1</sup>

### Abstract

Zárate, Quezada-Llanes, and Armenta (2024) examine whether Hispanic and Anglo voters change voting behavior if a political candidate speaks to them in Spanish and whether it matters how proficient in Spanish the candidate sounds. They find that Hispanic support for the Anglo and Hispanic candidates is higher in the native-like Spanish condition compared with the English-only condition. Relative to the English condition, non-native Spanish does not increase support for the Anglo candidate, but it decreases support for the Hispanic candidate. They find mixed effects for the Anglo participants. We conducted computational and robustness reproductions. First, we successfully computationally reproduced all the main results. Second, we constrained the analysis to participants that passed the manipulation checks (this was not done by the authors). We find the same results hold. Notice we constrained ourselves to study 1, which is based on data by Prolific.

---

<sup>1</sup> The authors did not have any financial support or conflict of interest. The authors' response is available here: <https://osf.io/xfp4k/>. This report was prepared for the Institute for Replication. This is a collaboration with Nature Human Behaviour (Brodeur et al., 2024).

# 1. Introduction

Zárate, Quezada-Llanes, and Armenta (2024) (hereinafter, the authors) conducted two studies. For study 1, they obtained a national sample of 503 Hispanic and 506 Anglo respondents through the survey firm Prolific. The survey was fielded from October 26-31, 2020. Overall, their Hispanic and Anglo samples were younger, had higher levels of education, were more liberal, and were more likely to identify as Democrat compared with estimates of the United States population. For study 2, they obtained a sample of 1000 Hispanic and 1000 Anglo respondents through Lucid. The survey was fielded from December 28, 2021 to January 11, 2022. The partisan composition of the sample corresponds to the two-party 2020 presidential vote share.

The authors use a between-subjects survey experiment. Prior to beginning both experiments, all participants were told that they were being asked to participate in a research study examining how people evaluate state legislature candidate campaign speeches. At the beginning of the survey, they asked several demographic questions including age, gender, ideology, and partisanship. In study 1, participants were randomly assigned to one of six conditions where they were instructed to listen to a short audio clip from a campaign speech of a candidate running for a seat in a state legislature. They varied whether the candidate is Hispanic or Anglo, and whether part of the campaign speech is spoken in non-native Spanish, native-like Spanish, or fully in English. To vary the ethnicity of the candidate, they gave a Hispanic- or Anglo-sounding name to the hypothetical candidate (Josué Martínez or Josh Martin) and explicitly stated the candidate's ethnicity or race. For the language treatment, respondents were assigned to one of three different audio recordings, all of which shared the same substantive content. The only differences between these audio files was in the language, English or Spanish, used in two sentences and the candidate's accent when speaking Spanish. For Study 2, they also vary whether the candidate is running in a nonpartisan or partisan primary election for a total of 12 treatment conditions. Participants in Study 2 either did not know the candidate's partisanship or were told that the candidate was running in their self-identified political party primary. After receiving the treatment, they included four questions to measure their outcome of interest: evaluations of the candidate. They

measure candidate evaluations by asking respondents how likely they are to vote for the candidate, whether they trust the candidate, how much they like or dislike the candidate, and if they would feel represented with the candidate in office. All items are measured on a 5-point scale. For their dependent variable, they construct an index of candidate evaluations by summing scores of these four variables. Following the candidate evaluation measures, they collected other relevant social and demographic information. In Study 2, they asked for racial identification information at the beginning of the survey and randomly assigned participants to receive the group strength questions either before or after the treatment.

In study 1, they “find that Hispanics prefer candidates who use Spanish-language appeals but only when the candidate speaks with native-like proficiency. This is true for both of the Anglo and Hispanic candidates in [their] experiment. [...] Spanish-language appeals do not have a consistent impact, positive or negative, on Anglos’ candidate evaluations [...]” (p. 364). Study 2 “successfully replicate[s] these findings” (p. 364).

For our reproduction, we obtained the data from the American Political Science Review Dataverse. We did not contact the authors for further data or clarifications.

We conducted computational and robustness reproductions. First, we successfully computationally reproduced all the main results (i.e., Tables 1 and 2 and Figures 1, 2, 3, and 4 of the paper) from the raw data and code provided. A small coding error was found, and is discussed below. Second we reproduced the results of hypotheses 2 and 3 for Study 1 (that Hispanic voters give higher evaluations to Hispanic candidates, and that relative to the Anglo candidate speaking English, they will punish Hispanic candidates speaking non-native Spanish) purely from the raw data, without looking at their code. Finally, we enforced manipulation checks on participants and discarded those who answered incorrectly (they indicated the incorrect candidate ethnicity or language), and additionally also enforced match between ethnicity self-identification and Prolific’s classification.

## 2. Computational Reproducibility

We used the replication package here:

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GT15SC>.

The replication package includes a script for all analyses. The authors conducted supplementary analyses with additional datasets that were not available in the replication package. One of those datasets was proprietary. We successfully computationally reproduced all the main results (i.e., Tables 1 and 2 and Figures 1, 2, 3, and 4 of the paper) from the raw data. Our code is available here: <https://osf.io/s4uhg/>. We found one coding error, which could have been impactful but by fortune wasn't: when the authors encode the candidate's ethnicity under the "treatment" field, they incorrectly assign Q10 to the Anglo-English condition. The correct question to use is Q10.1. Fortunately, this was the first encoded condition, and the following ones overwrote the "mistake", except for the cases for which it actually applied. Hence the error had no consequence. However, it could have resulted in different outcomes had the list not been exhausting. Also, notice that if they thought the candidate was identified as biracial, the order of encoding matters, since posterior encodings overwrite previous ones. We did not investigate this further, but we looked into a similar issue when it comes to participant's self-identification of ethnicity. See [Section 4.1](#) for further details.

Our results, compared with those of the original paper, can be seen in [Figures 1-4](#).

	Fully	Partial	No
Raw data provided	x		
Cleaning code provided	x		
Analysis data provided			x
Analysis code provided	x		
Reproducible from raw data			
Reproducible from analysis data			NA

## 2.1 Discrepancies Between Pre-analysis Plan and Article

The authors pre-registered both studies (study 1: <https://osf.io/6ebav>; study 2: <https://osf.io/g36q9>), including the analysis plans. The table below contrasts the pre-registrations with the published paper.

	<b>Study 1</b>	<b>Study 2</b>
Hypotheses	The pre-registered hypotheses are essentially the same as in the published paper (with different wording and order).	The pre-registration didn't re-specify the hypotheses.
Conditions	The conditions specified in the pre-registration are the same as in the published paper.	The conditions specified in the pre-registration are the same as in the published paper.
Respondents	The number of respondents specified in the pre-registration is essentially the same as in the published paper.	The number of respondents specified in the pre-registration is the same as in the published paper.
Dependent variable	The wording of the questions is the same as in the published paper. See "Analysis" below.	In this pre-registration, the authors focused on the index. They also mention that they added new dependent variables per the editor and reviewers' comments.
Analysis	All the main analyses are mentioned. The threshold for statistical significance is not specified. One difference is that the pre-registration says that they would run regressions for each of the four measures for the dependent variable and, in addition, one with an index. The paper only reports the index.	Same as for study 1.



Outliers	While in pre-registration they said they would remove or winsorize outliers, we did not identify any outlier detection and removing from the paper.	This pre-registration didn't specify anything about outliers.
----------	---	---

In addition to the information provided in table above, it is worth mentioning that the authors did not specify any data wrangling decisions in the pre-registration. This includes issues such as recoding variables and handling missing data. Also, the published paper did not report everything exactly as specified in the pre-registration. However, the authors mention this in footnote 8 and provide extensive supplementary materials. The supplementary materials also include additional analyses not specified in the pre-registration.

### 3. Recreate Reproduction

We conducted recreate reproducibility using the replication package. We focused on replicating the findings of Study 1. To achieve recreate reproducibility, one of us attempted to replicate the findings starting with the raw data without having looked at the code from the authors. We used the same linear regression models as the authors.

We followed the steps outlined in the “Experimental Design” section of the original article to create appropriate dependent and independent variables. The survey instruments in the replication package provided enough information to guide data processing. [Table A3](#) in the Appendix below shows the regression output using the data from Study 1. The first and second columns show the results from the original paper, while columns 3 and 4 are the result of our recreate reproduction. Both the direction and statistical significance of the effect remain the same, and the magnitudes are similar.

## 4 Robustness

Furthermore, we conducted robustness reproducibility tests. We did two independent analyses based on the two branches of our code: one from the code taken from the authors (used for computational reproduction) and one from the code we wrote from scratch (used for recreate reproduction). We made independent decisions during data processing, which resulted in our final dataset having different sample size and covariates. We used the same linear regression models as the authors. For the first branch, based on the computational reproduction code, we chose only the participants whose ethnic self-identification matched the one provided by prolific. This is discussed in [Section 4.1](#). For the second branch, based on the recreate reproduction code, we 1) enforced manipulation checks, and 2) jointly enforced manipulation checks and restricted analysis to those whose ethnicity self-identification matched Prolific's. This is discussed in [Section 4.2](#).

### 4.1 Self-Identification for Race/Ethnicity – Based on Computational Reproducibility

During data processing, we noticed that a series of questions were asked to report their ethnicity in multiple-choice questions Q53\_1, Q53\_2, Q53\_3, Q53\_4, Q53\_5, Q53\_6. The provided code used a unique strategy for coding race, based on exclusive self-nomination, where the last option overwrites all previous choices. For example, if a participant self-reported to be Black and White race column would reflect only white Race/Ethnicity. At the same time, if a participant self-reported to be Black, White and Latino then the Latino option will overwrite all previous ones. We consider it a mistake, since running the provided code in a different order will result in a different definition of race. To demonstrate the possible ambiguity of this choice we compared White and Latino samples by Race/Ethnicity distributions by running code in the original and reversed orders.

```
# Race/Ethnicity
prolific.df$race[prolific.df$Q53_1==1] <- 1 #Black
prolific.df$race[prolific.df$Q53_2==1] <- 2 #White
prolific.df$race[prolific.df$Q53_3==1] <- 3 #Asian
prolific.df$race[prolific.df$Q53_4==1] <- 4 #Latino
```

```

prolific.df$race[prolific.df$Q53_5==1] <- 5 #Native American
prolific.df$race[prolific.df$Q53_6==1] <- 6 #Marked other

```

Variable	Original order of executing			Reverse order of executing		
	N	Latino, N = 503 <sup>1</sup>	White, N = 505 <sup>1</sup>	N	Latino, N = 503 <sup>1</sup>	White, N = 505 <sup>1</sup>
race	1,006			1,006		
Black		4 (0.8%)	1 (0.2%)		11 (2.2%)	1 (0.2%)
White		38 (7.6%)	494 (98%)		112 (22%)	499 (99%)
Asian		2 (0.4%)	1 (0.2%)		4 (0.8%)	1 (0.2%)
Latino		452 (90%)	4 (0.8%)		374 (75%)	1 (0.2%)
Native American		1 (0.2%)	2 (0.4%)		0 (0%)	0 (0%)
Marked other		5 (1.0%)	2 (0.4%)		1 (0.2%)	2 (0.4%)

<sup>1</sup>n (%)

While the authors do not use the self-reported Race/Ethnicity to split the sample from Study 1 into Latino and White and rely on Race/Ethnicity that was assigned to participants by Prolific. Given the complexity of the concept of Race/Ethnicity and the possibility of mixed or combined identity, we decided to conduct a robustness check, where Latino and White groups will be identified based on a survey response only. For the simplicity of the study, we will assign participants who have exclusively self-identified as White in the survey as White, and participants who exclusively identified as Latino in the survey. If participants have reported more than one Race/Ethnicity or if they reported a Race/Ethnicity that is not exclusively Latino or not exclusively White, then we removed them from this robustness check. The difference in demographic characteristics is summarized in the table below.

Variable	Original groups				Groups by exclusive self-nomination				
	N	Latino, N = 503 <sup>1</sup>	White, N = 505 <sup>1</sup>	p-value <sup>2</sup>	N	Latino, N = 371 <sup>1</sup>	White, N = 532 <sup>1</sup>	other and mixed, N = 105 <sup>1</sup>	p-value <sup>3</sup>
<b>age</b>	1,005	27 (22, 32)	31 (25, 40)	<0.001	1,005	26 (22, 32)	31 (25, 40)	28 (22, 34)	<0.001
<b>female</b>	979	219 (45%)	253 (52%)	0.020	979	161 (44%)	262 (51%)	49 (48%)	0.2
<b>education</b>	1,008			<0.001	1,008				
1		6 (1.2%)	8 (1.6%)			4 (1.1%)	8 (1.5%)	2 (1.9%)	
2		77 (15%)	77 (15%)			60 (16%)	76 (14%)	18 (17%)	
3		149 (30%)	125 (25%)			111 (30%)	135 (25%)	28 (27%)	
4		63 (13%)	45 (8.9%)			45 (12%)	50 (9.4%)	13 (12%)	
5		169 (34%)	147 (29%)			124 (33%)	157 (30%)	35 (33%)	
6		31 (6.2%)	87 (17%)			23 (6.2%)	90 (17%)	5 (4.8%)	
7		7 (1.4%)	4 (0.8%)			3 (0.8%)	4 (0.8%)	4 (3.8%)	
8		1 (0.2%)	12 (2.4%)			1 (0.3%)	12 (2.3%)	0 (0%)	
<b>partyid</b>	970			<0.001	970				0.003
1		168 (35%)	179 (37%)			126 (35%)	188 (37%)	33 (32%)	
2		123 (25%)	71 (15%)			98 (28%)	78 (15%)	18 (17%)	
3		58 (12%)	72 (15%)			43 (12%)	73 (14%)	14 (14%)	
4		42 (8.7%)	40 (8.2%)			29 (8.2%)	42 (8.2%)	11 (11%)	
5		25 (5.2%)	34 (7.0%)			16 (4.5%)	35 (6.8%)	8 (7.8%)	
6		45 (9.3%)	42 (8.7%)			28 (7.9%)	47 (9.2%)	12 (12%)	
7		24 (4.9%)	47 (9.7%)			15 (4.2%)	49 (9.6%)	7 (6.8%)	
<b>ideology</b>	1,007			0.003	1,007				
1		95 (19%)	141 (28%)			73 (20%)	146 (27%)	17 (16%)	
2		151 (30%)	120 (24%)			115 (31%)	126 (24%)	30 (29%)	
3		73 (15%)	64 (13%)			53 (14%)	69 (13%)	15 (14%)	
4		92 (18%)	82 (16%)			71 (19%)	86 (16%)	17 (16%)	
5		55 (11%)	43 (8.5%)			36 (9.7%)	49 (9.2%)	13 (13%)	
6		26 (5.2%)	35 (6.9%)			16 (4.3%)	36 (6.8%)	9 (8.7%)	
7		10 (2.0%)	20 (4.0%)			7 (1.9%)	20 (3.8%)	3 (2.9%)	
<b>income</b>	1,008	6.0 (4.0, 9.0)	6.0 (4.0, 10.0)	0.2	1,008	6.0 (3.0, 9.0)	6.0 (4.0, 10.0)	6.0 (4.0, 9.0)	0.2
<b>treatment</b>	1,008			>0.9	1,008				0.056
1		84 (17%)	83 (16%)			54 (15%)	91 (17%)	22 (21%)	
2		84 (17%)	83 (16%)			70 (19%)	87 (16%)	10 (9.5%)	
3		84 (17%)	85 (17%)			64 (17%)	85 (16%)	20 (19%)	
4		85 (17%)	84 (17%)			68 (18%)	93 (17%)	8 (7.6%)	
5		84 (17%)	86 (17%)			58 (16%)	86 (16%)	26 (25%)	
6		82 (16%)	84 (17%)			57 (15%)	90 (17%)	19 (18%)	

<sup>1</sup>Median (IQR); n (%)<sup>2</sup>Wilcoxon rank sum test; Pearson's Chi-squared test<sup>3</sup>Kruskal-Wallis rank sum test; Pearson's Chi-squared test

We are happy to report that all reported results remained consistent with the original study. The direction and significance level of the results remained the same, given the alpha value at 0.1 and the same effect size. The only difference was found in Table 2 of

the paper “Effects of Language and Ethnicity on Candidate Evaluations among Anglos”, where interaction Hispanic x Spanish was not significant at level 0.1 ( $b = -0.065$ ,  $SE = 0.043$ , **Table A2**) in the original study, but became significant ( $b = -0.076$ ,  $SE = 0.041$ , **Table A2**) at level 0.1 in the robustness check analysis. A table comparing this table with our robustness scenario is found [Table A2](#). However, we decided not to interpret the interactive effect, as the direction of the effect was consistent with the original study and the uncertainty intervals overlap. We have included a comparison of the figures and tables reported in the paper for study 1 with those from the robustness check in Appendix A. See Figures [A1](#), [A2](#), and [A3](#), and [Table A1](#) and [Table A2](#).

## 4.2 Removing Participants that Failed Manipulation Checks, and Matching Ethnicity Self-identification with Prolific’s classification – Based on Recreate Reproducibility

A series of questions were asked to evaluate participants’ attention (manipulation checks) in the survey. Participants were told the name and ethnicity (White or Latino) of their assigned candidate in the survey vignette. After listening to the recording, participants were asked to answer the race of the candidate and the language spoken in the recording. In study 1, for the Hispanic participants dataset, we found that only 418 of 501 participants correctly identified the candidate’s ethnicity.

We remove the participants that did not pass the manipulation checks from the analyses. The candidate’s race is hypothesized to influence a participant’s evaluation of the candidate; therefore it is questionable if the effect of candidate’s race is truly reflected by those unable to correctly identify the candidate’s race. Our result is shown in [Table A4](#). Our results match the original results, with the effect directions remaining the same as well as almost all significant levels. The one difference is for hypothesis 2, non-fluent Spanish, in which the effect ceased to be significant.

It is worth pointing out that our analysis removing participants that did not pass the manipulation checks is not free of potential problems. As pointed out by the original authors in our communication with them, removing participants based on manipulation checks can lead to cherry picking and to non-randomness of the samples. This decision can subsequently lead to an increase in Type I and Type II errors (Kotzian, Stoeber, and

Weissenberger 2020). Despite this possibility, it is reassuring that the original findings remain nearly the same both removing and not removing participants who did not pass the manipulation checks.

A few words regarding those who did not pass the manipulation checks are in order. We noticed misidentification mostly happened from Latino to white and vice versa. Upon further investigation, restricting ourselves only to those participants who misidentified the Latino candidate as white and the white candidate as Latino, we found the following: for misidentifications of the Latino candidate, 60% happened when he spoke non-fluent Spanish, compared to 33% expected from random chance (vs fluent Spanish and English). Similarly, for misidentifications of the white candidate, 54% happened when he spoke fluent Spanish, again deviating from the 33% expected by chance. We did not pursue this further, but it may be due to participants encountering unexpected situations: a Latino candidate speaking non-fluent Spanish, and a white candidate speaking fluent Spanish.

<b>Number of misidentifications for the Latino and white candidate per Spanish fluency level.</b>		
<b>Candidate Ethnicity</b>	<b>Candidate Language</b>	<b>Number of misidentifications</b>
Latino	English	15
Latino	Fluent Spanish	9
Latino	Non-fluent Spanish	35
White	English	8
White	Fluent Spanish	28
white	Non-fluent Spanish	16

We further added the self-identification constraint. That is, besides passing the manipulation checks, participants were restricted to also match their ethnicity self-identification to that indicated by Prolific (see the discussion in [Section 4.1](#)). Only 383 of the original 501 remained. Our results are presented in [Table A5](#). We are happy to report that even after this constraint the effect direction and significance remains the same, except again for hypothesis 2, non-native Spanish, for which significance ceased.

## 5. Conclusion

To reproduce the results of the paper, we took two different paths, one based on the code provided by the authors (computational reproduction) and one based on code created from scratch (recreate reproduction). In the computational reproduction branch, we perfectly reproduced the figures and tables from the paper. See [Figures 1-4](#) and [Tables 1-2](#). Notice these involve both the Prolific and the Lucid datasets. We used this as basis to add a robustness check on the Prolific dataset: using only the data for which ethnicity self-identification matched Prolific's identification. The results are in [Figures A1-A3](#) and [Tables A1-A2](#). We happily report the general results hold. In the recreate reproduction branch, we limited ourselves to study 1 (the Prolific dataset). Our recreate reproduction is available in [Table A3](#), where we happily inform the same direction and significance was found. We further built on this to add two constraints, the first one was subsetting participants to only those who passed manipulation checks, this is shown in [Table A4](#). The second one further enforces participant's match between ethnicity self-identification and Prolific's identification. The results are shown in [Table A5](#). We are happy to report that for both these scenarios the directions remain the same, and with one exception the significance levels also remain the same.

## References

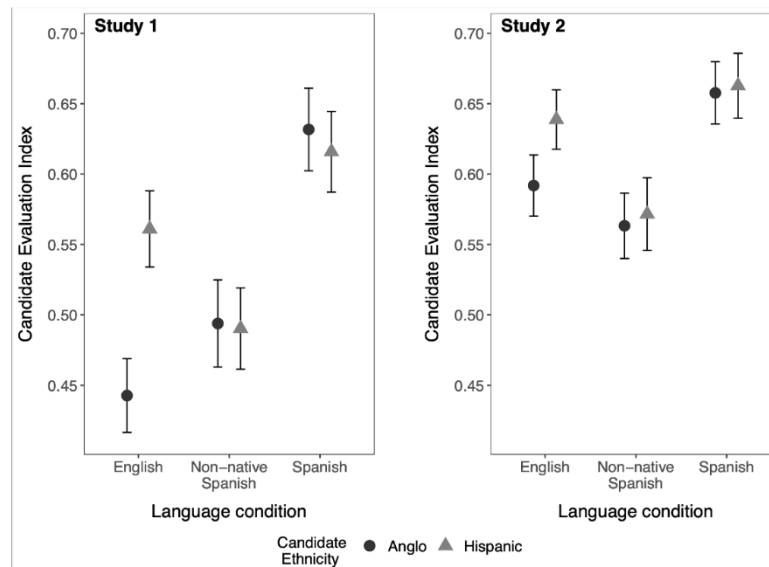
- Brodeur, A., Dreber, A., Hoces de la Guardia, F. et al. Reproduction and replication at scale. *Nat Hum Behav* 8, 2–3 (2024). <https://doi.org/10.1038/s41562-023-01807-2>
- Kotzian, P., Stoeber, T., Hoos, F., & Weissenberger, B. E. (2020). To be or not to be in the sample? On using manipulation checks in experimental accounting research. *Accounting Research Journal*, 33(3), 469-482.
- Zárate, M.G., Quezada-Llanes, E. & Armenta, A.D. (2024). Se Habla Español: Spanish-Language Appeals and Candidate Evaluations in the United States. *American Political Science Review*, 118(1), 363-379.



# Figures

Figure 1

## Computational Reproduction



## Original Paper

FIGURE 1. Mean Candidate Evaluation among Hispanics by Treatment Group

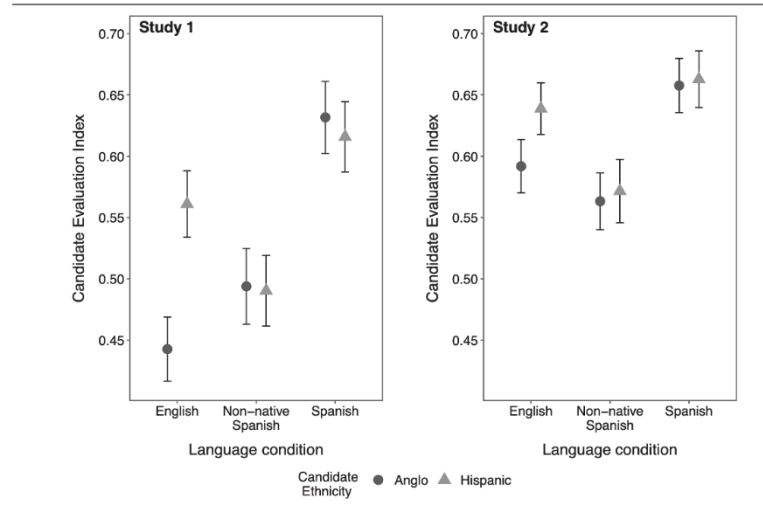


Figure 2

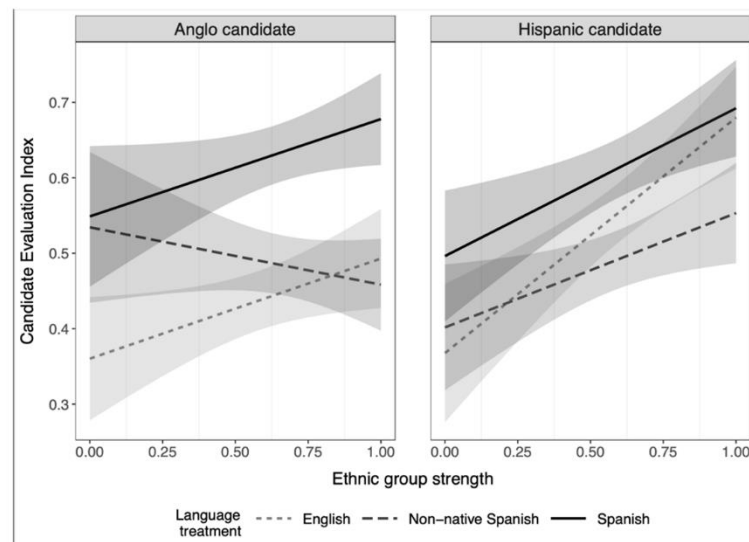
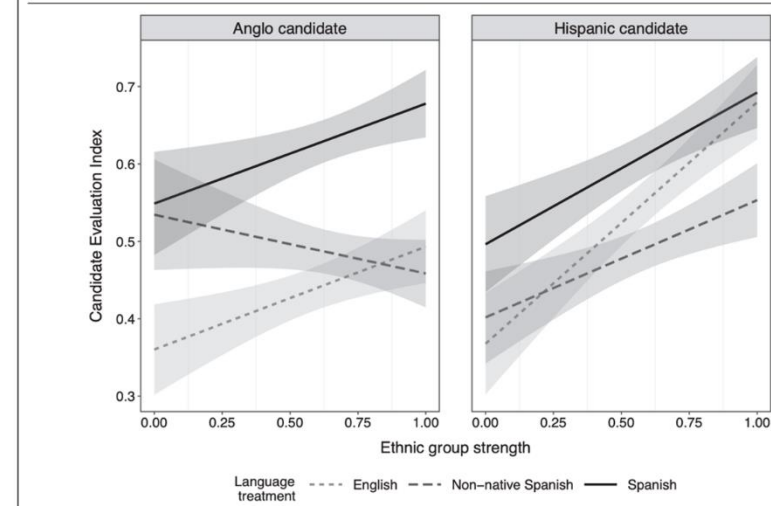
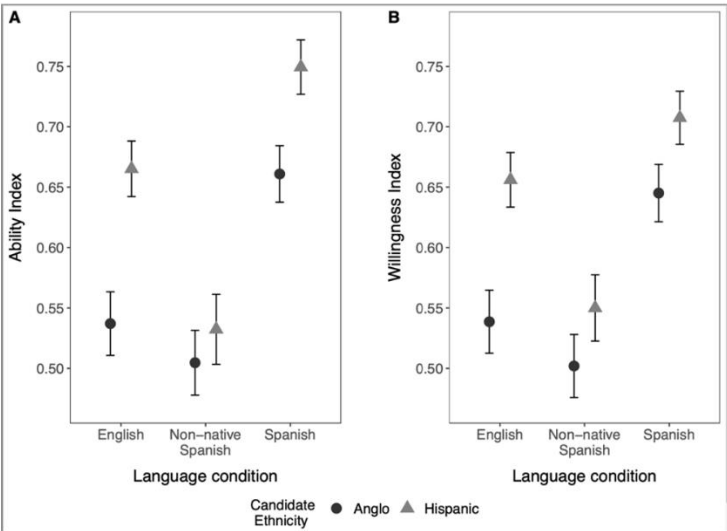
**Computational Reproduction****Original Paper****FIGURE 2. Predicted Candidate Evaluation among Hispanics by Treatment Group across Levels of Ethnic Identity (Study 1)**

Figure 3

Computational Reproduction



Original Paper

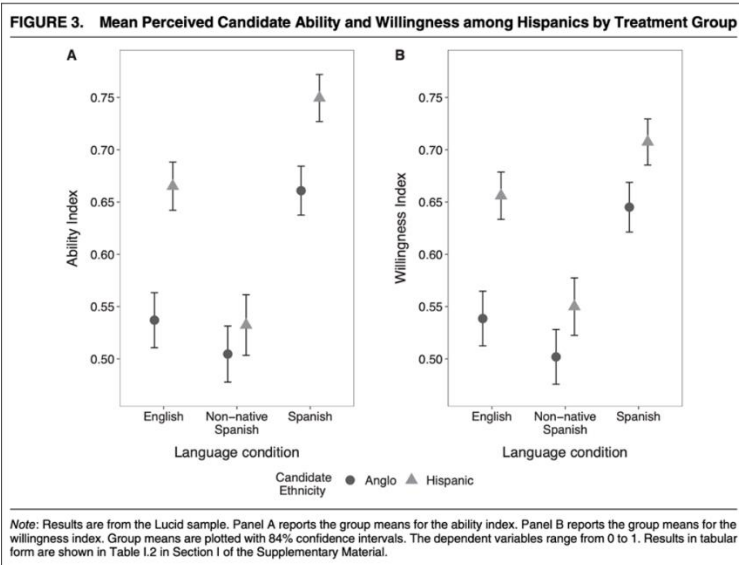
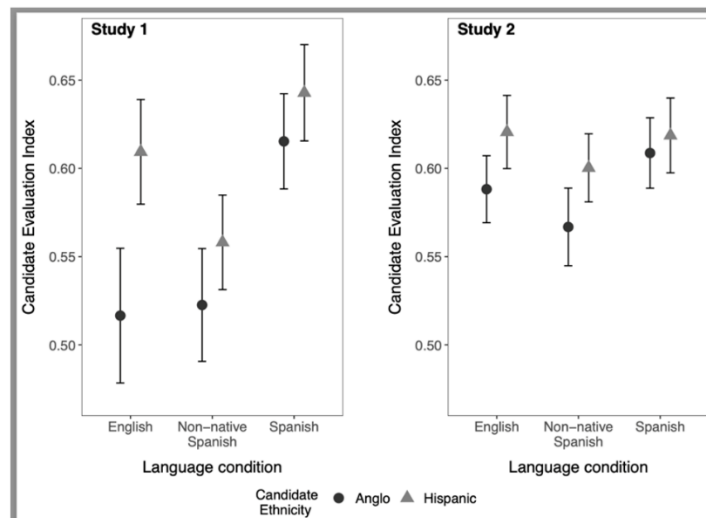
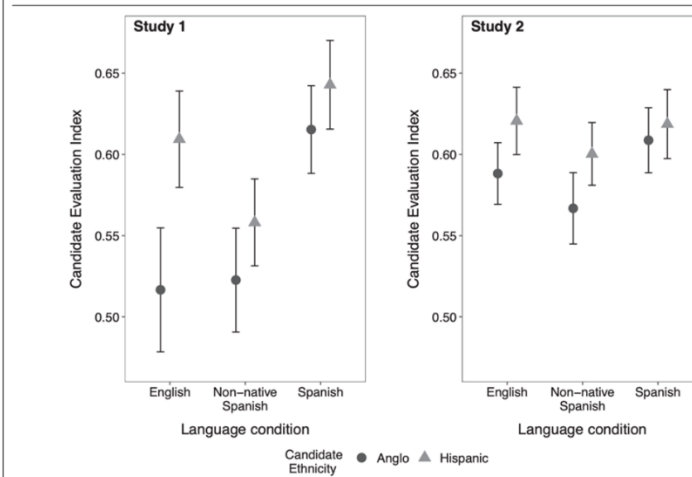


Figure 4

**Computational Reproduction****Original Paper****FIGURE 4. Mean Candidate Evaluation among Anglos by Treatment Group**

Note: Results from the Prolific sample are shown on the left; Lucid sample is shown on the right. Group means are plotted with 84% confidence intervals. Dependent variable ranges from 0 to 1. Results in tabular form are shown in Table I.3 in Appendix I of the Supplementary Material.

## Tables

Table 1:

Effects of Language and Ethnicity on Candidate Evaluations among Hispanics

	<i>Dependent variable:</i>			
	Candidate Evaluations		can.eval	
	Study 1		Study 2	
	(1)	(2)	(3)	(4)
Non-native Spanish	-0.010 (0.021)	0.051* (0.029)	-0.048*** (0.016)	-0.029 (0.023)
Spanish	0.122*** (0.021)	0.189*** (0.029)	0.044*** (0.016)	0.066*** (0.023)
Hispanic Candidate		0.118*** (0.028)		0.047** (0.023)
Hispanic x Non-native Spanish		-0.122*** (0.040)		-0.039 (0.033)
Hispanic x Spanish		-0.134*** (0.040)		-0.042 (0.032)
Constant	0.502*** (0.014)	0.443*** (0.020)	0.616*** (0.012)	0.592*** (0.016)
Observations	501	501	997	997
R <sup>2</sup>	0.092	0.124	0.032	0.036
Adjusted R <sup>2</sup>	0.089	0.115	0.030	0.031
<i>Note:</i>		* p<0.1; ** p<0.05; *** p<0.01		

Table 2:

Effects of Language and Ethnicity on Candidate Evaluations among Anglos

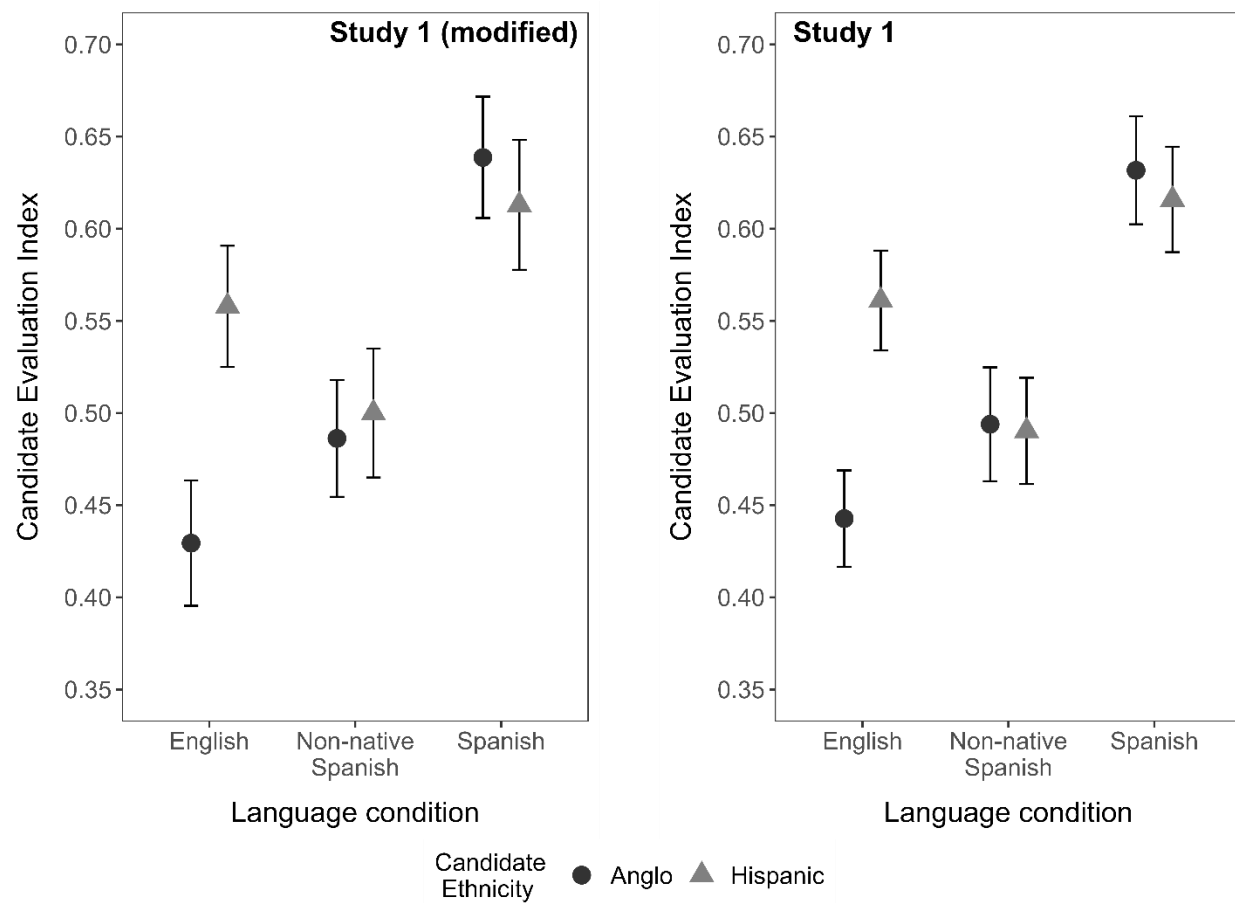
	<i>Dependent variable:</i>			
	Candidate Evaluations		can.eval	
	Study 1		Study 2	
	(1)	(2)	(3)	(4)
Non-native Spanish	-0.023	0.006	-0.020	-0.021
	(0.022)	(0.031)	(0.014)	(0.020)
Spanish	0.066***	0.099***	0.009	0.020
	(0.022)	(0.031)	(0.014)	(0.020)
Hispanic Candidate		0.093***		0.032
		(0.031)		(0.020)
Hispanic x Non-native Spanish		-0.057		0.001
		(0.043)		(0.029)
Hispanic x Spanish		-0.065		-0.022
		(0.043)		(0.029)
Constant	0.563***	0.517***	0.604***	0.588***
	(0.015)	(0.022)	(0.010)	(0.014)
Observations	503	503	996	996
R <sup>2</sup>	0.035	0.056	0.004	0.010
Adjusted R <sup>2</sup>	0.031	0.047	0.002	0.005
<i>Note:</i>		* p<0.1; ** p<0.05; *** p<0.01		

# Appendix A

## Appendix Figures

Figure A1

Comparison of Figure 1 between original study and modified version based on **definition of Latino and White groups**.



## Figure A2

Comparison between Figure 2 in original paper and modification based on **definition of Latino and White groups**.

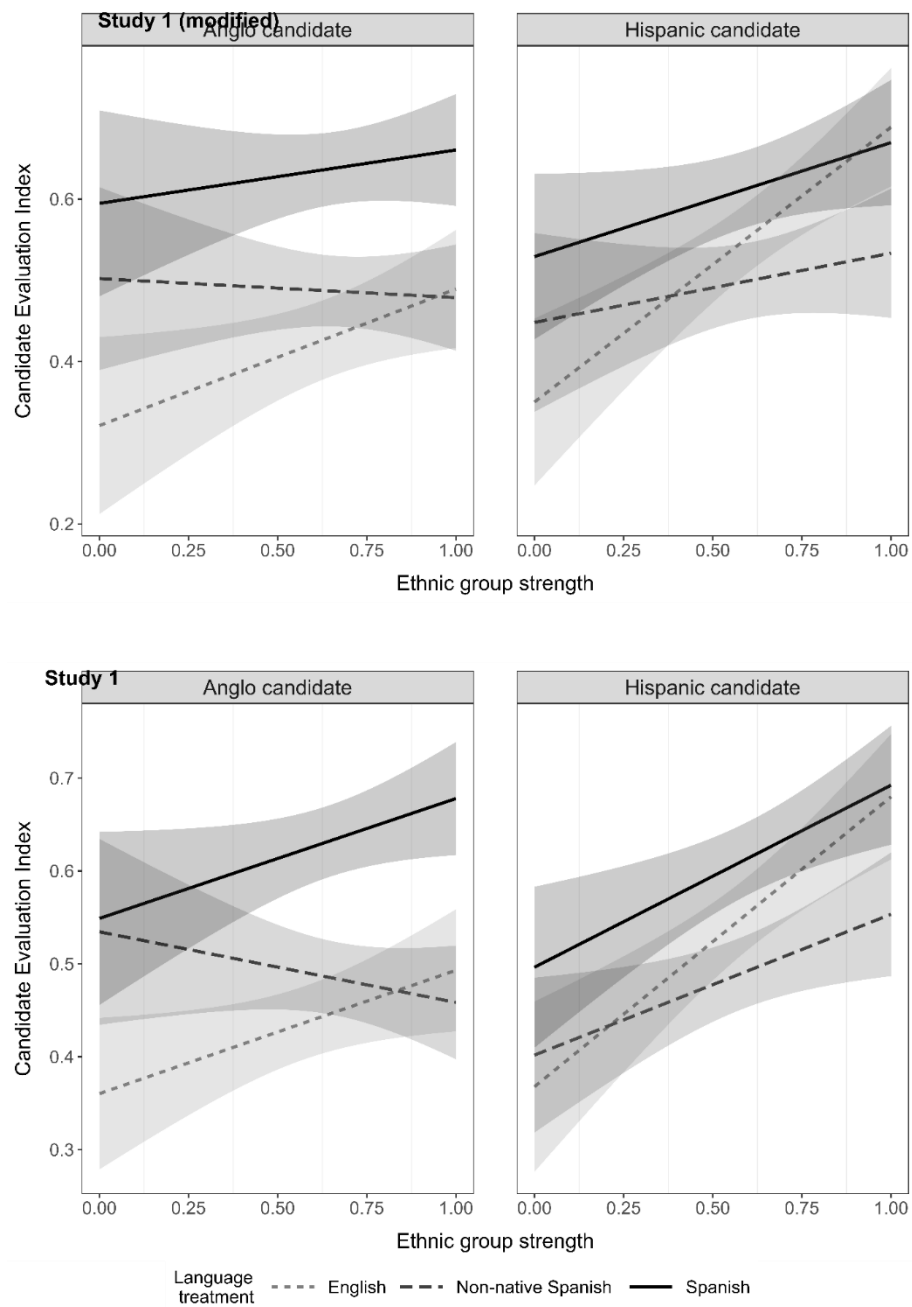
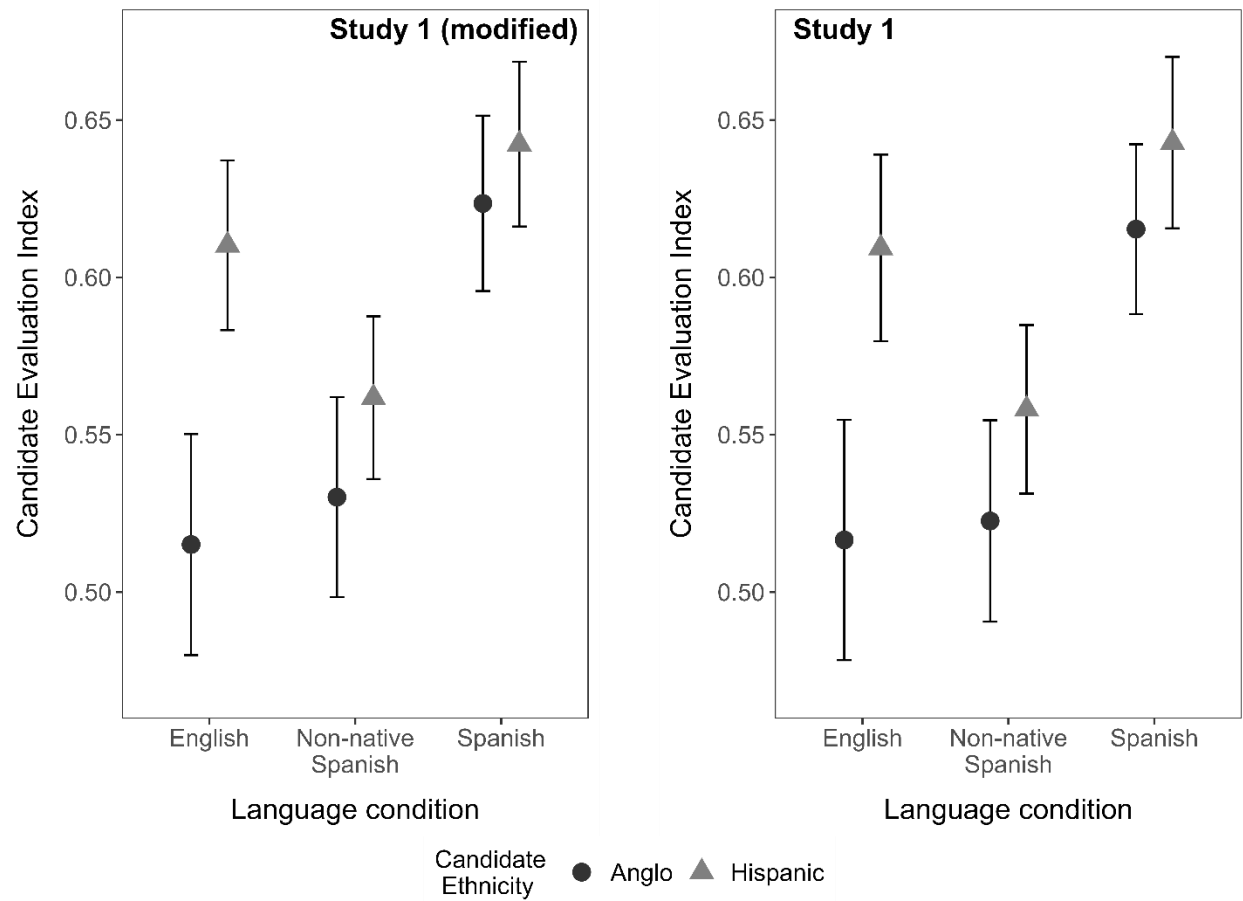




Figure A3

Comparison between Figure 4 of the original paper and our modification based on **definition of Latino and White groups.**



## Appendix Tables

Note: For each coefficient, standard errors are reported in parenthesis and p-value are reported in brackets.

Table A1:

Comparison between table 1 for the study 1 of the original paper and our modification based on **definition of Latino and White groups**.

### Effects of Language and Ethnicity on Candidate Evaluations among Hispanics

	<i>Dependent variable:</i>			
	Candidate Evaluations			
	Study 1 (modified)		Study 1	
	(1)	(2)	(3)	(4)
Non-native Spanish	-0.008 (0.024) [0.727]	0.057* (0.034) [0.098]	-0.010 (0.021) [0.623]	0.051* (0.029) [0.076]
Spanish	0.126*** (0.024) [<0.001]	0.209*** (0.035) [<0.001]	0.122*** (0.021) [<0.001]	0.189*** (0.029) [<0.001]
Hispanic Candidate		0.129*** (0.034) [<0.001]		0.118*** (0.028) [<0.001]
Hispanic x Non-native Spanish		-0.115** (0.048) [0.018]		-0.122*** (0.040) [0.003]
Hispanic x Spanish		-0.154*** (0.048) [0.002]		-0.134*** (0.040) [0.001]
Constant	0.501*** (0.017) [<0.001]	0.429*** (0.026) [<0.001]	0.502*** (0.014) [<0.001]	0.443*** (0.020) [<0.001]
Observations	369	369	501	501
R <sup>2</sup>	0.094	0.130	0.092	0.124
Adjusted R <sup>2</sup>	0.089	0.118	0.089	0.115

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01



Table A2:

Comparison between table 2 for the study 1 of the original paper and our modification based on **definition of Latino and White groups**.

### Effects of Language and Ethnicity on Candidate Evaluations among Anglos

	<i>Dependent variable:</i>			
	Candidate Evaluations			
	Study 1 (modified)		Study 1	
	(1)	(2)	(3)	(4)
Non-native Spanish	-0.017 (0.021) [0.406]	0.015 (0.029) [0.607]	-0.023 (0.022) [0.298]	0.006 (0.031) [0.845]
Spanish	0.070*** (0.021) [0.001]	0.108*** (0.030) [<0.001]	0.066*** (0.022) [0.003]	0.099*** (0.031) [0.002]
Hispanic Candidate		0.095*** (0.029) [0.002]		0.093*** (0.031) [0.003]
Hispanic x Non-native Spanish		-0.064 (0.041) [0.126]		-0.057 (0.043) [0.185]
Hispanic x Spanish		-0.076* (0.041) [0.066]		-0.065 (0.043) [0.131]
Constant	0.563*** (0.015) [<0.001]	0.515*** (0.020) [<0.001]	0.563*** (0.015) [<0.001]	0.517*** (0.022) [<0.001]
Observations	530	530	503	503
R <sup>2</sup>	0.035	0.058	0.035	0.056
Adjusted R <sup>2</sup>	0.032	0.049	0.031	0.047

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table A3:

**Effects of Language and Ethnicity on Candidate Evaluations among Hispanic with Recreate Reproduction**

	<i>Dependent variable: Candidate Evaluation</i>			
	Original Study		Recreate Reproduction	
	(1)	(2)	(3)	(4)
Non-native Spanish	-0.010 (0.021) [0.623]	0.051* (0.029) [0.076]	-0.008 (0.016) [0.623]	0.041 (0.023) [0.076]
Spanish	0.122*** (0.021) [<0.001]	0.189*** (0.029) [<0.001]	0.097*** (0.016) [<0.001]	0.151*** (0.023) [<0.001]
Hispanic Candidate		0.118*** (0.028) [<0.001]		0.095*** (0.023) [<0.001]
Hispanic x Non-native Spanish		-0.122*** (0.040) [0.003]		-0.098*** (0.032) [0.003]
Hispanic x Spanish		-0.134*** (0.040) [0.001]		-0.107*** (0.032) [0.001]
Constant	0.502*** (0.014) [<0.001]	0.443*** (0.020) [<0.001]	0.602*** (0.012) [<0.001]	0.554*** (0.016) [<0.001]
Observations	501	501	501	501
R <sup>2</sup>	0.092	0.124	0.092	0.124
Adjusted R <sup>2</sup>	0.089	0.115	0.089	0.115

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table A4

**Effects of Language and Ethnicity on Candidate Evaluations among Hispanic with Recreate Reproduction and for participants passing manipulation checks: they correctly identified 1) the candidate's ethnicity, and 2) the language spoken by the candidate.**

	<i>Dependent variable: Candidate Evaluation</i>			
	Original Study		Recreate Reproduction	
	(1)	(2)	(3)	(4)
Non-native Spanish	-0.010 (0.021) [0.623]	0.051* (0.029) [0.076]	-0.005 (0.018) [0.773]	0.023 (0.025) [0.268]
Spanish	0.122*** (0.021) [<0.001]	0.189*** (0.029) [<0.001]	0.098*** (0.018) [<0.001]	0.153*** (0.026) [<0.001]
Hispanic Candidate		0.118*** (0.028) [<0.001]		0.089*** (0.024) [<0.001]
Hispanic x Non-native Spanish		-0.122*** (0.040) [0.003]		-0.067* (0.035) [0.058]
Hispanic x Spanish		-0.134*** (0.040) [0.001]		-0.107*** (0.035) [0.003]
Constant	0.502*** (0.014) [<0.001]	0.443*** (0.020) [<0.001]	0.599*** (0.012) [<0.001]	0.555*** (0.017) [<0.001]
Observations	501	501	418	418
R <sup>2</sup>	0.092	0.124	.09	.122
Adjusted R <sup>2</sup>	0.089	0.115	.086	.112

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table A5

**Effects of Language and Ethnicity on Candidate Evaluations among Hispanic with Recreate Reproduction and for participants passing manipulation checks as well as matching ethnicity self-identification with Prolific's identification.**

	<i>Dependent variable: Candidate Evaluation</i>			
	Original Study		Recreate Reproduction	
	(1)	(2)	(3)	(4)
Non-native Spanish	-0.010 (0.021) [0.623]	0.051* (0.029) [0.076]	-0.002 (0.019) [0.910]	0.037 (0.026) [0.161]
Spanish	0.122*** (0.021) [<0.001]	0.189*** (0.029) [<0.001]	0.103*** (0.019) [<0.001]	0.154*** (0.027) [<0.001]
Hispanic Candidate		0.118*** (0.028) [<0.001]		0.091*** (0.026) [<0.001]
Hispanic x Non-native Spanish		-0.122*** (0.040) [0.003]		-0.078** (0.038) [0.039]
Hispanic x Spanish		-0.134*** (0.040) [0.001]		-0.100*** (0.037) [0.008]
Constant	0.502*** (0.014) [<0.001]	0.443*** (0.020) [<0.001]	0.592*** (0.013) [<0.001]	0.546*** (0.018) [<0.001]
Observations	501	501	383	383
R <sup>2</sup>	0.092	0.124	.094	.124
Adjusted R <sup>2</sup>	0.089	0.115	.089	.112

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01