Language Bias in Court

Marco Repetto

Department of Economics

University of California, Santa Barbara

Faculty Mentor: Dr. Shelly Lundberg

Undergraduate Research and Creative Activities (URCA) Grant

June 13, 2025

Abstract

In the US court system, interpreters are provided for anybody who does not speak English sufficiently proficiently to ease language difficulty and equalize opportunity under the law. Despite that precaution, jurors still hear the original language being translated, which could create extrajudicial biases in court cases. Thus far, there has been limited literature on whether language creates a juror bias and whether one language creates more bias than another. These studies have also been limited in scope and conflicted in findings, with some finding evidence of bias while others not. In order to fill this gap, I designed a pilot study to model for a larger scale study if biases were found. This included recording testimony for a fake court case with dubbed voices to maintain accuracy. I then created a survey asking for ratings of guilt, understanding, credibility, and reasoning which I sent out to a representative sample of the US residents. My findings show that while defendant language does not significantly impact the outcome of a clearly undisputed trial, it may influence the outcome in a highly contested one.

Background/Introduction

A defendant's right to fundamental fairness and due process in criminal proceedings is enshrined in the US Constitution through the Fifth, Sixth, and Fourteenth Amendments. Given that approximately 8.3% of people have Limited English Proficiency (LEP)(U.S. Census Bureau, 2023), these amendments have often been understood to require interpreters for those defendants (National Crime Victim Law Institute, 2024). Beyond the Constitution, a litany of state and federal laws ensures the right to an interpreter. For this reason, it is estimated that 225,000 judicial proceedings use interpreters each year in over 125 languages. While these interpreters do aid in mending the legal inequality that arises from language difficulty, the question of to what extent still stands.

In the US, court proceedings are typically interpreted using consecutive instead of simultaneous interpreting. Under consecutive interpreting, interpreters let the speaker talk uninterrupted and then repeat the statement in the target language. This means that when interpreting a witness on the stand, everyone in the courtroom hears the original statement in the foreign language and the interpreted version in English. Interpreters follow a code of ethics that obligates them to provide a complete, accurate, and impartial rendering of the original statement and are required to note when they make a mistake. For federal courts, there are two levels of interpreters in the US – Federally Certified Court Interpreters and Otherwise Qualified Interpreters (Administrative Office of the United States Courts, 2024). Federally certified interpreters have to pass a test, while otherwise qualified interpreters are approved in cases where no certified interpreter is reasonably available. This includes cases when the court has not developed a certification program for that language. On the state level, certification varies with each state having different certifications and processes.

In the past, there have been several related studies and court cases relating to linguistic bias in courtrooms. The first relevant study was in 1986 when two researchers at New Mexico State University studied the effects of Thai and Spanish translation on a mock assault case (Stephan & Stephan, 1986). To isolate language bias, the interpreters had Thai and Spanish defendants give testimony in both the foreign language and in English. A control group was also included, in which a white person gave the same testimony in English. They found that in both cases, a bias did exist against the translated defendants, with a stronger bias existing against the Spanish-speaking defendants, but also found the bias to be strongly reduced by a judge's jury instruction. The study also found that those who spoke the same language as the defendant evaluated the defendant more favorably. However, this study was limited by several factors. Firstly, the subjects were only students of this university who received class credit for participation. Secondly, the Thai study took place in Hawai'i instead of New Mexico with an entirely different group of individuals. This likely means that the bias between Spanish and Thai trials[1] is not comparable to each other, but simply shows that bias does exist. Several years later the Supreme Court held in Hernandez v. New York that a juror could be dismissed for speaking the same language as translated testimony (*Hernandez v. New York*, 1991). The reasoning was that jurors would have a different interpretation of the case since they would not have to take the interpreter's word like other jurors would have to. In contrast to Stephan and Stephan, a trial in 2011 found that Spanish-speaking and Spanish-accented speakers faced no negative bias and were often even awarded higher damages. A key qualification to this study, however, is that the accented and foreign language speakers were the plaintiffs in a civil matter instead of the

[1] Spanish and Thai trials in this case refer to trials in which the defendants speak Spanish and Thai, not that the trials were held in those languages. In many cases throughout the article, I will refer to a trial by the language of the defendant despite all trials being held in English.

defendants in a criminal matter. Along with that, the demographics of the jury were only representative of the Dallas Metropolitan area, which has proportionally more Hispanics than the US average (Shuman et. al, 2011). The authors also noted that they had a higher proportion of Women and educated people, which they claimed could be a reason for the outcomes of the study. Lastly, a recent study in Australia found that the competency of the interpreter has a large effect on juror perception (Hale et. al, 2024). They also found there to be language bias, but claimed it could be due to the difference in interpreter competency. This trial also did not account for other factors arising from the defendant that could influence the outcome, such as tone or appearance.

While results on this topic are inconclusive, closely related topics could give reason to believe that this bias does exist. Several studies have shown that accents and ethnicities have an impact on the perception of the defendant and the outcome of a case (Cantone et. al, 2019; Kurinec & Weaver, 2019). This leads me to hypothesize that a language bias does exist for juror trials and can vary from language to language. I believe that jurors most likely favor English-speaking defendants and disfavor Spanish-speaking defendants because of their connection with Latinos in the US, who are stereotypically associated with crime and low socioeconomic status (Kurinec & Weaver, 2019). My null hypothesis, therefore, was that there would be no statistically significant differences in defendant guilt or trust across the different trials.

Methodology

I designed a pilot study to model for a larger-scale study if biases are found and to serve as a model for future research in isolating a single variable. To do this, I created a survey with a fake court case of an alleged auto theft and questions on guilt and other factors. The languages I chose

to test in this trial were English, Spanish, French, and Chinese (Mandarin) since they are the most spoken languages in the US (Dietrich & Hernandez, 2022). I created the survey in Qualtrics and distributed it using Prolific to a representative sample of 330 US adults. The Qualtrics survey would randomly distribute each participant to hear one of the four languages.

This included a summary of the case, written opening statements from the defense and prosecution attorneys, written jury instructions, and a recording of the defendant's testimony. The evidence brought against the man who stole the car included a GPS signal on his phone, blurry surveillance footage, and an anonymous tip. The defense's opening statement focused on the lack of concrete evidence, while the prosecution's focused on the accumulation of circumstantial evidence. Participants were also shown a still image of the security footage depicting an individual near the vehicle. The jury instructions were brief but included 5 sections instructing participants on Presumption of Innocence and Burden of Proof, Charges and Elements, Evaluating Evidence and Identity, Defendant's Rights, and Verdict.

In order to isolate the bias against the LEP defendant and save resources, the only testimony I included in the trial was that of the defendant. The testimony included direct questioning from the defense attorney and cross-examination from the prosecution attorney, but no redirect examination from the defense attorney. The testimonies were only audio recorded but were overlaid onto a black screen with titles that displayed who was speaking at what time. Using audio recording instead of video recording helped eliminate the possibility of other biases, such as appearance, expressions, or tone that might vary when creating several video recordings. I recorded the audio for each attorney, the defendant, and the interpreter separately and stitched them together in an editing application. I then took the same recordings of the defendant and interpreter in English and ran them through an audio dubbing software to maintain the same tone

and voice across all languages. Audio dubbing software, while relatively novel technology, has made great advancements in its capabilities with the ability to accurately replicate voice, tone and content. Despite this, the software is still limited in its ability to replicate nuanced phrases or highly technical language (Oni 2025). To account for this, I consulted natives of each language and edited the translation recordings to accurately replicate the English version. Therefore, the attorney recordings were the exact same between all four versions, and the English section of the interpreter recording was the exact same across the three versions with foreign language defendants. With these factors accounted for, the only difference between the several trials was the language spoken by the defendant on the stand.

After hearing the trial, participants were asked various questions about their perception of the trial. The first questions I asked were comprehension check questions to ensure data quality. These included one question about the facts of the case and one about the presumption of innocence from the jury instructions. Both questions were straightforward and did not require high levels of legal or technical comprehension. Failure to answer either question correctly would indicate inattention or misinterpretation rather than a lack of technical understanding. After the comprehension check questions, I asked participants how likely they were to return a guilty verdict on a scale from one to ten, where one was very unlikely and ten was extremely likely. On the same page, I asked participants to rate how trustworthy they perceived the defendant to be, how credible they perceived the interpreter to be, and how well they comprehended the testimony. Each of these questions was on a scale of one to five, with one being very poor and five being very positive with respect to each question. On the next page, I asked participants to provide a written explanation for why they voted on the guilt of the defendant in the way they did. Lastly, I asked if the participants spoke more than one language,

and if so, which ones, what race they were, what race they perceived the defendant to be, and how often they interacted with individuals whose primary language is not English.

To analyze the data, I first cleaned out all responses that incorrectly answered either one of the comprehension check questions.[2] I then analyzed if there was a significant difference in guilt, trust, comprehension, and interpreter credibility between languages using a one-way ANOVA test. For the questions on multilingualism, participants' race, perceived race of the defendant, and frequency of interacting with non-English speakers, I used a combination of linear regressions and bivariate analysis to assess their relationship with guilt ratings.

Lastly, for the free response question, I created a Python script and used the Chat GPT API to run each free response through a Chat GPT 4.0 prompt that would categorize the response into one or more categories I created. I created these bins to represent the 4 sentiments that were mentioned in responses for guilt rating – Lack of Proof, Uncertainty, Legal Standard, or Credibility Assessment. I then created a joint distribution table summarizing the amount of each category per language. I used the binary independent variables in a series of ordinary least squares regressions with guilt rating as the dependent variable. To assess whether reasoning patterns varied systematically by trial language, I also created linear probability models and logistic regressions for each language group. This allowed for easy interpretation of whether the language was statistically predictable based on the participants' reasoning for guilt.

[2] There were several which answered it incorrectly but were still kept because they displayed a sufficient knowledge of the facts of the case in the free response question. You should report how many you ended up throwing out.

Results

Language Effect on Guilt

After cleaning the data, the total sample size was 301 responses, with the breakdown being 75 responses for the English trial, 72 for the Chinese trial, 78 for the French trial, and 76 for the Spanish trial. The average guilt rating on the scale from 1-10 across all languages amongst those who answered both comprehension check questions correctly was 2.454 with a standard deviation of 2.20. On average, this means that people were unlikely to find the defendant in this case guilty. The averages in guilt rating per language were 2.28 for English (SD=2.08), 2.39 for Chinese (SD=2.12), 2.51 for French (SD=2.35), and 2.61 for Spanish (SD=2.27). The average guilt rating for the trials with interpreters was 2.50 (SD=2.24), and when comparing them versus the English trial, I found the t-statistic to be -0.79 and the p-value to be 0.43 (d=0.10). When comparing across all languages together, the ANOVA test gave a F-statistic of 0.31 and a p-value of 0.82 (η^2=0.003). When comparing Spanish versus English, the t-statistic is -0.92 and the p-value is 0.36 (d=0.15). These results indicate that there were no statistically significant differences in guilt ratings across the different language trials. With respect to the free response questions, below I have added the table displaying the number of responses in each category per language.

Frequencies of Reasoning Categories Used in Guilt Explanations by Trial Language				
	Credibility	Lacking Proof	Legal Standard	Uncertainty
Chinese	16	62	17	42
English	13	65	21	44
French	13	63	17	43
Spanish	20	62	23	37

A key note from the table is the large majority of responses including the Lacking Proof sentiment.

To find whether the language influenced juror reasoning in deciding bias, I ran a series of linear probability and logarithmic models, regressing the presence of specific reasoning bins on the trial language. The results showed that none of the reasoning bins were significant predictors of the defendant language. The only somewhat noticeable predictor for any language was Credibility for the Spanish defendant trials. While not significant, it was still distinguishable when compared to the other predictors ($p=0.192$). This suggests the possibility that jurors were more likely to question the defendants credibility when deciding on the guilt of the Spanish trial. Tables for each of the linear probability and logarithmic models are in the Appendix.

Language Effect on Other Outcomes

To examine whether testimony language influenced juror perceptions of trust, comprehension, and interpreter credibility, I conducted separate one-way ANOVAs for each outcome. The first question we asked was to test whether the language of the defendant had an effect on their trustworthiness. The average rating of defendant trustworthiness, on a scale of one to five where one was extremely untrustworthy and five was extremely trustworthy, was 3.28 ($SD=0.85$).

Amongst the different languages, the average ratings were 3.28 for English (SD=0.94), 3.32 for Chinese (SD=0.78), 3.22 for French (SD=0.85), and 3.32 for Spanish (SD=0.85). This indicates that the defendants were viewed rather neutrally with a slight inclination towards more trustworthy. The ratings all showed minimal variation, with 3 of them being within 0.04 points of each other, and French being within 0.1 of Chinese and Spanish. The average trust rating of all the non-English trials together was 3.28 with a standard deviation of 0.83. Since the English and non-English trials had the same average trust rating, the t-statistic is -0.016 and the p-value is 0.987 (d=0.002) – displaying no significant difference. The ANOVA test of trust ratings between languages displayed a similar finding with the F-statistic being 0.22 and the p-value being 0.88 (η^2=0.002).

Next, we asked for participants from the three non-English trials to rate the credibility of the interpreter. The scale of this question was the same as the last, with five being extremely credible and one being extremely noncredible. The average rating amongst all the responses was 4.16 with a standard deviation of 0.89, indicating that, on the aggregate, people perceived the interpreter to be rather credible. The averages of the interpreter credibility for each language group were 4.08 for Chinese (SD=0.86), 4.13 for French (SD=0.94), and 4.25 for Spanish (SD=0.87). The ANOVA test reveals an F-statistic of 0.71 and a p-value of 0.49 (η^2=0.0063). So while the difference between the groups is slightly larger than the difference between groups in trust ratings, it still is not statistically significant.

Lastly, we asked participants to rate their comprehension of the testimony on the same one to five scale. The average comprehension across all languages was 4.35 (SD=0.76), indicating a high overall level of comprehension. Amongst the different languages, the averages were 4.49 for English (SD=0.70), 4.23 for Chinese (SD=0.75), 4.27 for French (SD=0.85), and 4.42 for

Spanish (SD=0.72). The average across the three interpreted trials (Chinese, French, and Spanish) was 4.31 (SD=0.78), indicating that most people claimed to comprehend less of the interpreted testimonies compared to the English-only testimonies. When comparing the English comprehension versus the non-English, I found the t-statistic to be 1.92 and the p-value to be 0.057 (d=0.24). This displays a statistically significant difference in comprehension between the two groups at a 0.10 threshold. Amongst all languages, the F-statistic is 1.98 and the p-value is 0.116 (η^2=0.020). Therefore, while the difference is noticeable, there is no statistically significant difference in testimony comprehension across the different trials.

Participant-Level Predictors of Guilt

I also asked for a variety of other factors to test if there were other predictors of guilt as well. I asked participants what race they perceive the defendant in the trial to be, what race they are, if they are multilingual, and lastly, how often they interact with LEP individuals. For the question asking participants what race they perceived the defendant to be, there was little variation within each language group, with most participants selecting the same race for a given language. Below, I have added a table displaying the distribution of race perceptions per language group. Because of the little variation per language, it was not possible to determine if race perception had a significant impact on guilt rating apart from language.

Variation In Perceived Race of the Defendant for Each Language Group							
Language	White	Black or African American	American Indian or Alaska Native	Asian	Native Hawaiian or Pacific Islander	Hispanic or Latino	Middle Eastern or North African
Chinese	10 (13.9%)	3 (4.2%)	0 (0.0%)	54 (75.0%)	3 (4.2%)	1 (1.4%)	1 (1.4%)
English	51 (68.0%)	16 (21.3%)	0 (0.0%)	1 (1.3%)	4 (5.3%)	0 (0.0%)	3 (4.0%)
French	39 (50.0%)	13 (16.7%)	0 (0.0%)	4 (5.1%)	8 (10.3%)	6 (7.7%)	8 (10.3%)
Spanish	7 (9.2%)	5 (6.6%)	1 (1.3%)	3 (3.9%)	60 (78.9%)	0 (0.0%)	0 (0.0%)

Race of the participant, on the other hand, was not directly tied to the language spoken by the defendant. Table of the distribution found in the Appendix. This allowed me to analyze if people of different races voted on guilt differently. The average guilt rating from White participants was 2.59 (SD=2.33), and the average guilt rating for non-white participants was 2.33 (SD=1.92). The t-statistic is 1.62 and the p-value is 0.11 (d=0.19), which is not statistically significant. I then filtered out the races that had very few responses and ran an OLS regression accounting for language and race to find an effect on guilt. Below is the summary of the regression findings. The regression shows that individually, no language or race had a statistically significant effect on the guilt rating of the defendant.

OLS Regression of Guilt Rating on Participant Race and Trial Language			
Predictor	Coefficient	t-statistic	p-value
Intercept (White, English)	2.42	8.74	< .001
Chinese (vs. English)	+0.10	0.26	0.79
French (vs. English)	+0.26	0.69	0.49
Spanish (vs. English)	+0.29	0.76	0.45
Black	+0.02	0.05	0.96
Asian	–0.40	–0.76	0.45
Hispanic or Latino	–0.59	–1.33	0.18

In response to the multilingual question, 53 participants responded that they speak more than one language. The average guilt rating for multilingual participants was 2.55 (SD=2.23) compared to 2.43 for single language participants (SD=2.20). The t-test gave a t-statistic of -0.36 and a p value of 0.72 (d=0.05). This shows that multilingualism had no significant effect on the guilt rating of the defendant. There were also 16 participants who had reported speaking the same language as the language of the defendant. The average guilt rating among them was 2.31 (SD=1.99). When comparing whether they are statistically different in their guilt rating, I found a t-statistic of -0.26 and p-value of 0.79 (d=-0.06), showing no significant difference in guilt rating between those who speak the same language and those who do not.

Lastly, the average value for how often the participant speaks to an LEP individual is 2.81 (SD=0.84). There were 4 options for this question, from Never to Often. Options 2 and 3 were Rarely and Sometimes, meaning that the average response was in between those two answers, but leaning more towards Sometimes. I found a -0.094 correlation coefficient between frequency of speaking with an LEP individual and guilt rating with a t statistic of -1.64 and a p value of

0.10 (r^2=0.009). This means that while the correlation between frequency of speaking with an LEP individual and guilt rating may be weak, it is still statistically significant.

Analysis

At face value, most of the results received in this study were not statistically significant, yet many still provide meaningful insight as to why and how to proceed. Language, and speaking the same language as the defendant, all showed to have no statistically significant effect on guilt rating. This is in contrast to other studies in which ingroup bias was proven to have an effect on guilt rating (Stephan & Stephan, 1986). A key explanation for this could be the design of the fake trial, however. With a mean guilt rating of 2.45 out of 10, people were much more inclined to vote not guilty rather than guilty. The open-ended responses back up this point with a large majority of each language group citing Lack of Evidence as part of their reason for voting in the way they did. With such a large inclination towards one side of the ruling, it appears that people were clearly set in their choices. This could therefore outweigh any potential bias that may exist, as a bias would have to prevail over an easily decisive factor. This is supported by the liberation hypothesis, set forth by Kalvin and Ziesel in 1966 (Kalven & Zeisel, 1966). This theory supports that when the evidence in a case is less conclusive, jurors are more likely to deviate from the facts of the case when deciding the verdict.

Despite this, there was still some variation for each category, which could serve as potential insight despite the skewed data. The average guilt ratings, for example, followed the order I set forth in my hypothesis, with the English trial being the least guilty and the Spanish trial being the most guilty. The open-ended responses also show that more people cited the credibility of the Spanish-speaking defendant when making choices, despite it not being a significant predictor of

guilt. Similarly, those who spoke the same language as the defendant did have a lower average guilt rating than those who did not. So while it is also not statistically significant, it is not entirely inconsistent with the Stephan and Stephan study of ingroup bias. The order of languages and non-statistically significant results does not provide reliable evidence of bias, however. As of now, the data does not support the idea that a language bias exists in criminal juror trials, and any apparent trends should be interpreted with caution. The question remains open pending further testing with a more balanced trial design and larger samples.

Apart from the predictors of guilt, language also had little effect on trust and interpreter credibility. The small effect on interpreter credibility is expected and displays a successful use of the dubbing technology to accurately control for factors across languages. The little effect on trustworthiness is in contrast to the case in Stephen and Stephen but does also reflect on the ability of the dubbing technology to control for factors. Lastly, the perceived level of comprehension did significantly vary between English and non-English testimony. This, however, is inconclusive as it could result from a misinterpretation of the question asked. Given that across all non-English trials, the English section of the interpreter speaking was the same audio clip, it is reasonable to assume that various participants voted based on how well they understood the defendant instead of the testimony. Especially since the order of comprehension ratings followed roughly the order of linguistic similarity to English with the exception of Spanish being comprehended more than French because of its overwhelming prevalence in US culture and society (Chiswick & Miller 2004).

Conclusion

This pilot study represents a possible way for future researchers to isolate specific variables in court and test for biases. The coupling of isolated audio recordings and open-ended response analysis allowed me to gather valuable insight despite difficulties in trial design. Despite not receiving statistically significant results that language influenced guilt rating, the differences and order of average guilt rating give reason to believe that the bias may exist in a more balanced trial. The same applies to several other variables, such as multilingualism, frequency of interaction with LEP individuals, or participant race. This, therefore, opens the door to future studies examining this same question with greater scrutiny. In order to gather valid results, I believe future researchers could replicate the study with a less skewed case or examine court data to compare aggregate rates of guilty verdicts across trials with different languages.

This trial comes with several other limiting factors as well. Firstly, isolating variables could ignore other factors that take place in a real courtroom throughout an entire case. It is possible that with the length and complexity of real in-person court cases, biases are heightened or nullified in comparison to a single person watching a portion of a case. Being on a panel of jurors where discussion is necessary may also influence results in a full-length case. The inherent connection between defendant language and perceived race also created another limitation. Without this variation, it was impossible to separate the effects of the perceived race from the effects of the language. Additionally, using a representative sample of the US may make the results applicable on aggregate across the country, but could limit their practicality as juror panels are gathered from the area of the courthouse. This means that individual cases would have different juror sample demographics from each other and from this study, which could therefore provide different results.

Nevertheless, it is imperative that this question get a definitive answer as it risks the justice of thousands of individuals every year. If a bias is found, it would violate the right to a fair trial enshrined in our constitution and would necessitate governmental action.

Bibliography

1. Administrative Office of the United States Courts. (2024). *Federal court interpreter orientation manual and glossary* (Rev. ed.). https://www.uscourts.gov/rules-policies/judiciary-policies/court-interpreting-guidance
2. Barreto, M. A., Manzano, S., & Segura, G. M. (2012). *The impact of media stereotypes on opinions and attitudes towards Latinos*. National Hispanic Media Coalition & Latino Decisions. https://www.latinodecisions.com
3. Cantone, J. A., Martinez, L. N., Willis-Esqueda, C., & Miller, T. (2019). Sounding guilty: How accent bias affects juror judgments of culpability. *Journal of Ethnicity in Criminal Justice*, *17*(3), 228–253. https://doi.org/10.1080/15377938.2019.1623963
4. Chiswick, B. R., & Miller, P. W. (2004). *Linguistic distance: A quantitative measure of the distance between English and other languages* (IZA Discussion Paper No. 1246). Institute for the Study of Labor (IZA). https://docs.iza.org/dp1246.pdf
5. Dietrich, S., & Hernandez, E. (2022). *Language use in the United States: 2019* (ACS-50). U.S. Census Bureau. https://www.census.gov/library/publications/2022/acs/acs-50.html
6. Hale, S., Martschuk, N., Goodman-Delahunty, J., & Lim, J. (2024). Juror perceptions in bilingual interpreted trials. *Perspectives: Studies in Translation Theory and Practice*, 1–24. https://doi.org/10.1080/0907676X.2024.2343772
7. *Hernandez v. New York*, 500 U.S. 352 (1991).
8. Kalven, H., & Zeisel, H. (1966). *The American jury*. Little, Brown.
9. Kurinec, C., & Weaver, C. (2019). *Dialect on trial: Use of African American Vernacular English influences juror appraisals*. *Psychology, Crime & Law, 25*, 1–53. https://doi.org/10.1080/1068316X.2019.1597086
10. National Crime Victim Law Institute. (2024). *Interpreters during court proceedings: A requirement for the meaningful exercise of rights and access to justice for victims in need of language assistance.* https://ncvli.org Administrative Office of the United States Courts. (2024). *Federal court interpreter orientation manual and glossary* (Rev. ed.). https://www.uscourts.gov/rules-policies/judiciary-policies/court-interpreting-guidance
11. Oni, S. B. (2025, April). *The impact of AI on the translation industry*. https://www.researchgate.net/publication/391050035_The_Impact_of_AI_on_the_Translation_Industry
12. Shuman, D. W., Stokes, L., & Martinez, G. (2011). *Stranger at the gate: The effect of the plaintiff's use of an interpreter on juror decision-making*. *Behavioral Sciences & the Law, 29*, 499–512. https://doi.org/10.1002/bsl.985
13. Stephan, C., & Stephan, W. G. (1986). *Habla inglés? The effects of language translation on simulated juror decisions*. *Journal of Applied Social Psychology, 16*, 577–589. https://doi.org/10.1111/j.1559-1816.1986.tb01160.x
14. U.S. Census Bureau, U.S. Department of Commerce. (2023). Selected Social Characteristics in the United States. *American Community Survey, ACS 5-Year Estimates Data Profiles, Table DP02*. Retrieved May 10, 2025, from https://data.census.gov/table/ACSDP5Y2023.DP02.

Appendix

Linear Probability Model for English			
Predictor	Coefficient	P val	T stat
Intercept	0.228	0.0014	3.22
Lacking proof	0.014	0.843	0.20
Uncertainty	0.026	0.643	0.48
Credibility	-0.042	0.517	-0.65
Legal standard	0.024	0.697	0.39

Logarithmic Regression for English			
Predictor	Coefficient	P val	Z val
Intercept	-1.23	0.0016	-3.16
Lacking proof	0.084	0.832	0.21
Uncertainty	0.138	0.629	0.48
Credibility	-0.236	0.508	-0.66
Legal standard	0.125	0.692	0.40

Linear Probability Model for French			
Predictor	Coefficient	P val	T stat
Intercept	0.341	0.000003	4.78
Lacking proof	-0.056	0.441	-0.77
Uncertainty	-0.009	0.865	-0.17
Credibility	-0.081	0.215	-1.24
Legal standard	-0.055	0.369	-0.90

Logarithmic Model for French			
Predictor	Coefficient	P val	Z val
Intercept	-0.63	0.079	-1.75
Lacking proof	-0.29	0.434	-0.78
Uncertainty	-0.05	0.855	-0.18
Credibility	-0.45	0.216	-1.24
Legal standard	-0.30	0.361	-0.91

Linear Probability Model for Spanish			
Predictor	Coefficient	P val	T stat
Intercept	0.252	0.0004	3.57
Lacking proof	-0.013	0.860	-0.18
Uncertainty	-0.040	0.454	-0.75
Credibility	0.084	0.192	1.31
Legal standard	0.059	0.331	0.97

Logarithmic Model for Spanish			
Predictor	Coefficient	P val	Z val
Intercept	-1.10	0.0028	-2.98
Lacking proof	-0.07	0.854	-0.18
Uncertainty	-0.22	0.447	-0.76
Credibility	0.43	0.192	1.30
Legal standard	0.31	0.325	0.99

Linear Probability Model for Chinese			
Predictor	Coefficient	P val	T stat
Intercept	0.179	0.0107	2.57
Lacking proof	0.055	0.443	0.77
Uncertainty	0.024	0.654	0.45
Credibility	0.039	0.543	0.61
Legal standard	-0.027	0.645	-0.46

Logarithmic Model for Chinese			
Predictor	Coefficient	P val	Z val
Intercept	-1.51	0.0002	-3.73
Lacking proof	0.31	0.443	0.77
Uncertainty	0.13	0.653	0.45
Credibility	0.21	0.536	0.62
Legal standard	-0.15	0.641	-0.47

Variation in Race of the Participant for Each Language Group							
Language	Race 1	Race 2	Race 3	Race 4	Race 6	Race 7	Race 8
Chinese	49 (67.1%)	6 (8.2%)	0 (0.0%)	3 (4.1%)	13 (17.8%)	0 (0.0%)	2 (2.7%)
English	53 (70.7%)	8 (10.7%)	0 (0.0%)	3 (4.0%)	7 (9.3%)	0 (0.0%)	4 (5.3%)
French	47 (60.3%)	11 (14.1%)	3 (3.8%)	8 (10.3%)	6 (7.7%)	2 (2.6%)	1 (1.3%)
Spanish	52 (68.4%)	10 (13.2%)	1 (1.3%)	7 (9.2%)	4 (5.3%)	1 (1.3%)	1 (1.3%)