

The Effects of Changing the Census Race and Ethnicity Question

An experiment looking at the effect of changing the format of the race and ethnicity question asked on the US Census to one that expands representation of minorities such as Hispanic and Middle Eastern and North African people.



Katie Mo & Kat Borojerdi

08.13.2019

MIDS W241 . Summer 2019 . Section 2

TABLE OF CONTENTS

ABSTRACT	2
BACKGROUND	2
RESEARCH QUESTION	3
EXPERIMENTAL DESIGN	4
TREATMENT AND CONTROL	4
SURVEY DESIGN	4
RANDOMIZATION PROCESS	5
DATA COLLECTION	5
THE PILOT	6
DATA	6
RESULTS	10
PRIMARY FINDINGS	10
SECONDARY FINDINGS	12
LIMITATIONS AND LESSONS LEARNED	13
REFERENCES	14
APPENDIX	15
APPENDIX A: 2010 CENSUS ETHNICITY AND RACE QUESTION	15
APPENDIX B: PROPOSED CENSUS ETHNICITY AND RACE QUESTION	16
APPENDIX C: QUALTRICS SURVEY	17
APPENDIX D: QUALTRICS SURVEY RANDOMIZATION	25
APPENDIX E: FACEBOOK AD	26
APPENDIX F: UC BERKELEY ETHNICITY CLUBS	28
APPENDIX G: TESTS FOR HETEROGENEOUS TREATMENT EFFECTS	30

ABSTRACT

The following study examined the potential impact of changing the current race and ethnicity question on the US Census Survey. By studying this topic, we evaluated whether changing the way data is collected will improve race and ethnic representation in the US Census. Based on our findings, we can not make a direct causal claim that the proposed race and ethnicity question will result in better representation in the Census data. Although a number of our findings such as identification as White and Middle Eastern and North African (MENA) had the expected direction of the treatment effect, we also observed identification as Hispanic and Other displayed treatment effects in the opposite direction of our hypothesis. However, the majority of these findings lacked statistical significance due to being underpowered. For one of our results, we found a causal effect of providing ethnic subcategory check boxes in addition to the race category, resulting in a 28% increase in disclosure of ethnic subgroups in the White race.

BACKGROUND

Every 10 years, those living in the United States are asked to take the US Census survey. Since the survey's inception in 1790, a race and ethnicity question has been included. While the demographic of the United States has evolved over time, so has the race and ethnicity question, reflecting the diverse history of this nation. To frame this study, we first provide a formal definition of race and ethnicity. While race is defined by people who have differences and similarities in biological traits (examples include Caucasian, African American, Asian, Hispanic), ethnicity refers to shared cultural practices, perspectives, and distinctions that set apart one group of people from another, often associated with specific country or region¹. It is helpful to think of race as the major category, while there may be a number of subcategories of ethnicities within each race.

Starting in the 1970s, the Census Bureau began studying the race and ethnicity question in further detail to better understand its relevance in society and to evaluate whether the classifications presented to the public makes sense. The Census Bureau's internal studies have indicated that more and more Americans do not identify with the race and ethnic categories in the most recent census² (Appendix A), and that many are confused when making a selection. As a result, the "Some Other Race" category has become the third largest group selected as a response for the race/ethnicity census question³.

The motivation behind our research project is to test an approach previously suggested by the US Office of Management and Budget (OMB) as an alternative for the race and ethnicity question to be used in the upcoming 2020 US Census (Appendix B). In the OMB's research, a suggested change to the survey question is to streamline the two separate ethnicity and race questions into one combined question. The proposed question will

include new checkboxes for Middle Eastern and North African (MENA) and Hispanic^{4,5}.

However for the time being, the Census Bureau has decided to retain the original race and ethnicity question format used in the 2010 census, citing a lack of research to make the change⁴. We would like to build on the previous research by the OMB to explore whether the proposed improvements to the race and ethnicity question could lead to better representation in the Census data.

The data from the race and ethnicity question in the census is used to ensure equal opportunity in a variety of programs, including housing, voting, language, employment, and education. In addition, this question's format serves as a standard for many forms and processes⁶. If distinct, contributing racial and ethnic communities in the United States are not properly accounted for, they may not be considered as individual groups in medical research⁷, in college or job applications, and social programs⁶. Clarifying the ethnicity question improves classification of these ethnicities, and may reduce bias where the US Census race and ethnicity question is used.

RESEARCH QUESTION

We are interested in whether changing the race and ethnicity question on the US Census allows for better representation in the Census data. The proposed version of the question will combine the Hispanic ethnicity question and the race question into a single race and ethnicity question, and will provide separate checkboxes for the Middle Eastern and North African (MENA) and Hispanic groups. We hypothesize that the proposed version of the question will reduce the number of people who select Some Other Race (SOR) and White as people in the MENA and Hispanic groups now have a representative checkbox to select. We believe that the number of people identifying as a race belonging to MENA will increase because of the addition of a specific MENA checkbox, while the number of people identifying as Hispanic will remain the same because the control question already has a Hispanic checkbox in a separate question.

We also identified secondary research questions that may also contribute to increased representation for multiracial and ethnic subgroups. Because Hispanic is not considered a race in the 2010 census question, it is not included when counting the number of people identifying as two or more races. With the proposed combined question, we hypothesize that the number of people selecting multiple races would increase.

The proposed version of the census question also provides subgroup checkboxes for the White category, whereas the current version of the census question only provides a fill-in space. We hypothesize that including these additional subgroup checkboxes will encourage people to disclose a more specific race.

The proposed census question has been designed to be less confusing and more comprehensive because it contains checkboxes with which people can readily identify

themselves. We use time spent on the race and ethnicity question as a proxy for confusion and hypothesize that the proposed census question will be quicker and easier to complete.

Collectively, the outcome variables we set out to measure help quantify what better representation would look like in various aspects.

EXPERIMENTAL DESIGN

TREATMENT AND CONTROL

Our approach was to use the current race and ethnicity question from the 2010 census as the control and the proposed version of the race and ethnicity question described as the treatment.

For the primary research question, we measured four outcome variables: White, Other, Hispanic, and MENA. The potential outcome for each outcome variable is 0 if the participant did not select that category's checkbox or 1 if the participant did select that category's checkbox. Because a MENA checkbox is not included in the control, the potential outcome for MENA in control is 1 if the participant had written a MENA race in a fill-in space or 0 otherwise.

For the secondary research questions, the outcome variables Mixed Race and White subgroups are also binarized. The potential outcome for Mixed Race is 0 if only one main group checkbox is selected or 1 if two or more main group checkboxes are selected. Out of those who select White, the potential outcome in control for White subgroups is 1 if there is a written response in the provided fill-in space or 0 otherwise. The potential outcome in treatment for White subgroups is 1 if there is a White subgroup checkbox selected or if there is a written response in the provided fill-in space, or 0 otherwise.

The potential outcome for our last outcome variable, Time, is the duration in seconds to complete the respective race and ethnicity question for both control and treatment.

SURVEY DESIGN

As we did not want to directly ask people for their race and ethnicity information because it may be a sensitive topic, we decided to conceal the purpose of the experiment by adding unrelated survey questions. We considered several options for concealing the objective of the experiment, including questions about ethnic foods, charities, or political cartoons. We chose to ask questions regarding people's ethnic food consumption habits and preferences because we believed that it would be an engaging and non-controversial topic. If the race and ethnicity questions was the focal point of the survey, we were afraid that the participants would over-analyze the questions, which might bias their responses.

Prior to our race and ethnicity question, we collected additional information for our covariates: gender, U.S. region, and family immigration history. We hypothesized that females may respond differently to the race and ethnicity question, as they may be more methodical. We also believed that different U.S. regions would have different ratios of ethnic groups. Finally, we also hypothesized that those who had recently immigrated to the US may respond differently than those who are several generations removed from when their family immigrated.

Although the US census has previously been completed by paper, the main method for conducting the US Census in 2020 will be through the internet. As such, we opted for an online survey format using the Qualtrics platform (Appendix C). Unfortunately, we had accidentally omitted the Native American and Alaska Native box into our productionalized treatment survey version, which may have obscured our measurement of the Other selection.

RANDOMIZATION PROCESS

Through the Qualtrics platform, participants were not randomized into the control and treatment versions of the race and ethnicity questions until they completed the previous covariate survey questions. We used the RXO and R_O randomization design to conduct our experiment. The “Evenly Present Elements” option was selected for the randomizer, which ensured an even number of participants were assigned to either the control or treatment survey (Appendix D). We used different survey links for each recruitment method to allow for blocked randomization.

DATA COLLECTION

To calculate the number of participants required to have sufficient statistical power, we used the treatment effect of 3% found in the prior OMB census study for respondents reporting White⁵. The calculation showed that we needed 5000 responses to achieve at least 80% statistical power. Because we needed such a high number of responses, we aimed to oversample Hispanic and MENA communities as we believe that they may have a larger treatment effect for the given study design.

Although we considered offline methods such as going to public places with a high percentage of Hispanic and MENA people, we chose online methods for increased efficiency in collecting responses. The online platforms we considered were Mechanical Turk, Facebook, Reddit and emailing UC Berkeley ethnic organizations. We eliminated Reddit because the majority of ethnic subreddits were active by those outside the US. We also decided against Mechanical Turk because research indicated that many survey takers may reside outside of the US and there was a lack of targeting capability towards Hispanic and MENA populations.

Out of the ones considered, we decided that Facebook was likely the most representative

of the general US population and provided a targeting capability. Through Facebook Ads, we used the following categories as our targeting criteria: interest in Hispanic culture or Middle East, having Hispanic multicultural affinity, and those who engage in Ramadan month content. We ran two Facebook Ad campaigns, each being a week long (Appendix E). The first campaign was advertised to the entire US population, and the second campaign was advertised only to the cities identified as having high MENA and Hispanic populations: Jersey City (+25 mi), Newark New Jersey; New York (+25 mi) New York; Wayne County Michigan; Orange County California; Los Angeles County California; Cook County Illinois.

We also reached out to the most active UC Berkeley MENA and Hispanic ethnicity clubs. After ensuring that they are still active using Facebook, Twitter or club website, we emailed 11 clubs asking whether they can forward the survey to their club members (Appendix F).

Finally, we also invited friends and family to complete our survey.

THE PILOT

Because we were unfamiliar with the mechanics of Facebook Ads, we used our pilot study to run a one day ad campaign with Mediterranean food interest as a targeting criteria in Austin, Texas, which we did not believe to have a large MENA and Hispanic population. The pilot run was valuable because we learned how to use Facebook Ads present our post to Facebook users who would click on links rather than those who would comment or only like posts.

DATA

From the 3 methods of recruitment, we had a total of 742 participants to have responses recorded in our Qualtrics survey. The flow diagram in Figure 1, shows that there was a 2.3% (17/742) non-compliance rate for those who dropped out before reaching the race and ethnicity question. The 725 participants who had reached our question of interest were randomized into the control or treatment version of the survey. We removed 3 respondents who did not reside in the U.S, as we were only interested in the responses from the U.S. population. From the remaining 722 participants, there was a 4.4% (32/722) attrition rate of those that did not complete the race and ethnicity question or had completed the survey and wrote in that they did not want to disclose their race or ethnicity. Nine participants included in the attrition rate were unknown on whether they were shown the control or treatment version because they completed the questions on the previous covariate page, but did not make a first click on the race and ethnic question page before dropping out. Ultimately, our final data set contained 344 responses from the control question and 346 responses from the treatment question.

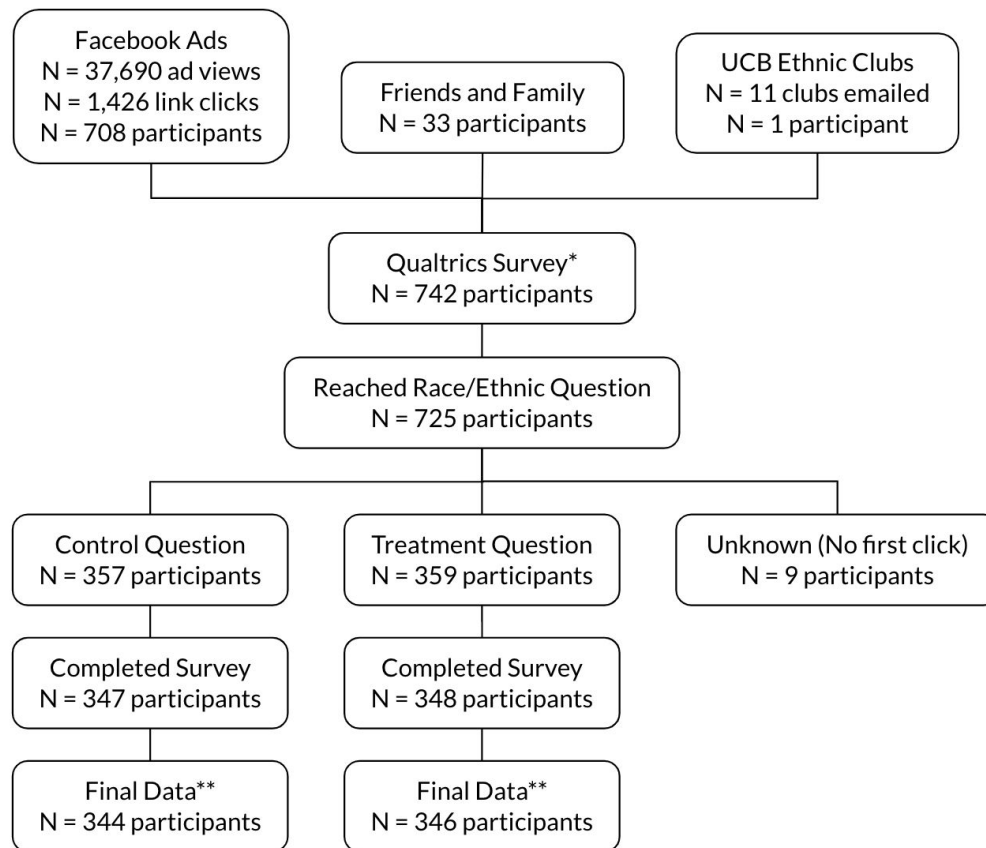


Figure 1: Ethnicity and Race Survey Flow Diagram

The flow diagram shows the number of participants at recruitment to inclusion in the final data set.

*There was a 2.3% (17/742) non-compliance rate and a 4.4% (32/722) attrition rate. *Each recruitment method had a separate survey to allow for blocked randomization. **Final data reflects participants removed who did not reside in the U.S. or did not want to disclose their race or ethnicity.*

We performed a covariate balance check to ensure the randomization done through the Qualtrics survey was executed correctly. Because our covariate questions were categorical, we one hot encoded each possible response, omitted one category to be the baseline, and regressed them on whether they were placed in control or treatment. Figure 2 shows that the covariates we captured were balanced between the control and treatment question, as none of the coefficients were statistically significant.

	<i>Dependent variable:</i>
	Treatment
Female	0.006
Region_Midwest	-0.051
Region_Northeast	-0.009
Region_South	-0.019
Immigrate_Self	0.045
Immigrate_Parents	-0.004
Immigrate_Grandparents	-0.041
Immigrate_GreatGrandparents	-0.050
Constant	0.533*** (0.060)
Observations	690
R ²	0.006
Adjusted R ²	-0.006
Residual Std. Error	0.502 (df = 681)
F Statistic	0.478 (df = 8; 681)
<i>Note:</i>	*p<0.05; **p<0.01; ***p<0.001

Figure 2: Covariate Balance Check

There were no statistically significant coefficients found when checking for covariate balance, indicating that randomization was executed correctly.

Because the race and ethnic question in the Qualtrics survey allowed for multiple selections as well as fill in responses, the resulting data columns were somewhat unstructured. New columns representing the major race and ethnic categories were created by binarizing whether they were selected or not. A correction was applied to the treatment group, as some participants had only selected a subgroup ethnicity checkbox and did not select the accompanying major race group category. We made the assumption that if participant selected the ethnicity subgroup checkbox, then they also would have selected the major race group category that the subgroup checkbox belonged to. The distribution of the major race and ethnic category responses in the Control and Treatment groups are shown in Figure 3.

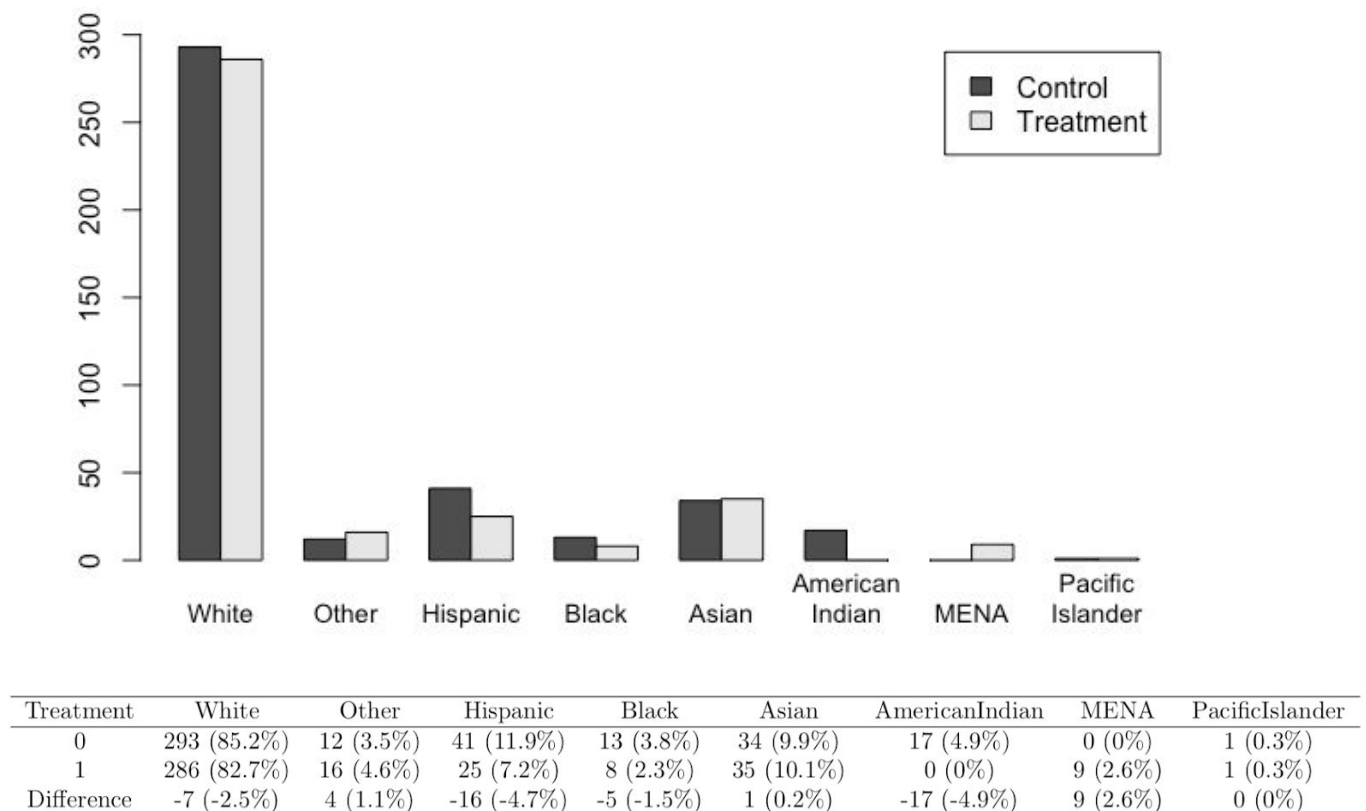


Figure 3: Distribution of Responses to Race and Ethnic Question

Counts and proportions of the major race and ethnic category responses are displayed in histogram and table format.

The proportions of race and ethnicities we had gathered through this experiment had some differences from the results of the 2010 census. We collected nearly 13% more White responses for the control question than in the 2010 census, which had been 72.4%. Hispanic responses were about 4% lower for the control question than in the 2010 census, although we had tried to oversample this community. We also had about double of the Asian responses, which was 4.8% in 2010, as well as a surprising percentage of American Indians for the control question as compared to 0.9% in 2010². We do not expect to achieve the same proportions as the 2010 census, as 9 years have passed and the makeup of the American ethnic and racial diversity is likely to have changed. However, the data that we had collected do make us question our recruitment method, especially as we attempted to oversample the Hispanic and MENA communities, which was unfortunately unsuccessful. Ultimately, it is difficult to say for sure, as we do not know the true race and ethnicity of our participants.

RESULTS

PRIMARY FINDINGS

For our primary research question, we regressed treatment on each of the binarized outcome variables. Our results for the proportions who selected White, Other, and Hispanic, and identified as MENA between the control and treatment version of the race and ethnic question are shown in Figure 4.

	<i>Dependent variable:</i>			
	White (1)	Other (2)	Hispanic (3)	MENA_writen (4)
Treatment	-0.025 (0.028)	0.011 (0.015)	-0.047* (0.022)	0.003 (0.012)
Constant	0.852*** (0.019)	0.035*** (0.010)	0.119*** (0.018)	0.023** (0.008)
Observations	690	690	690	690
R ²	0.001	0.001	0.006	0.0001
Adjusted R ²	-0.0003	-0.001	0.005	-0.001
Residual Std. Error (df = 688)	0.368	0.198	0.294	0.155
F Statistic (df = 1; 688)	0.807	0.570	4.408*	0.054
<i>Note:</i> *p<0.05; **p<0.01; ***p<0.001				

Figure 4: Regression Models for Primary Research Questions

The models for White and MENA resulted in coefficients in the expected direction. In contrast, the models for Other and Hispanic had coefficients in the direction opposite of the hypothesis. Out of the four outcome variables, only Hispanic had a significant treatment effect at the 0.05 significance level.

For the White specification, we found a treatment effect of -2.5% (2.8%). We had expected a negative coefficient, as people would be less likely to select White in the treatment if a more representative category is present, but this effect is not significant, likely due to not having enough statistical power. When accounting for our 4.4% attrition rate by substituting our outcome variable with the lower and upper extremes values, our true treatment effect lies between -5.6% to 0.9%, with the lower bound just barely becoming significant.

For the Other specification, we found a positive treatment effect of 1.1% (1.5%), which

was in the opposite direction of what we had expected. We believed that having more categories in which people can identify with, respondents would be less likely to select the Other category. When accounting for attrition, our treatment effect lies between -2.3% to 4.2%, with the upper bound becoming significant at the 0.05 significance level. However, our measurement for this outcome variable may have been compromised with the omission of the American Indian category, as 4 of the 16 people who selected Other in treatment had written that they were American Indian and may be the reason why the treatment effect was found to be positive. Nevertheless, we would have needed to recruit many more participants in order to achieve high statistical power.

For the Hispanic specification, we found a statistically significant 4.7% decrease in respondents selecting Hispanic in the treatment survey. When accounting for attrition, our treatment effect lies between -4.7% to -1.4%, in which the upper bound is no longer significant but still negative. We had hoped for no change in the proportion of Hispanic selections, such that the treatment survey does not deter people from selecting their most representative category. A possible explanation for this is that the control survey format is so common in society, people were not as used to this newer format and may have overlooked the category. Perhaps if this experiment was continued in the future, after people have become accustomed to the new format, this observed difference may diminish.

The MENA specification compared the proportion of people who wrote in a race in the MENA category in the control and the proportion of people who selected the MENA checkbox in the treatment. Although we expected an increase in people identifying as MENA, the model found no difference between the two as the coefficient is 0.3% (1.2%). When accounting for our attrition rate, both the lower and upper bounds of our treatment effect, -3.1% to 3.4%, become significant but are in opposite directions.

We also tested these outcome variables to see if there were any heterogeneous treatment effects for the three covariates we had collected. We did not find any heterogeneous treatment effect for gender, US region, and family's immigration history (Appendix G). Although we had 3 groups that had a significant finding at the 0.05 critical value, because we were conducting so many tests, we only considered positive results if the p-value was below 0.005, which was found by dividing 0.05 by 10, the number of tests run for each outcome variable.

Although we had regressed each outcome variable separately, in reality, they are likely intertwined and no one outcome variable may be sufficient to answer the question of whether the proposed race and ethnicity question has improved representation. As we look into the responses to the race question for those who selected Hispanic in the first ethnicity question in control, 68% selected White, 24% selected Other, 5% selected Asian, and 20% selected American Indian. This was surprising to us as we had thought that the majority of Hispanics would have selected Other, rather than White. Similarly, for those who had written in a MENA category in the control question, half had selected White and

the other half had selected Other Asian. We expected that most people who identify as MENA would have selected White in the control survey. Because the outcome space was not as distinct as we had anticipated, answering our research question may require more sophisticated models, such as multinomial logistic regression, in order to take all our outcome variables into account.

SECONDARY FINDINGS

We also had some secondary research questions that we wanted to explore using the data set we had collected. The regression models for Mixed Race, White Subgroups, and Time are shown in Figure 5.

	<i>Dependent variable:</i>		
	Mixed_Race (1)	White_Subgroups (2)	Time (3)
Treatment	0.034 (0.020)	0.282*** (0.031)	5.074 (9.656)
Constant	0.058*** (0.013)	0.662*** (0.028)	50.074*** (2.742)
Observations	690	579	690
R ²	0.004	0.125	0.0004
Adjusted R ²	0.003	0.123	-0.001
Residual Std. Error	0.264 (df = 688)	0.374 (df = 577)	126.941 (df = 688)
F Statistic	2.924 (df = 1; 688)	82.303*** (df = 1; 577)	0.276 (df = 1; 688)
<i>Note:</i>		*p<0.05; **p<0.01; ***p<0.001	

Figure 5: Regression Models for Secondary Research Questions

The models for Mixed Race and White Subgroups resulted in coefficients in the expected direction. White Subgroups had a significant p-value, even after accounting for the attrition rate. The model for Time had a coefficient in the direction opposite of the hypothesis but was not significant.

For the Mixed Race specification, we found that there were 3.4% (2.0%) more people who had selected two or more races in the treatment question than in the control question. We expected to find an increase, as the response to the Hispanic ethnicity question is not counted as an additional race in the control survey. However, the treatment effect found is not significant. If we account for the attrition rate, the true treatment effect would lie between 0% to 6.4%, in which the upper bound would be significant if all attriters had selected two or more races.

For the White Subgroups specification, we found a highly significant result, in which the treatment survey was able to cause 28.2% (3.1%) more White people to disclose their

specific ethnicity, on top of their major race category. This is what we had expected, since we thought that offering ethnicity checkboxes would be easier to select than actually taking the time to fill something in. When accounting for our attrition rate, our treatment effect would lie between 23% to 31%, in which both bounds remain significant at the 0.001 level. We can confidently say that including checkboxes has a causal effect in White respondents disclosing a more specific subgroup race.

Lastly, in the Time specification, our proxy for confusion, respondents took on average 5.1 seconds (9.7 seconds) longer to complete the treatment question, but this difference in completion time is not significant. Even when accounting for our attrition rate, by substituting the extreme values of 5.9 and 472.5 seconds, the minimum and maximum time from all observations (excluding the extreme outlier of 3199 seconds), the associated p-values remained insignificant at the 0.05 level. Although we did not find that the treatment survey question was faster or less confusing to complete, we take consolation that at least it did not take longer or was more confusing to complete than the control question.

LIMITATIONS AND LESSONS LEARNED

Looking back at our experience with this experiment, we had some discussions about what we would have done differently next time, and what we would have liked to do if we had unlimited time and resources.

The biggest limitation to our study was that we were unable to effectively oversample those who we set out to oversample, the Hispanic and MENA communities. If we had unlimited resources, one idea would be to first test if our Facebook Ads targeting strategy was working by using a different unaffiliated survey to see if we are able to reach our target audience. Another idea is that we would have preferred to use canvassers to help ensure that we are oversampling the right people. Ideally, we would like to be able to learn the actual race and ethnicity of our respondents, which could be done by conducting post-survey interviews.

If we were not so concerned with trying to achieve large sample sizes, we would have liked to use a factorial design to help identify mechanisms. One example would be randomizing the order of the categories to make sure people are not selecting White just because it is the first category shown.

We also learned that when using online surveys, we should try to optimize for mobile devices rather than for desktop. We learned that 97.9% of people targeted through Facebook Ads were using their mobile device.

Finally, if we were to use Facebook Ads in the future, we would prefer to disable comments and shares to minimize the potential for interference potentially created as a result of people commenting on the contents of the survey.

REFERENCES

1. Race and Ethnicity Defined
<https://www.cliffsnotes.com/study-guides/sociology/race-and-ethnicity/race-and-ethnicity-defined>
2. Overview of Race and Hispanic Origin: 2010
<https://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf>
3. Research to Improve Data on Race and Ethnicity
<https://www.census.gov/about/our-research/race-ethnicity.html>
4. 2020 Census To Keep Racial, Ethnic Categories Used In 2010
<https://www.npr.org/2018/01/26/580865378/census-request-suggests-no-race-ethnicity-data-changes-in-2020-experts-say>
5. 2015 National Content Test Race and Ethnicity Analysis Report
<https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2015nct-race-ethnicity-analysis.pdf>
6. Why We Ask Questions About... Race
<https://www.census.gov/acs/www/about/why-we-ask-each-question/race/>
7. Clinical Trials Have Far Too Little Racial and Ethnic Diversity
<https://www.scientificamerican.com/article/clinical-trials-have-far-too-little-racial-and-ethnic-diversity/>

APPENDIX

APPENDIX A: 2010 CENSUS ETHNICITY AND RACE QUESTION

→ **NOTE:** Please answer **BOTH** Question 6 about Hispanic origin and Question 7 about race. For this census, Hispanic origins are not races.

6. Are you of Hispanic, Latino, or Spanish origin?
Mark ☒ one or more boxes **AND** print origins.

☐ No, not of Hispanic, Latino, or Spanish origin

☐ Yes, Mexican, Mexican Am., Chicano

☐ Yes, Puerto Rican

☐ Yes, Cuban

☐ Yes, another Hispanic, Latino, or Spanish origin – *Print, for example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc.*

7. What is your race?
Mark ☒ one or more boxes **AND** print origins.

☐ White – *Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc.*

☐ Black or African Am. – *Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.*

☐ American Indian or Alaska Native – *Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.*

<input type="checkbox"/> Chinese	<input type="checkbox"/> Vietnamese	<input type="checkbox"/> Native Hawaiian
<input type="checkbox"/> Filipino	<input type="checkbox"/> Korean	<input type="checkbox"/> Samoan
<input type="checkbox"/> Asian Indian	<input type="checkbox"/> Japanese	<input type="checkbox"/> Chamorro
<input type="checkbox"/> Other Asian – <i>Print, for example, Pakistani, Cambodian, Hmong, etc.</i> <input type="text"/>	<input type="checkbox"/> Other Pacific Islander – <i>Print, for example, Tongan, Fijian, Marshallese, etc.</i> <input type="text"/>	

☐ Some other race – *Print race or origin.*

APPENDIX B: PROPOSED CENSUS ETHNICITY AND RACE QUESTION

8. What is Person 1's race or ethnicity?

Mark all boxes that apply **AND** print ethnicities in the spaces below.
Note, you may report more than one group.

☐ **WHITE** – Provide details below.

- | | | |
|----------------------------------|---------------------------------|----------------------------------|
| <input type="checkbox"/> German | <input type="checkbox"/> Irish | <input type="checkbox"/> English |
| <input type="checkbox"/> Italian | <input type="checkbox"/> Polish | <input type="checkbox"/> French |

Print, for example, Scottish, Norwegian, Dutch, etc.

☐ **HISPANIC, LATINO, OR SPANISH** – Provide details below.

- | | | |
|--|--|------------------------------------|
| <input type="checkbox"/> Mexican
or Mexican
American | <input type="checkbox"/> Puerto
Rican | <input type="checkbox"/> Cuban |
| <input type="checkbox"/> Salvadoran | <input type="checkbox"/> Dominican | <input type="checkbox"/> Colombian |

Print, for example, Guatemalan, Spaniard, Ecuadorian, etc.

☐ **BLACK OR AFRICAN AMERICAN** – Provide details below.

- | | | |
|--|------------------------------------|----------------------------------|
| <input type="checkbox"/> African
American | <input type="checkbox"/> Jamaican | <input type="checkbox"/> Haitian |
| <input type="checkbox"/> Nigerian | <input type="checkbox"/> Ethiopian | <input type="checkbox"/> Somali |

Print, for example, Ghanaian, South African, Barbadian, etc.

☐ **ASIAN** – Provide details below.

- | | | |
|-------------------------------------|-----------------------------------|---------------------------------------|
| <input type="checkbox"/> Chinese | <input type="checkbox"/> Filipino | <input type="checkbox"/> Asian Indian |
| <input type="checkbox"/> Vietnamese | <input type="checkbox"/> Korean | <input type="checkbox"/> Japanese |

Print, for example, Pakistani, Cambodian, Hmong, etc.

☐ **AMERICAN INDIAN OR ALASKA NATIVE** – Print, for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Tlingit, etc.

☐ **MIDDLE EASTERN OR NORTH AFRICAN** – Provide details below.

- | | | |
|-----------------------------------|-----------------------------------|-----------------------------------|
| <input type="checkbox"/> Lebanese | <input type="checkbox"/> Iranian | <input type="checkbox"/> Egyptian |
| <input type="checkbox"/> Syrian | <input type="checkbox"/> Moroccan | <input type="checkbox"/> Israeli |

Print, for example, Algerian, Iraqi, Kurdish, etc.

☐ **NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER** – Provide details below.

- | | | |
|---|---------------------------------|--------------------------------------|
| <input type="checkbox"/> Native
Hawaiian | <input type="checkbox"/> Samoan | <input type="checkbox"/> Chamorro |
| <input type="checkbox"/> Tongan | <input type="checkbox"/> Fijian | <input type="checkbox"/> Marshallese |

Print, for example, Palauan, Tahitian, Chuukese, etc.

☐ **SOME OTHER RACE OR ETHNICITY** – Print details.

APPENDIX C: QUALTRICS SURVEY

Page 1: Ethnic Food Questions



Welcome to our ethnic food survey!

This should take no longer for 3 minutes to complete. We appreciate your help!



How often do you consume ethnic food?

☐ Few times per week

☐ Few times per month

☐ Once a month

☐ Few times per year

Which ethnic food do you consume most often?

☐ Mediterranean

☐ Italian

☐ Chinese

☐ Mexican

☐ Japanese

☐ Other:

Where do you usually eat ethnic food?

☐ At home

☐ At a family member's house

☐ At a friend's house

☐ At a restaurant

Are there many ethnic food options near where you live?

☐ It's a diverse heaven

☐ There are some options

☐ I have to travel a bit for my preferred ethnic food options

On what occasions do you eat ethnic food?

☐ On special occasions (i.e. birthdays, celebrations)

☐ On special religious occasions

☐ No special occasions at all





What is your gender?

☐ Male

☐ Female

Which U.S. region are you from?



☐ West

☐ Midwest

☐ Northeast

☐ South

☐ I am not currently living in the U.S.

Which of the following best describes you?

☐ I immigrated to the U.S.

☐ My parents immigrated to the U.S.

☐ My grandparents immigrated to the U.S.

☐ My great-grandparents immigrated to the U.S.

☐ My great-grandparents were born in the U.S.





Are you of Hispanic, Latino, or Spanish origin?
Mark [X] one or more boxes **AND** print origins.

☐ **No**, not of Hispanic, Latino, or Spanish origin

☐ Yes, Mexican, Mexican Am., Chicano

☐ Yes, Puerto Rican

☐ Yes, Cuban

☐ Yes, another Hispanic, Latino, or Spanish origin – *Print, for example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc.*

What is your race?

Mark [X] one or more boxes **AND** print origins.

☐ White – Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc.

☐ Black or African Am. – Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.

☐ American Indian or Alaska Native – Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.

☐ Chinese

☐ Filipino

☐ Asian Indian

☐ Vietnamese

☐ Korean

☐ Japanese

☐ Other Asian – Print, for example, Pakistani, Cambodian, Hmong, etc.

☐ Native Hawaiian

☐ Samoan

☐ Chamorro

☐ Other Pacific Islander – Print, for example, Tongan, Fijian, Marshallese, etc.

☐ Some other race – Print race or origin.





What is your race or ethnicity?

Mark all boxes that apply **AND** enter ethnicities in the spaces below. Note, you may report more than one group.

☐ **WHITE** – Provide details below.

☐ German

☐ Irish

☐ English

☐ Italian

☐ Polish

☐ French

☐ Print, for example, Scottish, Norwegian, Dutch, etc.

☐ **HISPANIC, LATINO, OR SPANISH** – Provide details below.

☐ Mexican or Mexican American

☐ Puerto Rican

☐ Cuban

☐ Salvadoran

☐ Dominican

☐ Colombian

☐ Print, for example, Guatemalan, Spaniard, Ecuadorian, etc.

☐ **BLACK OR AFRICAN AMERICAN** – *Provide details below.*

☐ African American

☐ Jamaican

☐ Haitian

☐ Nigerian

☐ Ethiopian

☐ Somali

☐ *Print, for example, Ghanaian, South African, Barbadian, etc.*

☐ **ASIAN** – *Provide details below.*

☐ Chinese

☐ Filipino

☐ Asian Indian

☐ Vietnamese

☐ Korean

☐ Japanese

☐ *Print, for example, Pakistani, Cambodian, Hmong, etc.*

☐ **MIDDLE EASTERN OR NORTH AFRICAN** – *Provide details below.*

☐ Lebanese

☐ Iranian

☐ Egyptian

☐ Syrian

☐ Moroccan

☐ Israeli

☐ *Print, for example, Algerian, Iraqi, Kurdish, etc.*

☐ **NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER** – *Provide details below.*

☐ Native Hawaiian

☐ Samoan

☐ Chamorro

☐ Tongan

☐ Fijian

☐ Marshallese

☐ *Print, for example, Palauan, Tahitian, Chuukese, etc.*

☐ **SOME OTHER RACE OR ETHNICITY** – *Print details.*



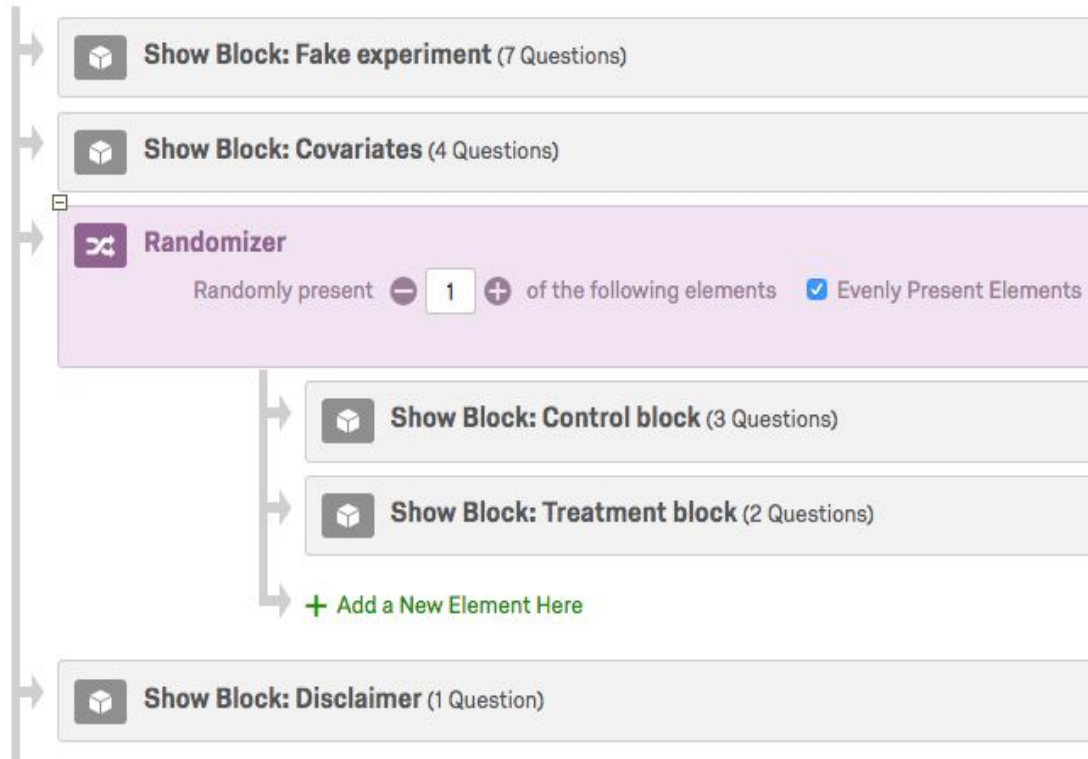
Page 4: Disclaimer Page



By submitting your response, you consent to the use of your anonymized and aggregated response to be used in a published study.



APPENDIX D: QUALTRICS SURVEY RANDOMIZATION



APPENDIX E: FACEBOOK AD

Facebook Ad Post

Page

Ad Center


Inbox 2

Notifications

Insights

Publishing Tools

More ▾



UC Berkeley I
School Grad
Students

Create Page @Username

Home

Posts

Reviews

▼ See more

Promote

Manage Promotions


Like

Follow

Share

...

Posts




UC Berkeley I School Grad Students

July 13 at 9:04 PM · 🌐

Hello! We are a team of UC Berkeley graduate students conducting a research project on the prominence of ethnic food in American culture. Please help us by taking a quick 3 minute survey. Link below

👉 https://berkeley.qualtrics.com/jfe/form/SV_26q0JlylrjNTN3v

We appreciate your help!



BERKELEY.QUALTRICS.COM

Online Survey Software | Qualtrics Survey Solutions

Qualtrics sophisticated online survey software solutions make...

26

People Reached and Engagement

38,802
People Reached

1,699
Engagements

Boost Again


Boosted on Jul 13, 2019
By Katayoun Borojerdi

Completed


People Reached **13.1K**


Link Clicks **512**


View Results


 93

23 Comments 28 Shares

 Like

 Comment

 Share



Targeting Criteria

Interests > Additional Interests

Hispanic culture

Middle East

Behaviors > Multicultural Affinity

Hispanic (US - Bilingual)

Hispanic (US - Spanish dominant)

Behaviors > Ramadan (Month)

Close friends of people celebrating Ramadan

Ramadan month (high content engagement)

Ramadan month (medium content engagement)

APPENDIX F: UC BERKELEY ETHNICITY CLUBS

Ethnicity Clubs Targeted

Student Organization	Followers/Members on Facebook	Ethnic Category
Arab Student Union	864 members	MENA
Afghan Student Association	680 followers	MENA
Armenian Students' Association	1086 members	MENA
Iranian Student Association in America	1831 followers	MENA
Yemeni Student Association of Berkeley	236 followers	MENA
Middle Eastern North African Recruitment and Retention Center	505 members	MENA
Chicano(a)s/Latino(a)s in Health Education (CHE)	1257 followers	Hispanic
Hermanos Unidos	633 followers	Hispanic
Hermanas Unidas	674 followers	Hispanic
Latin American Leadership Society	621 followers	Hispanic
Mexican Association of Students at Berkeley (MEXASB)	523 followers	Hispanic

Example of Email Sent

Help fellow UC Berkeley students with our Ethnicity Food Survey!

Katie Mo <kmo@berkeley.edu>
to ucberkeleyasu, Katayoun ▾

Sun, Jul 21, 1:06 PM   

Hi UC Berkeley Arab Student Union Organizers,

We are a group of graduate students at the UC Berkeley School of Information and we are currently conducting a research study on the prominence of ethnic food in American culture. We are looking for people to fill out this quick 3 minute survey on ethnic foods: https://berkeley.qualtrics.com/jfe/form/SV_9EQeq0ykHxI2vVX. Would you be willing to help us forward this survey to the members in your cultural club? We would really like to ensure all ethnic groups are represented in our study.

Your help is much appreciated!

Best regards,
Kat Borojerdi and Katie Mo

APPENDIX G: TESTS FOR HETEROGENEOUS TREATMENT EFFECTS

outcome	covariate	estimate	se	p-value	significant	significant_bonferroni
White	Female	0.06	0.08	0.43	0	0
White	Region_West	-0.03	0.07	0.62	0	0
White	Region_Midwest	-0.01	0.06	0.86	0	0
White	Region_Northeast	0.02	0.07	0.73	0	0
White	Region_South	0.04	0.06	0.52	0	0
White	Immigrate_Self	-0.02	0.13	0.87	0	0
White	Immigrate_Parents	0.16	0.11	0.13	0	0
White	Immigrate_Grandparents	-0.02	0.06	0.75	0	0
White	Immigrate_GreatGrandparents	-0.03	0.05	0.54	0	0
White	Immigrate_None	0.01	0.05	0.84	0	0
Other	Female	-0.00	0.04	0.91	0	0
Other	Region_West	-0.05	0.04	0.18	0	0
Other	Region_Midwest	-0.01	0.03	0.81	0	0
Other	Region_Northeast	0.07	0.03	0.04	1	0
Other	Region_South	0.00	0.03	0.96	0	0
Other	Immigrate_Self	-0.07	0.08	0.34	0	0
Other	Immigrate_Parents	-0.04	0.05	0.40	0	0
Other	Immigrate_Grandparents	-0.01	0.04	0.83	0	0
Other	Immigrate_GreatGrandparents	0.06	0.04	0.14	0	0
Other	Immigrate_None	0.01	0.03	0.82	0	0
Hispanic	Female	-0.02	0.06	0.69	0	0
Hispanic	Region_West	-0.11	0.06	0.07	0	0
Hispanic	Region_Midwest	0.09	0.04	0.04	1	0
Hispanic	Region_Northeast	-0.04	0.04	0.34	0	0
Hispanic	Region_South	0.06	0.05	0.28	0	0
Hispanic	Immigrate_Self	-0.23	0.11	0.04	1	0
Hispanic	Immigrate_Parents	0.02	0.08	0.82	0	0
Hispanic	Immigrate_Grandparents	0.01	0.06	0.84	0	0
Hispanic	Immigrate_GreatGrandparents	0.02	0.05	0.75	0	0
Hispanic	Immigrate_None	0.04	0.04	0.34	0	0
MENA_writein	Female	0.04	0.03	0.21	0	0
MENA_writein	Region_West	0.06	0.03	0.07	0	0
MENA_writein	Region_Midwest	-0.03	0.02	0.14	0	0
MENA_writein	Region_Northeast	-0.02	0.03	0.42	0	0
MENA_writein	Region_South	-0.02	0.02	0.30	0	0
MENA_writein	Immigrate_Self	0.05	0.06	0.43	0	0
MENA_writein	Immigrate_Parents	-0.01	0.05	0.89	0	0
MENA_writein	Immigrate_Grandparents	-0.02	0.03	0.50	0	0
MENA_writein	Immigrate_GreatGrandparents	-0.02	0.02	0.53	0	0
MENA_writein	Immigrate_None	0.01	0.02	0.70	0	0