Collaborating Center for Questionnaire Design and Evaluation Research

March 7, 2024

Cognitive Interview Evaluation of a Combined Race and Ethnicity Question including a Middle Eastern/North African Category

By: Stephanie Willson and Alisú Schoua-Glusberg

INTRODUCTION

The National Center for Health Statistics (NCHS), in collaboration with Research Support Services (RSS), conducted a cognitive interview evaluation of a proposed combined race and ethnicity question as presented in the Office of Management and Budget (OMB) Federal Register Notice (FRN) as part of the effort to revise OMB Statistical Policy Directive 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity (SPD 15). This report documents the performance of that question. It discusses how the framework of the question influences the question-response process in terms of how respondents go about choosing an answer and any difficulties they encounter in doing so. The study received both OMB and National Center for Health Statistics/Centers for Disease Control and Prevention Human Subjects approval.

Suggested citation:

Willson, S., Schoua-Glusberg, A. Cognitive Interview Evaluation of a combined Race and Ethnicity Question Including Middle Eastern/North African. National Center for Health Statistics -CCQDER. Q-Bank. 2024. Available from: https://wwwn.cdc.gov/gba

nk/report.aspx?1249

Combining race and ethnicity into one question is a departure from the current OMB standards for reporting race and ethnicity.² This study examined how respondents formulate answers to a question that combines race and ethnicity, and whether the categories offered are sufficient for people with varied backgrounds. Special attention was paid to Middle Eastern or North African (MENA) because it represents a new minimum reporting race category for federal surveys.

The methodology of the study is presented next, followed by study findings.

METHODOLOGY

The combined race and ethnicity question was administered in two languages (english and spanish) and in two modes (self-administered and interviewer-administered).³ half of the respondents received the interviewer-administered version and half received the self-administered version of the question.

³ The self-administered version can be seen in English and Spanish in Attachments 1 and 2.



¹ Federal Register / Vol. 88, No. 18 / Friday, January 27, 2023 / Notices

² Office of Management & Budget, Executive Office of the President, Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity, 62 Fed. Reg. 58,782 (Oct. 30, 1997), https://www.govinfo.gov/content/pkg/FR1997-10-30/pdf/97-28653.pdf

I. Sample

The race and ethnicity question was evaluated in the context of other questions, including gender, COVID-19, and cancer screening. All potential study participants were informed of the topics under investigation. Sample selection for the project was purposive. The aim was to choose respondents who met criteria relevant to the questions under investigation. Because this project included multiple topics, recruitment was based not only on race and ethnicity, but also gender, experiences with COVID-19, and age (to align with cancer screening recommendations). Recruitment was carried out through a combination of flyers, special interest groups, and respondent referrals. A demographic breakdown of respondents appears in Table 1. A total of 150 interviews were completed, 100 in English and 50 in Spanish. (Twenty-five of the Spanish interviews included gender minorities due to a separate evaluation of a gender question.)

Table 1: Respondent Demographics by Language

	English (n=100)	Spanish (n=50)	Total (n=150)
Race/Ethnicity*			
White	37	21	58
Black	12	2	14
Asian	20	0	20
American Indian or Alaska Native	2	1	3
Middle Eastern or North African	29	0	29
Hispanic	9	50	59
Gender			
Female	55	14	69
Male	43	11	54
Transgender, non-binary, or another gender	2	25	27
Age			
18-29	28	8	36
30-49	51	30	81
50-64	15	12	27
65 and older	6	0	6

^{*}Numbers do not add to denominators because respondents could select more than one category.

II. Data collection

Staff at the NCHS Collaborating Center for Questionnaire Design and Evaluation Research (CCQDER) and RSS conducted all 150 interviews. Interviews were conducted virtually (using the Zoom web conferencing platform) and lasted no longer than one hour. Upon completion of the interview, respondents received a \$50 remuneration.

Interviewers first administered the survey questions as intended under actual field conditions and obtained respondents' answers to all the questions. The self-administered version was completed on-screen by respondents, while the interviewer-administered version was read aloud to respondents. The combined race and ethnicity question appeared as the first question in the test instrument, the gender identity question was second, followed by questions on COVID-19 and, finally, cancer screening. The second part of the interview consisted of retrospective probing designed to capture contextual insight into the ways in which respondents interpreted the question, considered and weighed out relevant aspects of their lives, and formulated a response based on that consideration.

Upon completion of the interviews, all written summaries and partial transcripts were uploaded in English into Q-Notes⁴, a software application for data storage and analysis of cognitive interviews. Nine CCQDER interviewers conducted interviews in English, and three RSS interviewers conducted interviews in Spanish and English. All interviewers were skilled in qualitative interviewing techniques. Additionally, the use of Q-Notes allowed the Principal Investigator to monitor data quality as interviews were being completed.

III. Data analysis

Data analysis was conducted according to the grounded theory approach, which generates explanations of interpretive patterns and response error that are directly linked to the empirical data. Often referred to as the constant comparative method of analysis, the link between conclusions and data is achieved by comparing substantive findings against the original data across the entire research endeavor, from data collection to post-collection analysis (Lincoln and Guba 1985; Strauss and Corbin 1990; Suter 2012). Each phase represents a different level of analysis.

Multiple levels of analysis, per Miller et al. (2014), were performed for this study. First, analysts synthesized interview data into summaries, detailing how each respondent interpreted the question and formulated their answers. Next, analysts compared summaries across respondents, identifying common themes. Once themes were identified, analysts compared themes across subgroups, revealing ways in which different groups of respondents processed the question differently depending on their differing experiences and socio-cultural backgrounds. Finally, analysts drew conclusions, determining and explaining how the question performed as it functioned within the context of respondents' various experiences and socio-cultural locations. In each analytic step, data were reduced into a theoretical summary detailing the question's performance. As such, these different analytic steps achieve both data reduction and a movement toward larger conceptual themes. These findings are discussed in the next section

FINDINGS

I. The Question-Response Process for Reporting Race and Ethnicity

The question wording evaluated in this study reflects that which was posted in the OMB 2023 FRN and reads: What is your race or ethnicity? Select all that apply AND enter additional details in the spaces below. Note, you may report more than one group. ⁵ The full question had multiple categories as follows (refer to Attachments 1 and 2 for the full question and list of response options for each language):

⁴ Q-Bank - Q-Notes - Login (cdc.gov)

⁵ In Spanish, the question read: ¿Cuál es su raza o grupo étnico? Seleccione todas las opciones que correspondan Y ponga más detalles en los espacios que aparecen abajo. Se puede seleccionar más de un grupo.

What is your race or eth Select all that apply ANE Note, you may report me	enter additional details in the s	paces below.
WHITE - Provide de German Italian Enter, for example,	tails below. Irish Polish Scottish, Norwegian, Dutch, etc.	English French
Mexican or M American Salvadoran	NO - Provide details below. exican Puerto Ricar Dominican Guatemalan, Spaniard, Ecuador	Colombian
African American Nigerian	N AMERICAN - Provide details be can Jamaican Ethiopian Ghanaian, South African, Barba	Haitian Somali
ASIAN - Provide del Chinese Vietnamese Enter, for example,	tails below. Filipino Korean Pakistani, Cambodian, Hmong,	Asian Indian Japanese
	OR ALASKA NATIVE - Enter, for ve Village of Barrow Inupiat Trib	example, Navajo Nation, Blackfeet Tribe, al Government, Tlingit, etc.
Lebanese Syrian	OR NORTH AFRICAN - Provide de l'Iranian Moroccan Algerian, Iraqi, Kurdish, etc.	etails below. Egyptian Israeli
Native Hawaiia Tongan	OR PACIFIC ISLANDER - Provide In Samoan Fijian	details below. Chamorro Marshallese

Respondents went about answering the question in several ways. One group of respondents provided answers without deliberation, providing pre-conceived answers. A second group of respondents spent some time considering their answer, taking into account what they believed to be the purpose of the question.

Pre-Conceived, Reflexive Answers: For some respondents, reporting race and ethnicity was a straightforward task into which they put little thought. These respondents did not deliberate or go through a complex question-response process. Instead, their answers were more reflexive. In fact, when the process of answering was so automatic, it was difficult for respondents to explain the rationale behind their answers. This was often true for respondents who chose 'White.' For example, when one respondent was asked to explain his answer, his only reply was, "I'm White...Just White." Another respondent also struggled to articulate why he chose 'White,' but tried to offer a post hoc explanation. It seemed clear he had not given it much thought. He said, "Um. Well, I...[Thinks]...My parents are both White. My brothers are both White. My children are both White. Um. [Shakes head] I was brought up knowing I was White."

It was not only respondents who opted for the 'White' category. Other groups also had reflexive responses and struggled with providing a rationale. One respondent said, "So, it's just like – 'African American.' So, out of habit I just choose 'African American.'" It was similar for some respondents who marked 'Hispanic.' One Spanish-speaking respondent said, "I always answer that. And it's always been the same." Another Spanish-speaking respondent chose 'Hispanic' and said, "Basically that's how I've done it all my life."

Considered Answers: While some respondents offered reflexive responses to the combined race question, other respondents demonstrated more thought and nuance when providing an answer. Four patterns emerged that determined how respondents conceptualize their race and ethnicity. These patterns (or dimensions of race and ethnicity) are similar to those found by Miller and Willson (2002) and Willson and Dunston (2017) and are identified as cultural, social, administrative, and ancestral. Although the patterns are conceptually distinct, in practice they are not mutually exclusive and have elements that overlap.

⁶ Notes for Spanish interviews were translated into English; therefore, all direct quotes are presented in English.

The cultural dimension is characterized by a feeling of connectedness to a group which arises from shared ideas, language, or cultural practices. The social dimension of race and ethnicity refers to the way others in society view and define a person. In the US this is often based on physical features, most notably skin tone, but can also include cultural aspects (such as language, clothing, food, or religion). The administrative dimension is how a person answers in an official capacity on various forms. For example, reporting race and ethnicity for standardized questions can depend on the purpose of the form or survey. Finally, the ancestral pattern is based on genealogy, often colloquially referred to as a person's "family tree." The following discussion offers examples of each pattern.

<u>Cultural</u>: Many respondents thought of their race and ethnicity as an identity tied to cultural group affiliations and answered the question on this basis. This was true for many MENA respondents. For example, one respondent said, "I grew up in a community of Middle Eastern. Like, everyone is Middle Eastern in this community." Another respondent said, "I think there's also a lot of cultural customs and traditions that I'd say being Middle Eastern we pride ourselves on."

Hispanic respondents also thought of cultural dimensions of race and ethnicity when providing an answer. For example, one Spanish-speaking respondent described feeling part of a cultural subgroup and said, "We come from Mexico and find here [in the US] there are a lot of [Hispanic] people...and we all join together. And we all speak Spanish." In fact, the connection of language was cited by many Spanish-speaking Hispanic respondents as a central cultural factor in defining Hispanic identity in the US. The following are examples:

"I think just 'Hispanic.' First because there's no other culture or ethnic group I identify with. I'm from Latin America. I am Hispanic, Latino. I speak Spanish. My native language is Spanish."

"I am Latino because when the Spaniards colonized my country... In my country Spanish and Quechua are spoken, when the Spaniards conquered Ecuador, they took their language to our country, which is the Spanish language, that is why I am Latino, because it comes from Latin."

Respondents who thought of their race or ethnicity as defined by shared group experiences sometimes omitted racial and ethnic groups from their answer when they felt no cultural connection to those groups – even if those groups were technically part of their ancestry and lineage. For example, one respondent who told the interviewer that his mother was Asian did not include Asian in his answer. His explanation was because he was never exposed to his mother or to Asian culture. He said, "I'm not Asian, I don't speak Chinese, I don't speak Korean." Another respondent discussed her American Indian lineage (her mother was Black and American Indian) but did not include it in her answer. She explained "I don't really connect with it. The only time I talk about it is if people say something about my mom's hair or something like that. I might joke about it – like I have half my mom's hair and half my dad's hair." Another respondent chose only 'Black' but told the interviewer she also had White in her background. When asked why she reported only Black, she said:

"Yes. Because that is the way that I was raised. I was raised only by my mother. My mother, her grandfather was White, but her mother was half White and half Black and her father was Black, and I was raised by them, and our community was Black. You know what I mean, like the family was all Black the people I lived around were Black...I mean my neighborhood was actually pretty diverse, but the people I actually had dealings with outside of school were Black people. So, I just identify with being Black. I eat foods that Black people eat, I speak like a Black person, you know what I mean?"

<u>Social</u>: For some respondents, the dimension informing their answer to the combined question arose from perceptions that others have regarding their race. This was often driven by physical appearance. For example, one respondent talked about being defined a certain way based on physical attributes. He said, "I don't want to speak for all Black people, but some of us do know that we do have Caucasian American in us. But we don't identify with that, I guess. Because, if I come on here and say, 'Caucasian American' you might come on here like, 'Oh, yeah, sure you are [sarcastic tone]!'" Another respondent also chose 'Black' and linked her rationale to her treatment by society in general. She explained:

"I'm reminded that I'm Black daily. And because, the color of my skin doesn't allow me to even, even if I was Puerto Rican if I were to walk outside, I would definitely have to explain to everyone I walked past that I'm Puerto Rican instead of Black. So even if I did have Puerto Rican, even if I was Haitian and Puerto Rican, I would still identify as Black because it would take too much time and be too stressful to explain to somebody that I'm Puerto Rican."

An English-speaking respondent who identified as Afro-Latina echoed this perspective. She chose to report only 'Hispanic' even though she discussed her complex genealogy with the interviewer. She said, "I am Dominican because I was born in the Dominican Republic. However, from my mother and my father's side I have German, I have Spaniard, Italian and African. So, I have all those things. That I know of. God knows what else is my DNA breakdown." Despite this reality, her decision to report only Hispanic related to her perceptions of how she is defined in this country. She said:

"Because in this country it doesn't feel right for me to choose the other two [White and Black]. Because we're lumped into that sum of Hispanic – into that group. I live in a neighborhood where there's all cultures and I'm still considered Hispanic. White people don't consider me White. Black people don't consider me Black. Even Black people from the Islands don't consider me – they still see me as Hispanic."

Respondents with MENA backgrounds also considered the perceptions that others have of their physical appearance, especially if respondents were recent immigrants trying to decipher American notions of race. For example, one respondent described the logic of his response as one that, in the absence of a more applicable category, considers the American emphasis on skin color. He said, "I identify as 'White' in surveys because a) normally there is no option for Middle Eastern North African and b) I look White." Another respondent personally identified as White/European but expressed the idea that he is not seen that way in the US. As a result, he answered based on how others see him. He said, "Because I was born in Istanbul. Istanbul is in Europe border region. But for myself I am saying Middle Eastern [on the survey] because my looks are not European, that's why I chose Middle Eastern. Normally I would be from Europe, but I chose Middle Eastern because I am not white, I am brown. That is why."

Administrative: Respondents who see themselves as having multiple, valid ways of reporting their race and ethnicity often consider the purpose of the form or survey before they answer. For example, one respondent who answered only 'Black' talked to the interviewer about her White and Native American background. She said, "I went on a search to get my DNA tested or whatever because I know that we certainly have White in our family. We certainly have Native American in our family. We certainly have African in our family." But because the interview also included questions on COVID-19 and cancer screenings, she chose only 'Black,' "just to keep it simple, because I thought it was going to be more focused on health things, so I thought that for this purpose, I would just keep it simple with the 'African American' [category]." This respondent could authentically answer a question on race and ethnicity multiple ways, but chose her answer based on the dimension she saw as most salient in the context of health topics.

Additionally, experience in the US has taught some respondents that there are prescriptive ways for answering questions on race. This is particularly true when answering questions on administrative or official forms. This pattern was observed among respondents with a MENA background and was often due to experiences of being told to answer 'White' on forms when a MENA option is not proffered. For example, one respondent discussed how questionnaires in Iran do not ask about ethnicity, so she was unsure how to answer this question in the US until friends told her she should "always choose White." As a result, even when MENA is an option, some MENA respondents continue to include 'White' in an attempt to conform to US norms. As one respondent said, "When I was young, I just put down 'White.' I did not want to make a big deal of my race or ethnicity. [But] personally, I don't consider myself White."

Some Spanish-speaking Hispanic respondents also demonstrated this pattern. Similar to their MENA counterparts, they have often been instructed to mark either 'White' or 'Black' (in addition to 'Hispanic') on forms or questionnaires. Some respondents have adopted that practice even if they would not, on their own accord, make that choice. For example, one respondent chose both 'Hispanic' and 'White,' and explained that she has become accustomed to filling out 'White' on most surveys (even though she sees herself as only Hispanic).

Ancestral: A person's "family tree" is often seen as the most salient aspect for reporting race and ethnicity. This dimension served as the primary basis for answers among many respondents who chose MENA as their race and ethnicity. The rationale for their answers included genealogy, geography, or more typically a combination of the two. For example, one respondent chose both 'White' and 'MENA' because each represents a different parent. They said, "I am half French so that would be 'White,' but the other half is North African." Geography was often interwoven into notions of ancestry, particularly for respondents who were born abroad. The following are examples of respondents who perceive their race in terms of ancestral background and geographic history:

"Middle Eastern because I was born in Egypt. If there is an option for MENA, I go straight there because I was born there."

"I'm from the geographical location of the Eastern part of the world, but kind of being at the kind of crossroads of the world. That is why it is called the Middle East."

"Middle Eastern for sure, I'm Egyptian...I think of your roots, of where your parents are from. What country they came from."

Spanish-speaking Hispanic respondents demonstrated a similar pattern. Many in this group answered the question based on their ancestry, which was tied to or defined by where they were born, as the following examples illustrate:

"Because I belong to Latin America. I was born there."

"I was born in Puerto Rico. Being born in Puerto Rico makes me Latino."

"...I have Mexican roots, that is the reason why I answered that way. My ancestors are Mexican."

"Because I identify as Latino, so the word Hispanic/Latino takes me back to my geographic area of birth."

The ancestral pattern of response was apparent in other groups as well. For example, one Asian respondent said, "I heard that [question] to be [asking] where you're born, basically. Or your ancestors. And that's from Pakistan. I'm born there. I had my childhood there."

For another group of respondents, the decision to base their answer on ancestry was not necessarily tied to personally salient dimensions of their race and ethnicity or even to the context in which the question was asked. It was instead motivated by the structure of the question itself.

Some respondents noticed and commented on the unique format of the combined question. For example, one respondent said:

"I noticed with the race and ethnicity demographic that it's more specific. Because obviously you want to know which particular group or race that that person is...um...belongs to. Like for 'White,' because there are different subcultures or subraces of White people and same with 'Hispanic' and then 'Black or African American,' you know American Black or Jamaican. So, this is cool. It's good, you know?"

The previous respondent's narrative points not only to the fact that the new question format was noticed, but that it was asking for something different than most other questions on race and ethnicity. This observation caused some respondents to openly ponder (and sometimes ask about) question intent. It also caused some difficulty for

those who do not normally consider ancestry a salient part of their race and ethnicity. The next respondent reveals the complex response process (and difficulty) some people went through in determining how to best answer the question. She saw herself simply as White ("Because I look like I'm albino [laughs]. I have fair skin. I guess people acknowledge me to be White, Caucasian.") but she initially understood the question as asking about her ancestry. As a result, she began to answer in a way that included every known element of her genetic lineage. She said, "I just did my 23AndMe recently...German, Irish, French, English...that's pretty much what I am, and I'm like, um, some weird...0.6% of sub-Saharan African or something..." However, when she got to the 'Black or African American' category she began to rethink how she would answer the question because her DNA test results ran contrary to the way she normally defines herself. She said, "I don't think a .6% on my genealogy really counts as...[trailed off]." Similarly, when she saw 'Middle Eastern or North African,' she said, "It [the DNA test] says I'm like .6% Moroccan." Unsure whether to include these races, she asked the interviewer, "Should I add that?" The interviewer said it was her decision to make. Deciding it was not "enough" heritage to include, she justified omitting those categories by falling back on what she deemed a more authentic dimension of her race (a social dimension) regardless of what she thought the question may actually be asking. She said, "Clearly I don't look it [Black or MENA]." [Shows her arms to the camera, laughing.]

Not every respondent who pondered question intent arrived at the same understanding. Like the previous respondent, another respondent also saw herself only as White and did not have any personal sense of ethnic identity. But when she saw the subcategories, she wondered if she should answer based on her ancestry instead of her first inclination (which would entail leaving the ethnicity boxes blank). She said, "I have French and Irish in me. Would I put that?" The interviewer explained it was her decision. Unlike the previous respondent, she ultimately decided that, because of the enumerated subcategories, the question was asking specifically for genealogy. She, therefore, answered based on what she knew about hers. She said, "I think I'm ¾ French, ¼ Irish. I believe." Another respondent who also personally identified only as 'White' found himself choosing 'MENA' for the first time ever. He did not normally think of himself as Middle Eastern or North African, yet the structure of the question caused him to answer according to what he knew about his genealogic background. He said, "Put, actually, 'Italian'…and then go down to…I'm actually Armenian. So that's by Turkey, so I guess that would be the Middle East."

This pattern was also observed for categories other than 'White.' One example includes a respondent who marked 'White' and 'MENA' because her mother is White and her father is Palestinian. She interpreted and answered the question from the perspective of genealogy, but when asked how she normally describes herself, she said, "Palestinian American." She specified that her personal identity did not include 'White' because, "I'm never seen as a White person," but she included that category here because it is technically part of her genealogy. A Spanish-speaking respondent who marked 'Hispanic' normally does not think about reporting his subgroup affiliation. But he thought it was appropriate to do so here, given the structure of the question. He said, "So, I, myself, don't get offended if I'm just [able to] put the 'Hispanic or Latino' category. I don't have to say that I'm Puerto Rican. But, if you're going to ask that question, then I'm certainly going to note that."

In sum, the format and level of detail in the combined question prompted some respondents to see the intent of the question as asking specifically about their ancestry. Some respondents would answer questions on race or ethnicity based on ancestry to begin with, especially those for whom genealogy is a salient part of how they think of themselves. However, other respondents who typically think in more cultural or social ways about their race and ethnicity wound up providing an answer to this question based on their (known) ancestry – something they would not normally do. This was especially (but not exclusively) true of White respondents.

It is important to note that in no cases were respondents offering "incorrect" answers about their race or ethnicity because they "misunderstood" the question. It is more accurate to say that the complexity of the race and ethnicity construct, along with the detailed subcategories of the question, allows for a range of authentic individual representations. In this sense, the combined question performed well – everyone was able to choose categories that were acceptable to them and fit their realities, albeit in different ways.

The Impact of the Question Stem and Instructions on Question Response

One concern regarding the combined question was whether respondents would be confused by the merging of two concepts, race and ethnicity, into a single question. Therefore, attention was paid to the extent to which respondents understand both the question wording and instructions on how to answer. Recall that the question read: What is your race or ethnicity? Select all that apply AND enter additional details in the spaces below. Note, you may report more than one group.

As mentioned above, the detailed categories and the combined race and ethnicity feature of the question caused some respondents to reflect on the intent of the question. The patterns (or dimensions of race and ethnicity) that they chose to frame their responses were informed by their personal life experiences and/or by the structure of the question itself, specifically, the level of detail in the subcategories. For others, answering was a non-reflexive activity for which they had ready-made answers.

For these reasons, the phrase "race or ethnicity," along with the instructions, was not confusing and had little bearing on how respondents answered the question. In other words, how respondents may (or may not) define "race" and "ethnicity" in the abstract played little, if any, role in how they chose to answer. By far it was the categories themselves, and respondents' self-perceptions vis-à-vis the categories, that shaped how respondents answered.

Another reason that the wording of the question stem and instructions did not have much impact on question response is simply that many respondents did not focus on them. For example, one respondent commented that normally he skims through the directions and doesn't read such things verbatim ("Yeah I'm a skimmer."). This was typical; most respondents paid little attention to the question stem or the instructions. Additionally, respondents seemed to intuit that the question was 'mark all that apply,' though some did notice (or hear) this aspect of the instruction when answering.

II. Notable Group Patterns

Middle Eastern or North African (MENA) respondents: The MENA category represents a new minimum reporting race category as part of the combined question proposed for the SPD 15 revision. As a result, 29 interviews were conducted with respondents who identified as either Middle Eastern or North African. Specific patterns of response have been noted above; however, overall, respondents who identified as having a MENA background did choose this category and were able to find their appropriate subcategory. One respondent, however, did not initially check the MENA box because he was unaccustomed to seeing such an option for himself. He said, "Usually I mark 'Asian' because I was born in Iran, which is part of Asia. And I know my father and mother were both born in Iran, and my grandfather and grandma also. I was born in Iran." The interviewer pointed out the MENA race option and he immediately said, "Oh yeah, I did not see this! Yeah, I did not see, sorry. Because in all official forms we do not have 'Middle East' part, we just have 'Asia.' So, because of that, after seeing 'Asia' I just [marked the box]." The tendency of MENA respondents to miss the MENA category figured more prominently in prior work (See Willson and Dunston, 2017), but this example suggests the pattern may still occur, particularly in a self-administered format.

For those who did notice the category, the term 'Middle Eastern or North African' resonated well. For example, when asked if this was a term she would use herself, one respondent replied, "Yeah, it is, definitely." Another respondent was asked if she thought the MENA category felt more authentic than having to choose 'White,' which she typically does in the absence of MENA. She said, "I wouldn't describe it as authentic I would describe it as accurate. I feel like geographically, scientifically, to my mind this is very accurate." Other respondents also described a preference for having a MENA category as an option instead of having to choose, for example, 'White:'

"When I was applying to [college], all the forms I just checked 'White.' [But]...White doesn't quite – like I don't feel I have a lot in common with Irish or Italian [people]."

"I just saw 'Iranian' and, frankly, it's so unusual to see Iranian. Because I don't see my category whenever I fill something out. So, I'll either do Caucasian or Asian. Which neither of them are, really – I mean the Caucuses are right there, so technically we're Caucasian. But not the way you guys think about it."

"When there is no 'Middle Eastern' category [when recently applying for a job] there was 'African [American].' So I just chose that since Egypt's in Africa."

The above examples illustrate how the MENA option allowed respondents to opt for a category that more closely aligned with how they see themselves rather than having to choose what they saw as an ill-fitting category (such as 'White').

Black or African American Respondents: Respondents who are descendants of enslaved people (versus those who immigrated to the US voluntarily) often described themselves as simply "Black" (or "Black American") and did not view their lineage through the lens of an immigration narrative. However, the subcategories seem to imply such an experience. Although respondents who identified as Black were able to choose an answer as intended ("Black/African American" as the main category and "African American" as the subcategory), some respondents were somewhat confused by the subcategory options or admitted that it would be impossible to provide an answer because they had no way of knowing from what country their ancestors were taken. For example, one respondent demonstrated some confusion and said, "I would say 'African American' [thinks]...I'm sorry – that threw me off with all those [subcategories]. I would say 'African American." When asked about the source of his confusion he said:

"It's weird because I'm, like, I guess a natural born Black American. So, I don't – my lineage doesn't – unless you go WAY back – trace back to Africa. So, it's interesting that African American, the subset started listing African countries. Which technically...it's weird. Even though I identify as African American, I don't really see myself as African. I identify more as 'Black American.'"

The act of asking for specific African lineage of Black respondents can be awkward, and even potentially insensitive. The next two respondents illustrate why:

"I'm African American, I was born here, and according to history that was told before, we've basically not, you know, we're from Africa, but our predecessors...are from a particular place in Africa. But it's something we don't actually know, if we are from a particular part of Africa. Just know we're from Africa."

"I don't even say that I'm African American, I say that I'm Black, you know. But mostly I say that I'm an American because my family has been American for a very long time and even though they weren't considered American back then, I don't identify with anything beyond America. I just know that, because my family comes from slavery, there was a lot of mixing."

A Black respondent who was born in another country articulated the issue from the opposite perspective. He said, "Again, because I'm a naturalized American citizen, my exposure to different cultures and my origins are different than other Black people in America. So, I still identify as a Black person, but I know my roots are in the Caribbean." When asked why he chose that category if he did not think it fit, he explained, "Because that's the only one available that kind of captures part of my identity."

It seems that while the category 'Black or African American' is meant to capture both people who are descendants of enslaved people *and* people who more recently immigrated to the US by their own volition, it is not always seen as entirely befitting to merge both groups into this category.

<u>Hispanic Respondents</u>: Because 'Hispanic' is combined with the race groups as a minimum race reporting category (rather than a separate ethnicity) in the proposed question, 59 Hispanic respondents were interviewed for this study. Most (50) were interviewed in Spanish; nine were interviewed in English. Those who were interviewed in Spanish were born outside the US and Spanish was either their only or dominant language.

Patterns that were observed among non-Hispanic respondents were often mirrored in the Hispanic group. That is, Hispanic respondents also answered based on cultural, social, administrative, or ancestral understandings of race and ethnicity as examples in the previous section illustrate.

Overall, the combined question was not difficult or confusing for most Hispanic respondents. They were able to find a suitable response category. Sometimes this meant choosing more than one category (such as 'White' and 'Hispanic') but more often they chose just one category, 'Hispanic.' Additionally, this was true even when the term 'Hispanic' was seen as an American construct. These patterns are discussed next.

Choosing one versus more than one main category: The detailed subcategories under the 'White' category caused some Spanish-speaking respondents to refrain from choosing 'White' along with 'Hispanic.' This seemed, in part, due to the subcategories framing 'White' as a European category. For example, one Spanish-speaking respondent considered selecting 'White' because of her skin color but changed her mind when she saw the details listed. She said, "Here, under 'White' this is all Europe." Similarly, another Spanish-speaking respondent saw the 'White' subcategories and said, "I do not identify with any of those." In fact, many Spanish-speaking respondents chose only the 'Hispanic' category because it was the one they identified with culturally and/or in terms of their ancestry and/or country of birth. For example, one Spanish-speaking respondent chose only 'Hispanic.' When asked why, she said, "Because I am from the Latin American continent, and the country I belong to is Mexico." However, she went on to tell the interviewer, "We are also White. White race from Mexico." She did not select 'White,' however, because, she said, the 'Hispanic or Latino' option was available and made more sense to her. The following are more examples of Spanish-speaking respondents' rationales for choosing only 'Hispanic:'

"Because I come from a Hispanic country, a country in Latin America."

"Because my native language is Spanish, and I was born in Puerto Rico. Being born in Puerto Rico makes me Latino."

"Hispanic' because my language is Spanish, and my I have Mexican roots, that is the reason why I answered that way. My ancestors are Mexican."

Some Spanish-speaking respondents, however, did choose both 'Hispanic' and 'White.' For example, one Spanish-speaking respondent described a personal definition of race and said, "I consider myself White by race. Ethnically, I describe myself as a Hispanic or Latino person." The choice of including 'White' was sometimes motivated by a social dimension of race and ethnicity – that is, identities that are defined by others. For example, one Spanish-speaking respondent said that, when she fills out forms, she is classified as 'White' because she is from Mexico City, "not from some province, I'm from the city itself, and for that they classify me as 'White." But she also chose 'Hispanic/Latina,' "Because I belong to Latin America, I was born there." Another Spanish-speaking respondent said, "I answered 'White' because I am a little light-skinned. And Latina because I am from Central America." She further explained that she began the practice of answering 'White' and 'Latina' when she came to the US.

'Hispanic' as an American construct: Several Spanish-speaking respondents, though having no issue with choosing 'Hispanic,' identified it as an American construct that they have more-or-less adopted when answering questionnaires. In this way, their answers reflect the social dimension of answering race and ethnicity questions in the US. As one Spanish-speaking respondent said, "They call us all Hispanics." Similarly, another Spanish-speaking respondent said, "I decided to answer this because I consider or have heard that all of us from countries where Spanish is spoken are Hispanic because we speak Spanish and Latinos because it's Latin America, South America." Another Spanish-speaking respondent described how she came to understand the word Hispanic. She said, "Before, when I had a social worker, she would assign me the option of 'White,' but then I started to understand that it is supposed to be 'Hispanic or Latino' because I come from Mexico." Finally, another Spanish-speaking respondent explained, "When we are here, when we immigrate to a different country, here you hear 'Hispanics.' And they mean those who speak Spanish. You always hear and you always understand that is

Hispanics, Latinos. That is how others refer to the persons who have immigrated from South and Central America to the North."

Afro-Latino/a respondents: Of the 59 Hispanic respondents, only six with backgrounds they described as both Black and Hispanic were interviewed (three in English and three in Spanish). This was due to time constraints and recruitment difficulties.⁷ However, those who were interviewed offered important insights. Like other Hispanic respondents, those with Black heritage answered one of two ways. A couple of respondents chose to mark both 'Black' and 'Hispanic' as their answer, but the others chose only 'Hispanic.'

Multiple responses: For two respondents the question worked as intended; that is, to capture respondents with both Hispanic and Black backgrounds. For example, one English-speaking respondent chose both 'Black' and 'Hispanic' and explained, "I would say I'm Black (I'm American) and I have some – I'm mixed with Spanish. So Black American mixed with Puerto Rican." Similarly, a second Spanish-speaking respondent explained, "I chose 'Black' for my skin color and 'Hispanic' because I speak Spanish and I'm from Central America. I know we, as Blacks, we come from Africans. I'm Latina because of my culture and customs."

One response: However, four other respondents chose only one option ('Hispanic'). For two respondents this was because they answered from a cultural perspective regarding their sense of self. During follow-up discussion, one English-speaking respondent explained to the interviewer:

"Um, yeah, so I'm actually mixed. I'm Hispanic/Latino and I'm African American. And I know you said I could choose, like, both, but I really, if it comes to just identifying myself, and I have the option, I just choose 'Hispanic/Latino,' because that's, even though I'm mixed, that's the culture I most identify with. I was raised in a Hispanic household. And that's, like, the only family that I know."

The rationale was the same for another English-speaking respondent who chose only one option. While she understood that there were different races in her ancestry, she saw herself primarily as Hispanic, specifically, Dominican. She said, "No. I mean, I am Dominican, but I do have Italian in me. African. I'm multiracial. Afro-Latina. So, I don't know if I need to click more than just 'Dominican' – which is where I'm from." The interviewer asked how she would mark it on, say, a Census form. The respondent replied, "Just like I did now. Just Dominican."

Two other Spanish-speaking respondents chose only one option, not so much because it reflected their identity, but because they thought it was customary do so in the US. For example, one Spanish-speaking respondent chose only 'Hispanic' but during the interview discussed how he also has African ancestry. When asked why he did not choose 'Black/African American,' he described his impression that the American standard is for individuals to claim only one race. He said:

"There was a lot of mix in Venezuela so classifying someone with one race or ethnic group is too difficult. Here they force you to be one thing or another when you are a mix. You now do it automatically because you learn how the US system is...Many of us Latinos, we don't know what specific groups we come from. But I know I am a mix of African blacks and also Spaniard blood from Whites."

Another Spanish-speaking respondent had the same perspective. He thought of himself as Hispanic but talked about how he could answer 'Black/African American' if 'Hispanic' was not an option. He said, "In my country, Venezuela, dark skin people are called 'negro.' Before moving here, I used to live in (state). I had to complete a

⁷ People who answered both 'Black' and 'Hispanic' on the recruitment screener were included in the sample in an effort to interview Afro-Latino/a respondents. This was no guarantee that the respondents screened into the study self-identified as Afro-Latino/a; however, there was not sufficient time to work through Afro-Latino/a community organizations in order to recruit respondents who identified specifically as Afro-Latino/a.

⁸ Note that 'Negro' in Spanish refers to skin color, not to a group of people, per se.

health form, and the Hispanic/Latino option was not available, so, I answered 'Black.' But they told me that I could write Hispanic." His impression was that he should choose only one category.

In sum, although the structure of the question (containing many subcategories) prompted two respondents with Hispanic and Black backgrounds to answer the question based on their genealogy, this was not the case for four other respondents with Hispanic and Black backgrounds. They answered the question based on the way they understood race and ethnicity, which was shaped by either a personal cultural perspective or a social understanding of race (which was informed by the way they believe race is conceptualized in the US).

III. Reluctance to Answer

Racial and ethnic identification can be a politically sensitive topic for certain groups. The combined question seeks more information than usual from respondents in that they are asked to provide extensive detail about their race and ethnicity – including non-Hispanic ethnicities. But when the purpose of forms or surveys is not clear to respondents, the source may not be fully trusted. This creates the potential for either missing data or for respondents to change how they may normally answer. For example, one respondent marked 'Asian' but no subcategory. When asked about this omission, she explained:

"I would not specify...And now, like I said, I just don't put it because...knowing a lot of history in this country and what this country has done, like interned the Japanese Americans, and like I said, especially when they don't break out the other groups...Like for example for White, if it just says White. You know, if you are only going to target a certain group then..."

The respondent went on to admit to the interviewer that if she were to provide a subcategory, it would have been Chinese.

Another respondent with Russian lineage also chose to omit that from his answer. Similar to the previous respondent, he expressed concern about revealing such information. For him it was related to the war in Ukraine. He said:

"Well, I know about my parents and grandparents and where they came from. On my...father's side it was from Belarus, which at that time was Russian. So that's why I said that...I was a little hesitant. Because Russia is not a good place now. I wouldn't say it's shameful, but with what Russians are doing in Ukraine, I certainly wouldn't identify as a Russian."

MENA respondents also sometimes acknowledged the risk associated with reporting their race. As one respondent said, "There is a negative vibe to being Middle Eastern." However, this person *did* mark 'MENA,' citing the importance of inclusion and representation for the group as a whole. On the other hand, another respondent with reservations about the MENA category chose to mark only 'Hispanic/Latino' as his survey answer. It was only later in the interview that he revealed his MENA background to the interviewer. He said:

"I consciously avoid the ancestry of my dad because being born and raised in America, I still consciously want to belong. And by focusing on my mom's Latino side, I belong faster than talking about my Middle Eastern side or my dad's side because of the level of racism in America. And Iraq and Afghanistan and all these wars we had, people can be really nasty to me because of my dad's side. So, I don't really talk about that side that much. I avoid it."

When the interviewer confirmed that he would not mark MENA, he said, "No, because of racism and discrimination. A lot of people out there for some reason don't like Middle Eastern people. I don't like to label myself as that."9

⁹ This pattern was observed more often in previous work on the combined question (see Willson and Dunston, 2017) than in this study, which may be due to perceptions of a shifting political climate.

This phenomenon was also observed among Hispanic respondents. For example, one Spanish-speaking respondent discussed how sometimes people are afraid of giving details about their race or ethnicity for fear of discrimination. As a result, even though she did answer 'Mexican or Mexican American,' she does not always answer the same way. She said she prefers not to give details when she is talking to someone "scary" like a lawyer or the police. In such cases she just reports 'Latina' and declines to elaborate.

The above examples illustrate that the detailed nature of the combined question will not operate in a vacuum. The context in which it is asked may have an effect on how and whether people – particularly those who belong to certain racial or ethnic groups – decide to offer a complete answer.

IV. Noted Difficulties with Certain Subcategories

Confusion with 'Mexican or Mexican American' subcategory: While most of the subcategories refer to a non-US country of origin, 'Mexican or Mexican American' has a dissimilar framework. Because it is meant to include both American-born and Mexican-born people, this creates the potential for confusion, as was demonstrated in a couple of Spanish interviews. For example, one Spanish-speaking respondent did not understand this subcategory as including people of Mexican descent who were born in either the US or in Mexico. She left it blank and said, "I studied the list of all the races [options] and I'm 'Hispanic or Latino.' I was expecting to see Mexican, but I only see Mexican American." Although the subcategory does include 'Mexican,' combining it with 'Mexican American' shifted the meaning of the subcategory and made its intent unclear, especially in relation to logic of the other subcategories.

<u>Confusion with 'English' Subcategory</u>: Several respondents were confused over the term 'English' as a subcategory of 'White.' This confusion was observed for the self-administered test instrument only. Respondents could see the question in its entirety but did not usually read the question stem or instructions. Instead, respondents often just looked at the categories and clicked those that were relevant. As a result, 'English' was sometimes mistaken to mean language. For example, one respondent said, "I speak English, but I didn't think [to choose] 'White' because I'm not White, I just speak English as my basic language." Another respondent made the same error but did not realize her mistake until the probing portion of the interview. She said:

"You know what? Now that I'm looking at it, maybe I shouldn't have chosen 'English.' I'm thinking...it's 'White...what do I speak?' And then I'm like – oh 'race or ethnicity' that's dumb [meaning her interpretation as 'what language do you speak']. Now I'm looking at all this, wait...do I uncheck it? I'm not English, I'm White."

Confusion over the term 'English' seemed to be largely a function of the self-administered format. The next section discusses other observed mode effects.

V. Mode Effects

The survey instrument was tested in two modes – interviewer-administered and self-administered – for the combined race and ethnicity question. While this was primarily not a mode effect study, some mode issues were observed. Earlier it was noted that the response categories defined the intent of the question. Hence, when difficulties arose, they often centered around the response options (such as misunderstandings of the term 'English' in the previous examples). Moreover, similar patterns of difficulty were observed in both the English and Spanish interviews.

Interviewer-administered issues

Wording for the interviewer-administered version was adapted from the self-administered version and, as such, was somewhat awkward to read aloud. Improvements to the question stem and instructions could be made in order to allow for a smoother verbal delivery.

Most difficulties that arose, however, are related to the sheer number of subcategories, of which there were seven under each main race category. This made the question somewhat cumbersome to administer. As a result, there were occasions in which respondents either asked the interviewer to repeat the categories or, more often, offered answers before hearing all the choices. In the virtual interview, where the interviewer and respondent could see each other, this was not a significant impediment, but it might prove to be more problematic (in terms of comprehension and survey length) when administered over the telephone.

By comparison, however, the interviewer-administered version outperformed the self-administered version, which is discussed next.

Self-administered issues

The self-administered mode presented more and different challenges. Specifically, it was not clear to some respondents how they should complete the form. The form was offered as a fillable PDF. Respondents were given control of the computer screen so that they could complete it on their own. Some respondents had no difficulty with the task and filled in the form as intended. However, many respondents did not. For example, some respondents did not physically check a main race category; they only selected a subcategory and assumed that the main category was either implied or unrelated.

In addition, many respondents did not understand that they had the ability to type in a subcategory that was not offered as a check-box option (essentially, this open field was an 'other-specify'). The form listed each main category with an instruction to "provide details below" alongside the option (for example: WHITE - Provide details below...). Respondents did not always understand how to proceed. For example, one respondent checked the main category, 'White,' with no difficulty. However, he was confused about the subcategories because he did not see the open-text box below the subcategories which read: Enter, for example, Scottish, Norwegian, Dutch, etc. He only saw the categories with check boxes. Unsure of what to do, he said, "Now, when it says 'White' and all the different - Irish, English, and French - I'm trying to think the best thing to put there. Because my ancestry is half Russian and half French. But I'm just your basic White, Caucasian American. So should I just click "White?" In other words, because one of his ethnicities (Russian) was not represented with a check box, he thought he had no way to include it. Thus, to check a single box (French) would be a misrepresentation of his ethnic background. Another respondent illustrates the same dilemma. She said, "Well, it says enter additional details, so I guess I can click 'French,' but I'll have to leave the Scottish and the Russian ancestry blank, I guess. Because it's not an option. So, I'll click 'French' and 'White' and then we'll have to leave the Scottish and the Russian to the imagination." The interviewer asked whether the respondent saw the other-specify open field: "Oh! No, I didn't, actually." On the other hand, one respondent who did see the open-text field did not necessarily understand its purpose. She saw 'provide details' and thought it was asking for the entire family tree associated with her answer of 'Mexican or Mexican American.'

Many of the problems associated with the self-administered mode, as tested in this study, could be minimized in different formats. It is possible, for example, that a web format could more effectively guide respondents through the process of answering. For instance, a web survey could expose respondents to the main categories first (potentially eliminating missing data due to misunderstandings), followed by the corresponding subcategories. In addition, the purpose of the open-field, 'other-specify,' option could be better displayed as well.

REFERENCES

Lincoln, Y. and E. G. Guba. (1985) Naturalistic Inquiry. Beverly Hills, CA: Sage Publications.

Miller, K. and Willson, S. (2002) Cognitive Testing of NCHS Race Questions, Hyattsville, MD: National Center for Health Statistics.

Miller, K., Willson, S., Chepp, V., & Padilla, J.-L. (2014) Cognitive Interviewing Methodology: A Sociological Approach for Survey Question Evaluation. Hoboken, NJ: Wiley.

Strauss, A. C. and J. Corbin. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, 2nd edition. Newbury Park: Sage Publication.

Suter, W. N. (2012) "Qualitative Data, Analysis, and Design." In *An Introduction to Educational Research: A Critical Thinking Approach*, 2nd edition. Thousand Oaks, CA: Sage Publications.

Willis, G. B. (2005) *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage Publications.

Willson, S., & Dunston, S.K. (2017). Cognitive Interview Evaluation of the Revised Race Question, with Special Emphasis on the Newly Proposed Middle Eastern/North African Response Option. National Center for Health Statistics. Hyattsville, MD.

ATTACHMENT 1: ENGLISH SELF-ADMINISTERED TEST INSTRUMENT

	t is your race or ethnicity? et all that apply AND enter additi	onal details in the snaces h	nelow
	, you may report more than one		ciow.
	WHITE - Provide details below.		
Ш	German	Irish	English
	Italian	Polish	French
	Enter, for example, Scottish, Nor	wegian, Dutch, etc.	
	HISPANIC OR LATINO - Provide	details below.	
	Mexican or Mexican American	Puerto Rican	Cuban
	Salvadoran	Dominican	Colombian
	Enter, for example, Guatemalan	, Spaniard, Ecuadorian, etc	c.
	BLACK OR AFRICAN AMERICAN		- Haisian
	African American	Jamaican	Haitian
	Nigerian	Ethiopian	Somali
	Enter, for example, Ghanaian, S	outh African, Barbadian, e	tc.
	ASIAN - Provide details below.		Asian Indian
	Chinese	Filipino	Asian Indian
	Vietnamese	Korean	Japanese
	Enter, for example, Pakistani, Co	ımbodian, Hmong, etc.	
			le, Navajo Nation, Blackfeet Tribe,
	Mayan, Aztec, Native Village of	ваггом іпиріат Ігіваі Gove	ernment, Tiingit, etc.
_			
	MIDDLE EASTERN OR NORTH A		Egyptian
	Lebanese	Iranian	
	Syrian	Moroccan	Israeli
	Enter, for example, Algerian, Ira	ıqı, Kurdish, etc.	
\neg	NATIVE HAWAIIAN OR BACIFIC	SLANDER - Drouido dotoilo	halow
┙,	NATIVE HAWAIIAN OR PACIFIC I	Samoan Samoan	Chamorro
١	Native Hawaiian	Fijian	Marshallese
	Tongan		ividistidilese
	Enter, for example, Palouan, Tah	itian, Chuukese, etc.	

ATTACHMENT 2: SPANISH SELF-ADMINISTERED TEST INSTRUMENT

¿Cuál es su raza o grupo étnico? Seleccione todas las opciones que correspondan Y ponga más detalles en los espacios que aparecen abajo. Se puede seleccionar más de un grupo.



