

Do Race and Color Still Matter? -- Considerations on the Combined Question

Howard Hogan, Ph.D.

Chief Demographer (RET)

U.S. Census Bureau

Analytic Framework

**If you don't know what you need to measure,
any question will do.**

Statistics are gathered to aid in making decisions. Good statistical practice starts with an understanding of the decisions to be made (there are normally many) and the information needed to make these decisions. One must always start with a clear concept of what one would like to measure. Seldom can one measure exactly the desired concept; there are many steps in the measurement process. However, without a clear concept of what one is trying to measure, there is no way to evaluate the measurement design, including, in this instance, the question design.

Useful social, or at least statistical, constructs of race and ethnicity would have three properties:

- (1) be recognized by society and the individual;
- (2) categorize individuals into the same groups over a long period of time;
- (3) be predictive of social and economic opportunity.

(See Humes and Hogan, 2009)

Those proposing modifications to the race question need to show how these criteria are met.

Much of the current discussion on 'the combined question' is vague as to what needs to be measured: 'race' or 'ethnicity' or 'both race and ethnicity', or perhaps something vaguely described as 'identity,' 'community' or 'category.'

The Office of Management and Budget (OMB) has adopted the concept of national or ethnic self-identification. For example:

The category ‘White’ includes all individuals who identify with one or more nationalities or ethnic groups originating in Europe, the Middle East, or North Africa.

The only exception being:

The category ‘American Indian or Alaska Native’ includes all individuals who identify with any of the original peoples of North and South America (including Central America) **and who maintain tribal affiliation or community attachment.** (Emphasis added).

It is unclear why self-identification, as opposed to community attachment or lived experience is the most appropriate concept for the intended Federal purposes.¹ It is also unclear how nationality maps into a question purporting to be about race. To my knowledge neither OMB or the Census Bureau have provided such an analysis.

This ambiguity is further deepened with the proposed, and likely to be adopted, addition of Hispanic origin and Middle East/North African categories to the ‘race’ question. Implicit in much of the discussion is an assumption that since race is hard to measure, we no longer need data on it.²

A definition of ‘race’

Since the term ‘race’ can have different meaning to different people and in different contexts, I begin by defining how I will use the term:

Race refers to biologically inherited superficial physical characteristics perceived to be important by society.³

Thus, if a group of people is perceived and treated as different in American society because of their superficial appearance, that group is considered a race. It is consistent with how the terms race is defined in law. In the United States, the concept of race relates strongly with skin color. A primary purpose of the census is to help the

¹ With the increasing popularity of Ancestry.com and similar sites, the connection between perceived ethnicity and lived experience may become considerably weakened.

² I prefer the term Hispanic to Latino for a specific reason. Latino would seem to include Brazilians, who are excluded from the Federal definition and exclude Spaniards, who are included.

³ ‘Superficial’ is used here to include especially skin color, but also hair texture, nose shape, etc. Male/female, tall/short, etc. are also inherited characteristics perceived to be important, but do not define race.

enforcement of civil rights and voting rights laws where ‘race or color’ occupy a unique position.

‘Race’ may indeed be a ‘social construct,’ but that doesn't imply that it is not real. One can always argue that there are no precise dividing lines between the races. However, that doesn't imply that American society no longer distinguishes White from Black, or Asian Indian from American Indian. Of course, whatever groups are defined, there will always be sub-groups that can be further and further divided. Meaningful social statistics must always classify people into groups for useful social policy. Compromises are always necessary.

Does Hispanic Constitute a Separate Race?

The 1997 OMB standards and common sense both agree that people of Hispanic origin can be of different races. Of course, a large percentage of Hispanic people are descendants of the South and Central American native peoples. For lack of a better term, I will refer to this group as Indígena/Mestizos⁴. The term ‘Hispanic/Latino’ is often used to refer to only this one segment of all Hispanics.⁵ This misuse of the term ‘Hispanic/Latino’ confuses the issue as to when one is discussing all people of Hispanic ethnicity and when one is discussing only those who are viewed as racially Indígena/Mestizos.

Is Race a Meaningful Concept for Hispanics?

Here one must be very careful to distinguish between three quite different issues.

The first is whether Hispanics of different races are treated differently by American society. Is a 3rd generation Hispanic American of Indígena appearance treated differently from a 3rd generation Hispanic American of European appearance? Are White Hispanics treated differently from Brown or Black Hispanics? The answer is clearly yes; the statistics demonstrate this. (See for example Hogan, 2016.)

⁴ If anyone can suggest better terms, I would be interested.

⁵ Hence the headlines ‘Genetic variant helps protect Latinas from breast cancer,’ when the research only applied to those of indigenous ancestry.
(<http://www.sfgate.com/health/article/Genetic-variant-helps-protect-Latinas-from-breast-5835931.php>)

A different question is whether Hispanics understand the concept of race. Here the answer is also clearly yes. Mexicans can distinguish Indígena Mexicans from White Mexicans; Cubans can distinguish Black Cubans from White Cubans. The lines may be drawn differently, but it is a myth that South Americans don't see race.

The third question is whether many Hispanics see themselves in the particular wording and options offered to them on official Census Bureau forms. This is discussed in the next section.

What about the large number of Hispanics who respond with 'some other race'?

We need to address two questions. Why many Hispanics choose to write in a response under 'some other race'; and why this matters.

Many immigrants think that 'African American' does not refer to them; some do not think of themselves as being 'black.'⁶ Few Indígena/Mestizos identify with the term 'American Indian,' especially when asked then to write down their 'tribe.' Thus, having not given many Hispanics racial terms to which they can relate, the Census Bureau is surprised that they don't relate to the terms it gives them.

There is an additional factor at play. The list of 'races' includes many nationalities, for example Chinese, Filipino, Guamanian. In the 2020 Census, all race categories were mapped to particular nationalities or ethnic groups. Many people, including a large number of immigrants, not unreasonably infer that the question is about national origin, and thus write in the name of their home country. Haitians do this; Jamaicans do this; Iranians do this. However, these non-Hispanic groups have their answers re-coded. Nearly all (~95%) of Hispanics coded into 'some-other-race' have responded with a specific national origin.

This problem was made worse in 2020 by the way that the Census Bureau asked and coded the race question. If, say, someone from Portugal checked White and wrote under it 'Portuguese,' the person was coded as 'White Alone.' However, if someone from Spain checked White and wrote under it 'Spanish,' that person was coded 'White and Some-Other-Race.' The same coding rule was applied to all people marking only

⁶ Interestingly, evidence suggests that some Spanish speakers may have identified with the now-eliminated term 'Negro,' as this is the cognate to the Spanish term for black.

one race but writing in a Hispanic origin under that race category, the response was coded with that race and some-other-race. The coding rules made it difficult and confusing for people to report that they were single race Hispanics. So, while it is true that nearly all the people the Census Bureau tabulates in some-other-race are Hispanic, this only reflects its own editing and coding process.

Why is the number marking ‘Some other Race’ a problem?

One answer is that this is only a symptom of the real problem: the failure to record a race for many Hispanics. However, if this is the concern, the combined question approach cannot be the solution, as it records a race for significantly (both statistically and substantively) fewer Hispanics.

The assessment of non-response must be done with respect to the tabulations or analysis desired. Further, it should be obvious that when comparing the combined one-question approach with the separate two-question approach, one must use the information gathered from both of the two questions. Looking at the results of the ACQ and NCT correctly shows that the two-question approach does not increase non-response. That is to say, that the two-question approach does not reduce the responses that can be placed in any usual tabulation categories. Consider the two most common ‘race’ tabulation categories:

By race and Hispanic origin:

- Hispanic
- Non-Hispanic White
- Non-Hispanic Black
- Non-Hispanic Asian
- Etc.

By the five OMB race categories.

- White
- Black
- Asian
- Etc.

It is clear that the two-question design does equally well on the first and considerably better on the second. Switching to the combined approach treats the symptom (many

responses in some-other-race) at the cost of making the actual problem (too few people with recorded races) worse. Clearly, this cannot be the real justification for a change. It could be that the responses in some-other-race are a problem in and of themselves. Perhaps they cost too much to code or are too difficult to tabulate and report. This seems an unlikely justification for such a major change. Finally, perhaps needing to report in this category constitutes an undue and unnecessary burden on respondents. If we could gather the same information with lower burden, we should. This justification is discussed in more detail below.

Two Justifications for the Combined Question Approach.

It seems that there are two logical justifications for the combined question approach. One is that, in this ‘post-racial’ America, we no longer need data on race or color; what is important is ethnic-group identification or ‘community.’ This justification is implicit in the often-stated justification for the combined approach ‘Hispanics tell us that they see themselves better in the new approach.’ If we are trying to measure race, and if Hispanics can be of any race, then the relevance of this statement is not clear. This is a valid argument only if the underlying concept we are trying to measure is ethnic-group self-identification, and we no longer need data on race or color.

However, this position doesn’t seem to be the one that advocates for the combined question approach are willing to embrace, or to articulate the underlying concept that they are attempting to measure. So, the justification must lie elsewhere.

The underlying argument seems to be:

‘The Combined Question Approach allows us to collect data on race that is at least as good as we currently have and at lower cost, lower burden and with increased information on ethnicity.’

This argument is highly attractive. We gain much and lose nothing: data on race of equal or better quality is available. It means that OMB can adopt a combined one-question with no loss of useful information. However, this assertion is premature.

There are two parts to this reasoning. First is the assertion that reporting of race among Black, AIAN, etc. Hispanics will remain the same or even perhaps increase. Thus, the Census Bureau’s statements that:

‘Results for the [AQE] Combined Question: Did not reduce Black, AIAN, Asian, NHPI.’

The same claim was made for the 2015 National Content test. On a purely statistical basis, **this statement is not true**. The AQE was inconclusive on this point. A careful analysis of the NCT results indicates that the combined question probably did reduce the reporting of Black.

Equally important, but often neglected, is the fact that AQE only tested reporting among households who responded by mail. This was also true for the National Content Test. We have no data on how these questions will work in non-response follow up (NRFU), or how they might work in Group Quarters. This will depend on many things, including training. How these concepts will work with administrative records is also not clear.

The second part of the argument is more subtle. It is along the lines:

‘Although there will be a large reduction of Hispanics who mark White, those Hispanics who would mark White in a separate question, but not in the combined context, are not really White or at least don’t consider themselves as White.

If true, this is good. Black Hispanics will consistently mark Black and Hispanic; Asian Hispanics will consistently mark Asian and Hispanic; White Hispanics will consistently mark White and Hispanic. The data on Whites become more accurate by eliminating non-White Hispanics from ‘white’ response category. The ones marking only Hispanic would all be Indígena/Mestizos who do not maintain tribal or community identification. The ‘Hispanic Alone’ category would define a race.

There are two serious concerns. First, as mentioned above, there is evidence that the combined question will reduce the reporting of Black. Secondly, as mentioned above, the AQE and the NCT only tested household responses by mail. As the propensity to mail back a questionnaire is highly correlated with race, color, and ethnicity, it is likely that the NRFU universe will react differently.

Secondly, there is a problem caused by the inclusion of examples. The combined question places Hispanic ethnicity in parallel with race. Those who care deeply and

think carefully will know that they can and should mark both, as appropriate. These are the people most likely to respond by mail, including to the AQE, the NCT and the Census. However, a person of, say, Mexican origin will see ‘Mexican’ as an example only under Hispanic. The person will not see it as an example under, say, AIAN. Many people will, quite logically, decide that this is the category that the U. S. Government wishes them to choose.⁷

The Overwhelming Importance of Race and Color in US Law and Society

Race and color play a unique role in the Constitution and the civil rights laws, and the court cases that have enforced them. It may be true that the dividing lines between ‘races’ are not as distinct as they were 100 years ago. It may be that perception of ‘color’ is beginning to replace ideas of distinct races.⁸ Still, these rights will be harder to enforce if the Federal Government is no longer able to provide tabulations by race.

For the past several censuses, the Census Bureau has provided the Department of Justice tabulation of the population using the five recognized race groups, the so-called ‘MARS’ files. These tabulations will no longer be possible if a combined question is adopted. Obviously, the tabulations of the population by race provided to the states under PL-94-171 will also not be possible.

In the United States for the past 400 years, two factors have determined social and economic advancement: immigration status and race/color.

Waves of European immigrants have entered the U.S. Each has suffered hardship and discrimination. However, by the third generation, when the members speak American English and have received an American schooling, the discrimination has tended to lessen. Immigrant status changes over the generations.

This has not been true for people of color. There have been Black Americans here for about sixteen generations, and they still face racial discrimination. American Indians have been here even longer. East Asians have lived here for six generations. Race and color continue to matter.

⁷ This issue was raised at a National Advisory Committee and during the AQE focus groups.

⁸ See for example, Kevin Brown ‘The Rise and Fall of One-Drop Rule’ in COLOR MATTERS: SKIN TONE BIASES & THE MYTH OF A POST-RACIAL AMERICA (Kimberly Jade Norwood ed. 2013).

One of the most important social, economic and political questions for our future will be whether third or fourth generation Hispanics will be treated as a group, regardless of their race and color, or whether their opportunities will depend on their races, as it does for non-Hispanic Americans. Will Hispanics of European origin advance socially and economically, while those of African or Indígena ancestry are left behind? There is evidence that this is already happening. But, only a separate question can answer this question⁹.

Middle East and North African (MENA)

The discussion above has focused on combining the Hispanic origin with the ‘race’ question. Similar issues will arise with the addition of a MENA group. Historically, many immigrants to the US from the middle east were Lebanese Christians, who fit neatly into the ‘White’ category. (Think of the Ali Hakim character in the musical Oklahoma.) Now, the MENA immigrants and their descendants represents a spectrum of nationalities, colors and races. The issue is not how well these groups get along in their countries of origin. Rather, it is whether members of this group, on the whole, are treated similarly by U.S. society. Do Ashkenazi Israelis share common experiences in America with, say, Somalis? Other the OMB race groups are defined as those descending from ‘the original peoples’ of a particular geographic area. Would American Jews identify a descending from the Land of Israel, and thus as MENA¹⁰? Should the Parsis of India be classified as MENA or Asian? Is the real purpose to measure Islamophobia and antisemitism? There is no doubt that many of the peoples form this region face very real discrimination in American society. One still needs to ask what is the best way to capture this reality.

Conclusion

Whether one agrees with the continued centrality of race in American law and society, it should be incumbent on those advocating a combined question approach to articulate

⁹ For many decades the Census asked about ‘Parental Place of Birth.’ This was dropped in favor of ‘Ancestry’ in the 1980 census. With nationality now being ask as part of the ‘race’ question, asking ‘ancestry’ on the ACS seems to add little new information. Serious work needs to be done to bring back ‘parental place of birth.’

¹⁰ See for example

<https://blogs.timesofisrael.com/what-a-mena-racial-classification-would-mean-for-american-jews/>

clearly what they believe needs to be measured, and then construct a consistent analytic framework to assess how well the different approaches achieve their goals. In order to best enforce our civil rights laws, if trade-offs need to be made, do we need data on ethnic self-identification or on race and color?

REFERENCES

- Hogan, H. (2017). Reporting of Race Among Hispanics: Analysis of ACS Data. In: Swanson, D. (eds) *The Frontiers of Applied Demography*. Applied Demography Series, vol 9. Springer, Cham. https://doi.org/10.1007/978-3-319-43329-5_92009
- Karen Humes and Howard Hogan, 2009, "Measurement of Race and Ethnicity in a Changing, Multicultural America," *Journal of Race and Social Problems*, DOI 10.1007/S12552-009-9011-5.