

Maintaining Race and Ethnicity Trend Lines in U.S. Government Surveys

Elizabeth Greenberg¹, Jon Cohen², and Dan Skidmore³

In 1997 the U.S. Office of Management and Budget issued guidelines that require federal government surveys to allow respondents to report multi-racial identifications. Maintaining trend lines when reporting racial breakdowns requires a linkage between the old and new measures of race and ethnicity. For reasons of both cost and statistical efficiency, linking studies are best based on data in which the same sample is asked both the old and new race/ethnicity questions. However, if respondents are influenced by the order in which the questions are asked, complicated statistical modeling will be required to estimate and adjust for the influence of the joint administration of the items to a common sample. This article uses data from the National Assessment of Educational Progress 2000 field test to show that the sequence in which race/ethnicity questions are administered has a small but important effect on the likelihood that respondents will identify more than one race/ethnicity.

Key words: Race and ethnicity linking; racial identification; statistical trend; priming effect.

1. Introduction

In the United States virtually all government social statistics are reported by racial categories. This reporting convention reflects the history of U.S. race relations, and a body of laws such as those derived from the Civil Rights Act of 1964 that requires equal access, opportunity, and treatment regardless of racial identity. Active interest groups and civil rights laws ensure that any disparate effect of governmental decisions across racial groups will be identified and questioned.

This system of accountability presumes the ability to track data by racial groupings over time. Maintaining trend data allows stakeholders to monitor changes in racial disparities and identify changes that may correspond with changes in policies or demographic conditions in the country. For example, the Departments of Labor and Commerce join forces to track employment and income by racial groups using the Current Population Survey (Current Population Survey 2000); the Department of Commerce reports income, employment, and participation in federal programs by racial groups through the Survey of Income and Program Participation (U.S. Bureau of the Census 1998); data from

¹ American Institutes for Research, 1000 Thomas Jefferson St., NW, Washington, DC 20007, U.S.A.
Email: egreenberg@air.org

² American Institutes for Research, 1000 Thomas Jefferson St., NW, Washington, DC 20007, U.S.A.
Email: jcohen@air.org

³ Le Moyne College, 1419 Salt Springs Rd., Grewen Hall 402, Syracuse, New York 13214, U.S.A.
Email: skidmodl@lemoyne.edu

Acknowledgment: The work on which this article was based was performed pursuant to Cooperative Agreement Number R902 F 980001 of the Institute of Education Sciences, National Center for Education Statistics. It does not, however, necessarily reflect the views of that agency.

the decennial census is used to review racial disparities that may arise in political representation due to redrawing of congressional districts mandated by Public Law 94-171; the Department of Health and Human Services tracks and reports racial disparities in health status through surveys such as the National Health Interview Survey (Adams, Hendershot, and Marano 1999); and the Department of Education tracks racial disparities in educational achievement through the National Assessment of Educational Progress (U.S. Department of Education 1999).

In 1997 the U.S. Office of Management and Budget (OMB) issued a directive that changes the way government agencies gather data on racial identification. An increasingly vocal minority, who identifies with multiple racial groups, has brought attention to the fact that the survey questions currently used by federal agencies require respondents to select a single racial identification (U.S. Office of Management and Budget 1997). This constraint both alienates multi-racial respondents and inaccurately characterizes the racial identification of the U.S. population. Therefore, OMB has issued guidelines that require federal government surveys to allow respondents to report multi-racial identifications.

Whatever the merits of the intended changes with regard to the validity of the data generated, virtually all U.S. agencies will face the challenge of maintaining viable trend lines as the racial classification measures change. This problem will appear in two incarnations: first, agencies will be faced with the challenge of developing some sort of mapping, linkage, or interpolation between the older and newer measurement schemes; second, as the public becomes more accustomed to seeing multi-racial groupings in reported data, individuals' self-identifications may shift. In its first incarnation, the problem is simply one of analysis and reporting. In the second, agencies face a more difficult challenge of interpretation as membership in the groups across which we look for disparities shift and change. Here, we address only the first issue.

We see two general approaches to maintaining the trend lines that the system of racial accountability requires:

- Maintain two parallel trend lines – one based on the old item, the other based on the newer one, for a finite number of years; or
- Develop statistical models of the relationship between the old and the new items based on data collection in a base year, and use this model to reveal the “latent trend” in the new reporting categories for older data sets.

Both approaches require collection of parallel data using the two different versions of the racial identification question. Both cost and statistical efficiency favor gathering the required data from a common sample – that is, asking both items of each respondent.

Maintaining two parallel trend lines using a common sample is easier if the response to each item is unaffected by the response to the other item. Failing this, more complicated statistical modeling, along with attendant assumptions, is required to estimate and adjust for the influence of the joint administration of the items to a common sample. Alternatively, the old and new versions of the race and ethnicity items could be administered to two separate randomly equivalent groups. However, this strategy would greatly increase the required sample size and therefore greatly increase the cost of data collection.

Past work on this issue has presumed independence of the responses to two different items during the same administration. A previous study (Tucker and Kojetin 1996) based

on data from the May 1995 Current Population Survey (CPS) supplement reported on the properties of four different versions of a race and ethnicity measure. All respondents to the May 1995 CPS were first asked about their race and ethnicity using the CPS question in effect in 1995, and subsequently asked one of four new versions of the race and ethnicity question.⁴

Although all respondents to the May 1995 CPS supplement were asked the questions in the same order, preliminary cognitive testing of race and ethnicity items by the CPS in preparation for the supplement indicated that respondents might be influenced by the order in which questions were asked. As part of the cognitive testing process, respondents were asked to identify the race and ethnicity of remote ancestors. Although the sample size for the cognitive testing was not large enough to consider statistical significance (82 participants), respondents who were asked about the race and ethnicity of remote ancestors before being asked a question about their own race were more likely to consider saying they were multi-racial when answering the race question than respondents who were asked about remote ancestors after answering a question about their own racial identity (McKay and de la Puente 1996).

Given the nature of racial identification combined with the long history of single-race reporting, one might expect a priming effect from the first question asked on the second question asked.⁵ It seems reasonable to expect that some respondents, after being forced to identify with a single race group, would hesitate to identify with multiple groups in a subsequent response. Similarly, respondents who are offered the multi-racial version first might chafe at a subsequent question that eliminates the multi-racial option.

Does administering one version of the race question influence responses to a different, subsequently administered version? The present article reports the results of a study designed to answer this question. A 2000 field test for the U.S. Department of Education's National Assessment of Educational Progress (NAEP) included an experiment to test for dependence among responses to different versions of the race question. Specifically, respondents in grades 8 and 12 answered both the new (multi-racial) item and the older (single-race) item, with the order of administration randomly assigned. The two responses were separated by an extensive distractor task (a challenging 50-minute exam in history and geography). As we will present below, we found small but important effects from the order of administration.

We have no way of determining whether the results of this study can be generalized beyond the population included in the NAEP field test: students in eighth and twelfth grades in the spring of 2000. However, this is an important population to examine in this type of study, because the incidence of multi-racial identity is higher among younger cohorts in the United States than among the population as a whole. In the 2000 Census, 25.7 percent of the population was under age 18, but 41.9 percent of individuals reporting two or more races were under age 18 (U.S. Census Bureau 2001).

⁴Participants were divided into four panels and answered questions as follows: Panel 1: separate race and Hispanic-origin questions, no multi-racial category; Panel 2: separate race and Hispanic-origin questions, with a multi-racial category; Panel 3: a combined race and Hispanic-origin question, no multi-racial category; and Panel 4: a combined race and Hispanic-origin question, with a multi-racial category.

⁵See Sudman, Bradburn, and Schwarz (1996) for a discussion of priming effects.

The remainder of this article proceeds in three sections. The next section, Sample and Method, describes the experiment built into the NAEP field test and some of the response characteristics of the race/ethnicity items tested. Section 3, Results, presents findings from our analysis of the effects of the order in which the questions were administered. The final section discusses the implication of these findings for maintaining trend lines as measures of the classification variable change.

2. Sample and Method

The NAEP 2000 field test included an experiment that was designed to examine whether the administration of a single-response version of the race/ethnicity question influenced responses to a subsequently administered multiple-response version. This study compares the single-response version currently in use with two alternative multiple-response versions: a one-question version, which includes Hispanic among the racial categories, and a two-question version which asks about Hispanic ethnicity in a separate question.⁶

2.1. Sample

The sample for the NAEP 2000 field test was a multistage probability sample based upon the sampling frame used for the 2000 operational assessment. Counties or groups of counties were the first-stage primary sampling units (PSUs). Fifty-two PSUs were selected for the field test sample. In the second stage of sampling, schools in those 52 PSUs were then randomly selected for the field test sample, with schools with more than 15 percent black plus Hispanic enrollment sampled at twice the rate of other schools. Schools that had already been selected for national or state operational samples or had fewer than 25 students per grade were eliminated from the field test sample. In the third stage of sampling, students were selected within schools. The product of the response rates for the student and school samples exceeded 70 percent.

2.2. Method

Eighth and twelfth grade students were randomly assigned booklets with one version of the race/ethnicity item administered before the 50-minute NAEP assessment and a second version of the race/ethnicity item administered after the NAEP assessment. This random assignment of booklets was implemented at the student level, so that some students in each sampled school received each possible combination of the new and old race/ethnicity items. The majority of students received booklets that paired one of the two new versions of the race/ethnicity item with the current version of the item that NAEP has administered since 1986. The current version is a two-question version in which Hispanic is included among the race choices in the first question, followed by a second question that asks Hispanic students to identify a particular Hispanic ethnicity (Mexican, Puerto Rican, Cuban, or other). In the current NAEP item, students are permitted to identify only one race, in contrast to the two new versions in which students are permitted to select more

⁶The complete text of the questions is shown in Appendix A.

than one race. Table 1 shows the number of students in the field test sample who received each version of the race/ethnicity item before and after completing a history or geography assessment.

Table 1. Booklet design for NAEP 2000 field test race/ethnicity items

Grade	Number of students	Race/Ethnicity item administered before assessment	Race/Ethnicity item administered after assessment
8	500	N1	C
8	500	N1	N2
8	500	N2	C
8	500	N2	N1
8	500	C	N1
8	500	C	N2
12	500	N1	C
12	500	N1	N2
12	500	N2	C
12	500	N2	N1
12	500	C	N1
12	1,000	C	N2
Total	9,500		

Coding:

C: Current race/ethnicity item, two-question format with Hispanic listed among the races and a separate question about specific Hispanic ethnicity asked after the race question, students asked to select only one race

N1: New race/ethnicity item, one-question format with Hispanic listed among races, students permitted to select multiple races

N2: New race/ethnicity item, two-question format with separate Hispanic-ethnicity question asked before the race question, students permitted to select multiple races

The data generated from this design support a straightforward test of the null hypothesis that the items are independent when administered before and after a substantial distractor task. If a dependence exists we would expect that students offered a single-response version first would be less likely to report multiple races on the subsequent item. Similarly, we might expect respondents receiving the multiple-response version first to violate the single-response restriction on the subsequent, single-response item.

To test these hypotheses we rely on simple *t*-tests of the difference of means for independent groups. These tests compare the percent of respondents identifying multiple race categories when the multiple response question is presented first to the percent of respondents identifying multiple race categories when the multiple response question is presented second.

All comparisons are based on design-consistent standard error estimates. Standard errors were computed using a Taylor-Series estimation procedure which accounted for unequal probabilities of selected and treated schools as primary sampling units.

3. Results

The data collected for this study provided some insights into the different types of response patterns that are likely to be seen with each of the three versions of the

race/ethnicity item that were tested, as well as data that could be used to test our hypothesis regarding the priming effects associated with presenting respondents with two different versions of the race/ethnicity item. Before we discuss our tests of the priming hypothesis, we will discuss some of the response characteristics of the three race/ethnicity items.

3.1. Response characteristics of the three versions of the race/ethnicity item

In this section of the article, we review some of the response characteristics of the current and two new versions of the student race/ethnicity item. Specifically, we look at:

- Item nonresponse rates;
- Single and multiple race response frequencies; and
- Inconsistent responses to the two-question format.

3.1.1. Item nonresponse rates

Item nonresponse rates were higher for the two new versions of the race/ethnicity item than for the current version (Table 2). Looking at eighth and twelfth graders combined, 1.5 percent did not answer either question in the current version, compared with 2.3 percent who did not answer either question in the new two-question version and 2.9 percent who did not answer the new one-question version. If this difference in response rates between the two new versions and the current version of the race/ethnicity item carries through to the first operational assessment that uses one of the new versions of the race/ethnicity question, it may have implications for NAEP student-based race/ethnicity reporting.

Given the sample sizes, the differences in nonresponse rates between the new one-question version and the new two-question version are not large enough to be concerned about, although it should be noted that at each grade level more students are coded as missing on the new one-question version than on the new two-question version, that is, nonresponse rates were slightly higher for the new one-question version.

Table 2. Nonresponse rates by race/ethnicity item and grade level

	New one-question version	New two-question version, both questions skipped	Current version, both questions skipped
8th grade	2.00	1.00	0.88
12th grade	3.69	3.11	1.88
8th and 12th grades combined	2.91	2.27	1.48

We also compared the number of missing responses on the separate race and ethnicity questions on the new two-question version and the current version of the student race/ethnicity item (Table 3). At all grade levels, more students left the ethnicity question blank on the current version of the race/ethnicity item than on the new two-question version. However, more students left the race question blank on the new two-question version of the race/ethnicity item than did on the current item. We suspect that this difference is caused by the change in order of the two questions: in the current version

Table 3. Nonresponse rates for separate race and ethnicity items by grade level

	New two-question version, race question	Current version, race question	New two-question version, ethnicity question	Current version, ethnicity question
8th grade	10.70	0.93	2.20	8.66
12th grade	9.66	1.95	4.27	15.15
8th and 12th grades combined	10.07	1.54	3.44	12.53

the race question is asked first, but in the new two-question version the ethnicity question is asked first. More students are missing on the second question in both versions.

The students who skipped the second question in the two-question versions of the race/ethnicity item are primarily students who had already identified a race or ethnicity. Seventy-five percent of eighth and twelfth graders combined who left the race question blank on the new two-question race/ethnicity item chose an Hispanic ethnicity in response to the ethnicity question. Fewer than one percent of eighth and twelfth graders who left the Hispanic ethnicity question blank in the second question of the current race/ethnicity item chose Hispanic in response to the first question. Nonresponse rates were not affected by whether a race/ethnicity question appeared before or after the assessment.

It appears that once Hispanic students have answered the ethnicity question in the new two-question version of the race/ethnicity item, many of them do not think that they need to answer the second part of the item – the race question – because they have already indicated the group with which they identify.

If the NAEP program chooses to report race in addition to ethnicity for Hispanics, moving the race question to the second position in the race/ethnicity item may have implications for their ability to do so, because of the large number of students who did not answer the race question when it appeared in this position. However, if the field of education continues the current practice of reporting only Hispanic ethnicity for Hispanic students, this should not be a major issue in that field. It could potentially be an issue in other fields where race is reported in addition to ethnicity for Hispanics.

3.1.2. Single and multiple race response frequencies

The data from the NAEP field test support the findings of Tucker and Kojetin 1996 with regard to race frequencies. Specifically, students were more likely to identify themselves as Hispanic when answering a two-question version of the race/ethnicity item with a separate question asking about Hispanic origin than they were when answering a one-question version of the race/ethnicity item that listed Hispanic among the races. When the eighth and twelfth grade samples were combined, 21.7 percent of students who answered the two-question version chose Hispanic, compared with 19.3 percent of students who answered the one-question version.

Students who identified themselves as Hispanic and who answered the two-question version of the new race/ethnicity item were also more likely to choose a race in addition to Hispanic than students who identified themselves as Hispanic and answered the one-question version of the new race/ethnicity item. In the combined sample of eighth and

twelfth grade students, 16.6% of students who answered the one-question version of the new race/ethnicity item chose Hispanic and did not choose a race, compared to only 7.8% of students who answered the two-question version of the race/ethnicity item. Among students who chose Hispanic, 86% of those answering the one-question version did not choose a race in addition to Hispanic, compared to 35.9% of those answering the two-question version.

3.1.3. Inconsistent responses

The new two-question version of the race/ethnicity item had one other interesting feature that affected Hispanic students. The first question allowed students to choose, “No, I am not Hispanic or Latino,” and at the same time choose an Hispanic background. There were .9 percent of eighth grade students and .4 percent of twelfth grade students answering the two-question version of the new race/ethnicity item who chose these two seemingly contradictory statements. It is possible that these students did not understand the term “Hispanic or Latino,” but knew that they or their parents/ancestors were from Mexico, Puerto Rica, or Cuba, and therefore chose one of these ethnic origins. It is also possible that these students were careless or deliberately answering the items inconsistently. For the next set of analyses discussed in this report, we accepted the first explanation for the competing responses and coded the students who chose “No, I am not Hispanic or Latino,” and also chose an Hispanic ethnicity, as Hispanic. Given the small number of cases involved, including these students as Hispanic should not noticeably affect our results.

These inconsistent cases explain about one fourth of the difference discussed above in the percentage of students who identify themselves as Hispanic on the one- and two-question versions of the race/ethnicity item. The remaining difference may be attributable to the fact that having a separate question asking about Hispanic ethnicity could trigger a positive response in students for whom Hispanic is not a primary identity. Some of these students might not choose Hispanic if given a choice of other racial categories, but do respond that they are Hispanic when asked a targeted question about Hispanic ethnicity.

3.2. Priming effects

The data from this study indicate that the responses to different versions of the race/ethnicity items vary depending on whether other versions of the item were asked previously. We present these results in two parts. First we present evidence that the proportion of respondents reporting multiple races differs across conditions. *A priori* we expected that respondents who are first forced to select a single race category (on the current version of the question) might be more inclined to select only a single category on the subsequent new version that solicits multiple responses. Similarly we expected that students who are first offered the opportunity to select multiple categories might be more inclined to “double grid” an item intended to force a single response. The second part of this section takes an exploratory look at the data to help understand which students might be more inclined to report differently depending on the relative position of the old and new items.

Each of the analyses presented in this section is presented for the combined eighth and twelfth grade samples. When looking at distributions across categories of race/ethnicity

we are often interested in relatively rare populations. Detecting shifts in these distributions requires as large a sample size as possible; hence we combined the samples from the two grades.

3.2.1. Effects of the sequence of administration

Tables 4 and 5 compare the percent of respondents reporting multiple races when the new (multiple response) race item is presented before and after the original (single response)

Table 4. *Effects of the sequence of administration: Grades 8 and 12 – New one-question version*

	New item administered before current item: percent	Std. error	New item administered after current item: percent	Std. error	<i>T</i>	<i>p</i>
% choosing more than one race, all Hispanics coded single race	3.2	0.4	2.0	0.4	2.1	0.036
% choosing more than one race, Hispanics who chose one or more races in addition to Hispanic coded multi-race	6.3	0.5	4.5	0.6	2.1	0.033
% choosing more than one race, Hispanics who chose two or more races in addition to Hispanic coded multi-race	3.6	0.4	2.7	0.5	1.3	0.187

Table 5. *Effects of the sequence of administration: Grades 8 and 12 – New two-question version*

	New item administered before current item: percent	Std. error	New item administered after current item: percent	Std. error	<i>T</i>	<i>p</i>
% choosing more than one race, all Hispanics coded single race	2.2	0.3	1.9	0.3	0.5	0.592
% choosing more than one race, Hispanics who chose one or more races in addition to Hispanic coded multi-race	17.6	0.8	12.6	0.8	4.2	0.000
% choosing more than one race, Hispanics who chose two or more races in addition to Hispanic coded multi-race	3.2	0.4	2.5	0.4	1.3	0.211

race item. Table 4 is based on the one-question version of the new race/ethnicity item and Table 5 is based on the two-question version of the new race/ethnicity item. For the purposes of this analysis we treated Hispanics in three different ways. For the analyses presented in the first row of each table Hispanics are coded single race no matter how many races they chose in addition to choosing an Hispanic ethnicity. The first row of the table represents the method of reporting most commonly used for Education surveys. In the second row of each table Hispanic is treated as a race, and Hispanics who chose one or more races in addition to Hispanic are coded as choosing more than one race. In the third row of each table Hispanics are coded as choosing more than one race only if they chose two or more races in addition to Hispanic. The third row of the table represents the method of reporting used in many government agencies other than Education.

For all three coding methods the percentage of students choosing more than one race/ethnicity if the new race/ethnicity item is administered after the current version of the race/ethnicity item is lower than the percentage of students choosing more than one race/ethnicity when the new version of the race/ethnicity item is asked before the current version. If there is an order effect this is the direction in which we would expect to see the effect, i.e., we would expect that after being asked to choose only one race/ethnicity students would be less likely to choose more than one race/ethnicity in response to a second question. In three of these comparisons the difference is statistically significant at the customary confidence level of .95. The fact that all of the differences are in the same, and expected, direction, leads us to believe that the order of administration systematically influences the responses obtained and that with a large enough sample the other comparisons would probably also be significant.

Since students who answered one of the new versions of the race/ethnicity item after answering the current version of the race/ethnicity item appeared somewhat more likely to limit themselves to choosing only one race than students who answered a new version of the race/ethnicity item before answering the current version, we wondered if the reverse was also true. Specifically, we explored whether or not students who answered the current version of the race/ethnicity item, in which they were supposed to choose only one race, were more likely to double grid the current version of the race/ethnicity item, i.e., fill in more than one bubble even though instructed not to do so.

As shown in Table 6, the percentage of students who chose more than one race in response to the current version of the race/ethnicity item is small. However, as we would expect in the case of an order effect, it increases if the current version of the race/ethnicity

Table 6. Percentage of students who double grid: Current version of race/ethnicity item

	New item administered before current item: percent	Std. error	New item administered after current item: percent	Std. error	<i>T</i>	<i>p</i>
8th grade only	0.4	0.20	1.4	0.38	-2.2	0.025
12th grade only	0.2	0.10	0.8	0.27	-2.2	0.027
8th and 12th grade combined	0.3	0.10	1.1	0.23	-3.3	0.001

item is asked after a new version of the race/ethnicity item. The difference in the number of students double gridding before and after the cognitive items meets the common standard of statistical significance ($p < .05$) at both grade levels. The probabilities of statistical significance are slightly lower at the twelfth grade level, but still large enough to suggest that there may be an order effect.

Based on the preceding analyses we can reject the hypothesis that administration of one race item is unaffected by responses to the other, previously administered, version of the item. In particular, when the first item denies respondents the opportunity to identify multiple races, respondents are less likely to identify multiple races on a subsequent question that does offer that opportunity. Also, when the first item offers the opportunity to identify multiple races, responses are more likely to violate a single-response instruction in a subsequent item.

However, the fact that the order effect is small and not statistically significant for the new two-question version when Hispanics are all coded as a single race (line 1 of Table 5) or when Hispanics who choose more than one race in addition to Hispanic are coded as multi-race (line 3 of Table 5), indicates that priming may not currently be of serious concern if these methods of reporting are used. Although the method of reporting represented by line 1 of Table 5 is commonly used in Education, and the method of reporting represented by line 3 of Table 5 is most commonly used in other government agencies, and the results obtained with both these methods were not statistically significant, we are still concerned about the possibility of priming. The small size of this effect may be a result of the small number of non-Hispanic students choosing more than one race. As the number of non-Hispanic multi-race students increases, the small order effects associated with the question and this method of coding may well increase and become statistically significant.

3.2.2. Effects of the sequence of administration by racial and ethnic group: New items
The findings above beg the question: "Which groups of students are most affected by the order in which items are administered?" Examination of the effects of the sequence of racial/ethnic groups provides a preliminary answer to this question. Tables 7 and 8 provide this information broken down by the largest racial/ethnic categories.

Almost none of the shifts among racial and ethnic group that occur when students answer the new item second, rather than first, are statistically significant. However, students answering the current one-race-only version first are more likely to identify themselves as only white when the two-question version of the new race/ethnicity item is subsequently administered (Table 8). Perhaps predictably, this is not the case with Hispanics. Respondents identifying themselves as Hispanic on the current version of the item face a somewhat weaker constraint, since the current version of the item used a two-question format, which offers a convenient opportunity to indicate both Hispanic and some other race/ethnicity.

Of course, cell sizes here prevent a confident interpretation that these differences are not due to sampling error. However, given that we found evidence for an order effect in the entire sample, these findings by race suggest that a larger sample would reveal significant effects of order by racial group as well.

Table 7. Effects of the sequence of administration by racial and ethnic group: Grades 8 and 12 – New one-question version

	New item administered before current item: percent	Std. error	New item administered after current item: percent	Std. error	<i>T</i>	<i>p</i>
White	56.5	1.09	58.1	1.54	-0.9	0.395
Black	13.9	0.76	14.5	1.10	-0.4	0.657
Hispanic	16.8	0.82	16.6	1.16	0.1	0.923
Hispanic/white	2.0	0.31	1.2	0.34	1.8	0.065
Hispanic/nonwhite race	0.7	0.18	0.6	0.24	0.3	0.774
Hispanic/multiple race	0.4	0.14	0.8	0.28	-1.1	0.260
Asian	4.5	0.45	4.2	0.63	0.3	0.736
Nat Hawaiian/Pacific Is	0.6	0.17	1.2	0.34	-1.5	0.147
American Indian	1.5	0.27	1.0	0.31	1.4	0.172
White/Black	0.6	0.17	0.2	0.14	1.8	0.078
White/Asian	0.5	0.16	0.1	0.10	2.3	0.021
White/American Indian	1.1	0.22	1.0	0.31	0.2	0.836
Black/American Indian	0.4	0.14	0.3	0.17	0.4	0.675
Other multiple races	0.6	0.17	0.4	0.20	0.9	0.371

3.2.3. Effects of the sequence of administration by racial and ethnic group: Current item
Earlier in this article we discussed the fact that students who answered the current version of the race/ethnicity item after answering a new version of the race/ethnicity item were somewhat more likely to double grid, or select more than one response category for the current version of the race/ethnicity item, than students who answered the current version of the race/ethnicity item before answering a new version of the item. In this section of the article we turn to an examination of the extent to which there is an order effect that influences student responses to the current item other than double gridding.

In performing this analysis we coded students who chose an Hispanic ethnicity as Hispanic regardless of any race they chose in response to the race question. Table 9 shows the results of this analysis done using the data set that combined the eighth and twelfth grade samples.⁷

Consistent with our findings about the multiple response version, the number of students who chose the most common racial category, “white,” was lower among students who answered the current version of the race/ethnicity item after being allowed to choose multiple races in one of the new versions of the race/ethnicity item, and the number of students who selected “other” was higher. This suggests that, in addition to being more likely to double grid, students who have responded to a question that permitted them to

⁷ Note that the number of students coded as double gridding in Table 9 is lower than the number of students coded as double gridding in Table 6. This is because some students who doubled gridded on the race question in the current version of the race/ethnicity item, chose an Hispanic ethnicity in response to the ethnicity question. Therefore, following current NAEP conventions, we coded them as Hispanic in Table 9. However, in Table 6 they were coded as double gridding.

Table 8. Effects of the sequence of administration by racial and ethnic group: Grades 8 and 12 – New two-question version

	New item administered before current item: percent	Std. error	New item administered after current item: percent	Std. error	<i>T</i>	<i>p</i>
White	55.5	1.09	58.8	1.25	-2.0	0.042
Black	13.6	0.76	14.7	0.90	-1.0	0.326
Hispanic	7.2	0.57	8.3	0.70	-1.2	0.212
Hispanic/white	7.9	0.59	6.0	0.60	2.2	0.028
Hispanic/nonwhite race	6.6	0.54	4.1	0.50	3.3	0.001
Hispanic/multiple race	1.0	0.22	0.6	0.19	1.5	0.133
Asian	4.0	0.43	3.5	0.47	0.7	0.487
Nat Hawaiian/Pacific Is	0.7	0.18	1.2	0.27	-1.5	0.143
American Indian	1.5	0.27	0.8	0.23	1.4	0.059
White/Black	0.3	0.13	0.3	0.14	0.1	0.925
White/Asian	0.3	0.13	0.1	0.09	1.3	0.179
White/American Indian	0.7	0.18	0.4	0.17	0.9	0.356
Black/American Indian	0.2	0.10	0.3	0.14	-0.7	0.462
Other multiple races	0.6	0.17	0.7	0.21	-0.3	0.781

Table 9. Effects of the sequence of administration by racial and ethnic group: Grades 8 and 12 – Current version of race/ethnicity item

	New item administered before current item (percent)	Std. error	New item administered after current item (percent)	Std. error	<i>T</i>	<i>p</i>
White	58.8	0.95	54.9	1.10	2.7	0.0008
Black	14.1	0.67	14.2	0.77	-0.1	0.950
Hispanic	18.9	0.76	20.8	0.90	-1.6	0.110
Asian	3.5	0.36	4.1	0.44	-1.1	0.287
Pacific Islander	1.2	0.21	1.1	0.23	0.5	0.614
American Indian	1.2	0.21	1.4	0.26	-0.4	0.684
Other	1.9	0.27	2.7	0.36	-1.7	0.095
Double grid	0.2	0.09	0.8	0.20	-2.8	0.006

select more than one race may be less likely to select white in a subsequent question, and consequently select “other” or double grid.

4. Implications for Continued Reporting of Trend

In response to OMB’s mandate to offer respondents the opportunity to identify multiple races, federal agencies will have to change the survey items used to collect data. Each survey program will face the challenge of maintaining valid and informative trend information as the race classification variables change. It will be tempting to administer both the old and the new items to the same respondents, effecting a bridge between the two measurement instruments. While the data for this study were collected from a survey of

middle and high school students, and generalization beyond this population remains an open question, the results of the study highlight the fact that the potential risks associated with that strategy depend upon plans for coding the data for reporting purposes. Although the results shown in lines one and three of Table 5 are not statistically significant, the data in Table 5 suggest that with larger sample sizes – particularly as the percentage of people identifying themselves as multi-racial increases – we will probably see differences based on the order in which the questions are administered.

The priming effects associated with asking the same students two different race/ethnicity items are not uniform across all students. In particular, priming will be most acute among Hispanics and those who identify with multiple groups. Therefore, the reporting method chosen may determine the seriousness of the problem associated with the priming effects.

If Hispanics are classified only by their ethnicity, regardless of which racial category or categories they select, the order effect associated with the new two-question version is small and not statistically significant using the common standard of statistical significance ($p < .05$). Therefore, if the convention of classifying Hispanics only by their ethnicity is followed, linkage through a strategy of asking the same sample of students both the old and new two-question version may be possible. This is the reporting strategy that has most commonly been used by Education. However, as the number of non-Hispanic multi-race students increases, the small order effects associated with this question and this method of coding may well increase and become statistically significant.

If Hispanics are classified as a race and ethnicity, as in line two of Table 5, any mapping between the old and the new race/ethnicity items based on a common sample will reflect the dependence among responses. Unless the mapping explicitly captures and models this dependence, inferences generalized to future surveys that include only the new item, or past surveys that included only the older one, will be biased. This bias is likely to grow as the population of the United States becomes increasingly multi-racial.

Above we outlined two general strategies that federal agencies might apply to provide useful trend information as the measurement of race changes:

- Maintaining parallel trend lines, one reporting on the current race/ethnicity measure and one reporting on the new race/ethnicity measure, for a finite number of years; and
- Developing a linkage or mapping between the current and new items, and reporting the “latent” trend associated with the old item based on responses to the new item.

Given the difficulty of using a common sample to gather data using both items, these options become fairly unattractive. If results are reported for Hispanics using both their race and ethnicity, the first approach (maintaining two parallel trend lines) would require selecting a larger sample so that two different forms could be administered in order to maintain the overlapping trend lines for some number of years. If the same level of measurement precision is desired, the sample would need to be twice as large as the current sample, with each form administered to half the sample. This implies a significant and durable increase in data collection costs. If results for Hispanics are reported using

only their ethnicity, the order effects associated with the strategy are likely to be small and may be acceptable depending upon the level of precision required. However, as the non-Hispanic multi-racial population increases, the order effects are likely to grow and separate; overlapping trend lines may be required at that time.

Under the second approach, both versions could be administered to a common sample, and the specific priming effects could be modeled. This strategy would require complex statistical modeling that took into account priming effects, particularly if Hispanics are classified by both their race and their ethnicity. Even if Hispanics are classified only by their ethnicity, as the non-Hispanic multi-race population increases more complex statistical modeling is likely to become necessary.

The second approach has two drawbacks. First the “linking sample” would probably have to be quite large to estimate priming effects that may vary across groups. Second, if the priming effect is relevant (that is, if it is correlated with the survey’s target variables), the target variables themselves would probably have to be included in the linking model. Despite any technical merits of this approach, it would look as if the data were being “adjusted” for race – an appearance that is often controversial. Ironically this controversy arises for the same reasons that tracking key statistics by race is important.

Appendix A

Question Wording for Three Versions of the Race/Ethnicity Item

New One-Question Version of Race/Ethnicity Item

Which of the following describes you? Fill in **one or more ovals**.

- White
- Black or African American
- Hispanic or Latino
- Asian
- American Indian or Alaska Native
- Native Hawaiian or other Pacific Islander

New Two-Question Version of Race/Ethnicity Item

Are you Hispanic or Latino? Fill in **one or more ovals**.

- No, I am not Hispanic or Latino.
- Yes, I am Mexican, Mexican American, or Chicano.
- Yes, I am Puerto Rican or Puerto Rican American.
- Yes, I am Cuban or Cuban American.
- Yes, I am from some other Hispanic or Latino background.

Which of the following describes you? Fill in **one or more ovals**.

- White
- Black or African American
- Asian
- American Indian or Alaska Native
- Native Hawaiian or other Pacific Islander

Current Version of Race/Ethnicity Item

Which best describes you?

- White (not Hispanic)
- Black or African American (not Hispanic)
- Hispanic (“Hispanic” means someone who is from a Mexicano, Mexican American, Chicano, Puerto Rican, Cuban, or other Spanish or Hispanic background.)
- Asian (“Asian” means someone who is from a Chinese, Japanese, Vietnamese, or other Asian background.)
- Pacific Islander (“Pacific Islander” means someone who is from a Filipino, Hawaiian, or other Pacific Island background.)
- American Indian or Alaska Native (“American Indian or Alaskan Native” means someone who is from one of the American Indian tribes, or one of the original people of Alaska.)
- Other

If you are Hispanic, what is your Hispanic background?

- I am not Hispanic
- Mexican, Mexican American, or Chicano
- Puerto Rican
- Cuban
- Other Spanish or Hispanic background

5. References

- Adams, P.F., Hendershot, G.E., and Marano, M.A. (1999). Current Estimates from the National Health Survey, 1996. National Center for Health Statistics. *Vital Health Statistics*, 10(200).
- Current Population Survey (2000). Technical Paper 63: Current Population Survey Design and Methodology. Washington, DC: Current Population Survey.
- McKay, R.B. and de la Puente, M. (1996). Cognitive Testing of Racial and Ethnic Questions for the CPS Supplement. *Monthly Labor Review*, 119(9), 8–12.
- Sudman, S., Bradburn, N.M., and Schwarz, N. (1996). *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, CA: Jossey-Bass, Inc.
- Tucker, C. and Kojetin, B. (1996). Testing Racial and Ethnic Origin Questions in the CPS Supplement. *Monthly Labor Review*, 119(9), 3–7.
- U.S. Bureau of the Census (1998). *Survey of Income and Program Participation User’s Guide*. Washington, DC: U.S. Bureau of the Census.
- U.S. Census Bureau (2001). *The Two or More Races Population: 2000 (Census 2000 Brief)*. Washington, DC: U.S. Census Bureau.
- U.S. Department of Education (1999). *The NAEP 1996 Technical Report, NCES 1999-452*, by Allen, N.L., Carlson, J.E., and Zelenak, C.A. Washington, DC: National Center for Education Statistics.

U.S. Office of Management and Budget (1997). Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. The Federal Register, October 30. Washington, DC: Government Printing Office.

Received October 2000

Revised July 2004