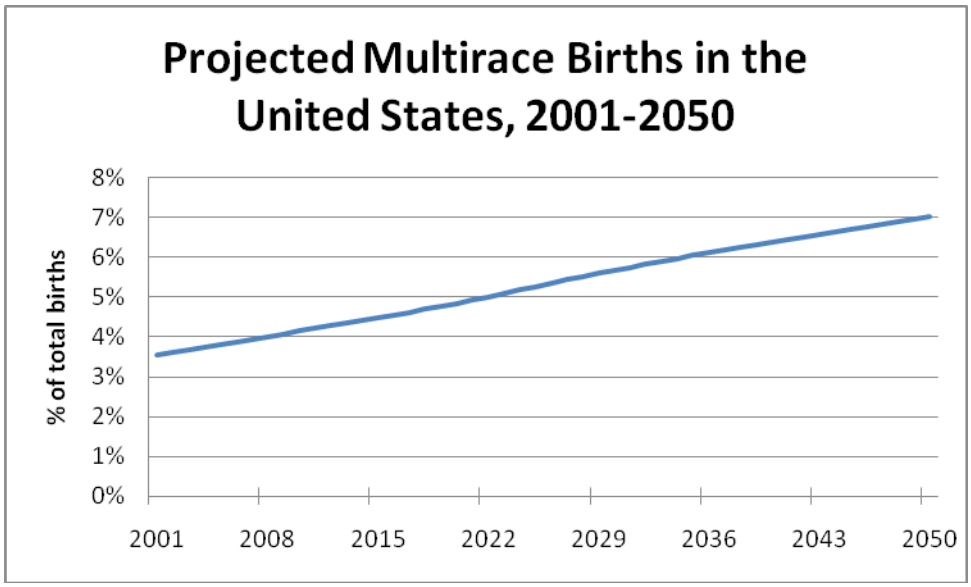
Shifting Demographics: Preparing for a New Race and Ethnicity Classification Scheme in NAEP

Statement of the problem

On September 24, 2007, the U.S. Department of Education (USDE) issued final guidance for collecting and reporting race/ethnicity information. The USDE guidance addresses three issues that will affect the National Assessment of Educational Progress study (NAEP): 1) How educational institutions and other data providers will collect and maintain racial and ethnicity data from students and staff; 2) How educational institutions and other recipients will aggregate racial and ethnicity data when reporting those data to the Department; and 3) How data on multiple races will be reported and aggregated under the Elementary and Secondary Education Act of 1965 (ESEA), as reauthorized by the No Child Left Behind Act of 2001 (NCLB). Final implementation of these guidelines should take place no later than the 2010–2011 school-year. But, what the USDE guidance does not address is how these changes will affect NAEP trend reporting, if at all.

In 2000, the U.S. Bureau of the Census estimated the multi-racial population at about 3.8 million, or 1.4 percent of the total population. By 2004, Census found the number of people who identified as multi-racial had grown by almost 14 percent, to nearly 4.5 million individuals. While these numbers and proportions may not seem significant and perhaps in some part an artifact of using a new classification scheme, the figure below shows the expected growth in the multirace population over the next fifty years[1].



More to the point, however, all jurisdictions under the purview of the Department of Education are being asked to collect and report race and ethnicity information in one standardized way. This one change alone is likely to yield importantly different results. Up until now, jurisdictions were only obligated to report race and ethnicity information in set categories, but were under no obligation to collect this information in any particular manner. Therefore, the combination of

1 Estimates are from the combination of single-race individuals only.

standardizing the method of collection and the allowing of individuals to "mark all that apply" in regards to race, is very likely to shift our understanding of educational performance in the very near future. Suggesting that it is important to understand sooner, rather than later, how the shifting demographics around race and ethnicity classifications are likely to affect our understanding of academic performance in our Nation's Report Card.

Methods and Data

Currently, NAEP collects student race/ethnicity information from two sources: school records and student self-reports via the NAEP student background questionnaire. To be clear, NAEP collects school-recorded race/ethnicity as part of the student sampling process; but at present time, states vary widely on how this information is collected and in the amount of detail collected on multi-race/ethnicity categories.

The student self-reported information, however, is collected using a two-part race/ethnicity question that permits multi-race responses. The structure of this two-part question was specified by OMB in 1997. As of now, the student self-reported variable has been a trend variable since 2003 and thus provides a special window to anticipate how differences in achievement as related to race/ethnicity classification might change or be affected post 2011 academic year. While the student self-reported information has and does differ from the source of race/ethnicity information (i.e., school-reported) these data have nevertheless played an important role in quality control, which for the foreseeable future will continue to be useful as states make the transition.

By using the student-reported race/ethnicity information as proxies of what NAEP can expect to see after the full-implementation of the Department's guidelines, we propose to explore the possible effects of the coming change in the racial/ethnic classification scheme.

Data will come from NAEP's 2003, 2005, and 2007 Reading and Mathematics assessments of Grade 8. We plan not to examine 4th grade data because their racial and ethnic self-identification is likely to be less reliable than their older counterparts and Grade 12 data is much too sparse to be reliable. In addition to using NAEP data there is a possibility of using other data sources to further the investigation such as the American Community Survey (ACS) to complement and corroborate NAEP. Will explore this possibility and make use of it were possible to contextualize our findings.

Identification of multiracial students will be accomplished using data from NAEP's student background questionnaire. Students are allowed to identify themselves with more than one racial/ethnic group. Such information is recorded in NAEP databases via dummy variables that indicate what race(s) and/or ethnicity the student chose. For the purposes of this study students that marked more than one option will be regarded as self-identified multiracial. To be clear, the mutually exclusive racial/ethnic categories we intend to work with are White (non-Hispanic), Black (non-Hispanic), Hispanic, Asian/Pacific Islander, American Indian/Alaska Native, and Unclassified. Table 1 below shows preliminary estimates of grade 8 students that can be identified as being more than one race. In the case of white students for example, about 3.1 to 5.0 percent report being white and some other race; whereas for black students that percent ranges from about 6.2 to about 9.0. For the other groups, however, the estimates do not appear to match as well between NAEP and ACS; further analyses will be carried out to better understand these discrepancies.

Table 1. Percent of grade 8 students identified as being of two or more races in NAEP and the American Community Survey for 2003, 2005, and 2007.

	2003		2005		2007	
	NAEP	ACS	NAEP	ACS	NAEP	ACS
American Indian / Alaska Native	12.5	45.0	15.0	41.7	18.0	50.0
Asian	6.5	19.8	6.5	17.7	8.0	18.6
Black	7.5	6.2	8.0	6.4	9.0	8.4
Hispanic	--	--	--	--	--	--
White	4.0	3.1	4.5	3.2	5.0	4.0
Other	29.0	30.1	26.0	25.6	29.0	27.4

After multiracials are identified and accounted for, all following analyses will be carried out using a combination of Stata and AM (American Institutes for Research and Cohen, 2005) procedures. Stata is a well known statistical package that allows for easy data management with great graphing capabilities. AM is less well known, but it is a statistical software package for analyzing large-scale assessment data from complex samples like NAEP. It allows estimation of target population parameters (e.g. means, variances) using Marginal Maximum Likelihood (MML). Users can utilize plausible values readily available in NAEP data sets with AM. An additional strength of this software is that it allows for direct estimation of population parameters (bypassing plausible values) if item response data and item parameter estimates (IRT-based) are available. Taking into all relevant design variables (strata, sampling unit, sampling weights) we will carry out our analyses either using plausible values or via direct estimation wherever necessary. More specifically, AM's MML Composite Means and MML Composite Regression procedures will be utilized in order to answer most of the questions listed above. Other questions will require simpler procedures such as frequency analysis.

Research Questions

The aim of this project is to investigate how and to what extent the change in racial/ethnic classification schemes might affect NAEP trend reporting, especially in relation to previously established racial/ethnic gaps in achievement. Below are the major questions we intend to explore:

1) What percent of the students identify themselves as multiracial?
2) To what extent is there correspondence between school- and student-reported race/ethnicity information?
3) Does correspondence depend on factors such as gender, socioeconomic status, and school type (i.e., public vs. private)?
4) Is there significant difference in achievement scores between 'corresponders' and 'non-corresponders' and does the difference vary by major race group? Does the difference vary gender, socioeconomic status, school type, or geographic region?
5) What is the net effect of the new classification scheme on the achievement gaps? Once we adjust for 'correspondence', do the gaps change? If so, how?