DOCUMENT RESUME

ED 292 848 TM 011 228

AUTHOR Rivera, Charlene; Pennock-Roman, Maria
TITLE Issues in Race/Ethnicity Identification Procedures in the National Assessment of Educational Progress, Part I: A Comparison of Observer Reports and Self-Identification. Research Report No. 87-19.
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE Apr 87
NOTE 65p.; Appendices contain small print.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Rating; Black Achievement; *Black Students; Elementary Secondary Education; *Ethnicity; Grade 4; Grade 8; Grade 11; Hispanic Americans; *Racial Identification; Secondary School Students; Self Evaluation (Individuals); *White Students
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

In this investigation, the validity of observer reports versus student self-reports of race and ethnicity in the National Assessment of Educational Progress (NAEP) for ages 9, 13, and 17 years were evaluated. Data from seven surveys (1975-1984) were analyzed, with sample sizes ranging from 15,859 to 38,899. While the concordance between the two methods for both White and Black students was very high (95 percent or better), the two classifications gave very disparate results for the other four racial/ethnic groups. For example, observers undercounted self-identified Hispanic 17-year-olds by 25 percent to 48 percent. Language background was significantly more consistent with self-identification than with observer reports for 17-year-old Hispanic students, which showed that self-reports were more valid than observed ethnicity in this age group. However, the results for ages 9 and 13 years were less clear-cut. The implications of the findings for continuity of NAEP data sets and the validity of reported group achievement data for Hispanics are discussed. (Author)

RR-87-19

RESEARCH REPORT

ISSUES IN RACE/ETHNICITY IDENTIFICATION PROCEDURES IN THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS, PART I: A COMPARISON OF OBSERVER REPORTS AND SELF-IDENTIFICATION

Charlene Rivera
Maria Pennock-Roman

Educational Testing Service
Princeton, New Jersey
April 1987

Research Report

Issues in Race/Ethnicity Identification Procedures in the National Assessment of Educational Progress, Part I: A Comparison of Observer Reports and Self-Identification

Charlene Rivera

Development Associates, Arlington, Virginia

Maria Pennock-Roman

Educational Testing Service, Princeton, NJ

April 1987

Acknowledgements

We gratefully acknowledge extensive comments to earlier drafts by Vilma Ortiz and reviews by Albert E. Beaton, Benjamin King, Rebecca Zwick, Phyllis Teitelbaum, and Richard W. Adams. Computer programming was carried out by Bruce Kaplan, Laurie Barnett, Steven Vail, and David Freud at ETS. Roberto Fernandez suggested valuable readings. Marion Epstein gave administrative and moral support throughout the execution of the work. Archie Lapointe, Douglas Rhodes, Nancy Mead, David Wright, Rebecca Zwick, Albert Beaton, Gene Johnson, and other members of the NAEP staff provided administrative and technical assistance and essential background information for this project. Terrie St. Clair, Beverly Cisney, Jo Ann Farr, Mary Evelyn Runyon and Terri Stirling patiently typed drafts and revisions of this report. This research was supported by a NAEP Visiting Scholar Postdoctoral Fellowship awarded to Rivera and postdoctoral fellowships from the Rockefeller Foundation and the National Research Council/Ford Foundation program, awarded to Pennock-Roman. Additional funds were provided by Ernest Anastasio and Victor Bunderson at ETS.

Abstract

In this investigation, the validity of observer reports versus student self-reports of race and ethnicity in the National Assessment of Educational Progress (NAEP) for ages 9, 13, and 17 were evaluated. Data from seven surveys (1975-1984) were analyzed, with sample sizes ranging from 15,859 to 38,899. While the concordance between the two methods for both White and Black students was very high (95% or better), the two classifications gave very disparate results for the other four racial/ethnic groups. For example, observers undercounted self-identified Hispanic 17-year-olds by 25% to 48%. Language background was significantly ($p < .0001$) more consistent with self-identification than with observer reports for 17-year-old Hispanic students which showed that self reports were more valid than observed ethnicity in this age group. However, the results for ages 9 and 13 were less clear-cut. The implications of the findings for continuity in NAEP data sets and the validity of reported group achievement data for Hispanic students are discussed.

Agreement between Race/Ethnicity Identification Methods in the National Assessment of Educational Progress

The procedures used to classify an individual s race and ethnicity in national surveys, population censuses, and vital statistics records have increasingly come under scrutiny during the last two decades. Although all currently used approaches have flaws (Fernandez, 1975; Hernandez, Estrada, & Alvirez, 1973; McKenny, Farley, & Levin, 1983; "Minutes," 1979; Smith, 1980, 1981), self-report methods have been found to be among the most practical and accurate procedures employed with adults and older adolescents. Consequently, there has been a gradual shift from reliance on observational and indirect methods in race/ethnicity identification (e.g., surnames) to self-identification.

This trend is evident in the design of the National Assessment of Educational Progress (NAEP), the survey that generated the data analyzed here. From 1969 to 1982, elementary and secondary school students were categorized with respect to race/ethnicity on the basis of test administrators' observations. In 1983-4, the procedure was changed so that students were classified primarily on the basis of their self-reports rather than the observer's reports, although assessment administrators continued to note race/ethnicity. In this article, we evaluate the effects of this change on the validity and continuity of the NAEP data sets with a special focus on the younger cohorts, Ages 9 and 13, and the difficulties of conducting survey research on school-aged children.

Review of the Literature

The revisions in NAEP procedures parallel earlier changes in methods in the Census and Current Population Surveys (CPS). Prior to the 1960 Census,

the race of respondents was coded by census enumerators based on personal observation, and ethnicity was determined by responses to questions such as surname, birth place, foreign parentage, and language background (P. Johnson, Bureau of the Census, personal communication, August 9, 1985; Lampe, 1984). Because of the problems and limitations associated with these observational and indirect measures of race and ethnicity, demographers and other social scientists began to employ self-report procedures. Self-identified race was first collected for the 1960 Census and self-reported ethnicity emerged in the 1980 Census (Linn, McKenney, & Berman, 1984). National surveys of secondary school students such as the ongoing High School and Beyond and the 1972 National Longitudinal Study (NLS) have also used self-report methods.

The widespread use of self-identification procedures reflects a shift in the social and political conceptualization of race/ethnicity and its measurement. While self-report measures that require selection among ethnic labels are not without their problems (see Buriel, 1984; Garcia, 1982; Hernandez, et al., 1973; Lampe, 1984; "Panel", 1978), they are "considered by many observers [to be] the closest approximation to the sociological concept of ethnic group identity" (Hernandez, et al., 1973, p. 679). Furthermore, they are more successful than other methods in providing an ethnic identifier for a substantial number of respondents (Levin, et al., 1984; Smith, 1981; McKenney, et al., 1983). The prevailing view is that, for adults and older adolescents, ethnicity is best defined by self-report which can be validated by indicators that are considered less subjective, such as surname, birthplace, etc., when these are available. Despite this acceptance, there is still concern about the reliability of self-report data ("Panel on Decennial", 1978) and its effect on the comparability across datasets (Hernandez, et al., 1973; Lampe, 1984; McKenney, et al., 1983; Sullivan, Gillespie, and Rogers,

1984). Lowry (1980) points to the possible bias of self-reports if political pressures are allowed to predominate; thus, he calls for "a general empirical study of ethnic self-identification and its objective correlatives...to resolve classification problems" (p. 6).

Third-party reports and other non-self-report measures also continue to be of interest because it is believed that they can serve to verify subjective methods more objectively as well as to provide classifications in cases where self-reports are not feasible. For birth and death records, for example, family members or a third-party (e.g., physicians) generally identify an individual's ethnicity, although, in the absence of a third party, proxies such as surname and birthplace are sometimes used.

Thus, the research suggested by Lowry to link various procedures would be desirable since no single method suffices for all applications. Although various indicators overlap, they are far from interchangeable. Moreover, the type of measure used can have a large effect on the reported size and characteristics of a group (McKenney, et al., 1983). This type of research with Hispanic populations is of special interest because the choice of procedures has a major impact on Hispanics (Hernandez, et al., 1973), the most rapidly growing segment of the U. S. population. Given the current reliance on self-identification methods in national data sets, it is encouraging to note that self-report methods have proven to be reliable, especially for Hispanic adults. For example, Johnson (1974) found a 64.7% concordance rate in ethnic reports for a matched sample of respondents in the 1971 and 1972 CPS. For Hispanic Americans, which included individuals reporting themselves to be of Puerto Rican, Mexican, Cuban, Central American, South American, or other Spanish origin, there was an 80% to 95% consistency rate. In a later study, Tienda and Ortiz (1985) validated self-report indicators from another

perspective. They analyzed the degree of correspondence among six indicators of Spanish/Hispanic origin or descent available in the 1980 census data set. These measures were: self-reported ethnicity, country of origin of ancestors, Spanish language use, indication of Spanish/Hispanic origin when responding to the race question, birthplace, and surname. They found "Hispanics to be quite consistent in reporting characteristics which flag their Hispanic origin or descent" (p. 17).

Agreement between indicators of Hispanicity in death certificates was found in a study which compared ethnicity as classified by surname with a third party report (Sullivan, Gillespie, Hout, & Greeley, 1983). Surnames of Hispanic origin were identified through lists compiled by the Texas Bureau of Vital Statistics and the U. S. Bureau of the Census. Unlike Census enumerators or NAEP assessment administrators who have virtually no acquaintance with the respondent, the third party in this study was someone (usually a relative or parish priest) who knew the decedents well. The authors found that, although the consensus varied according to the database (state or national), overall third-party reports for men agreed with surnames 92.4% of the time and for women 88.1% of the time. These findings indicate that surname is a better indicator of Hispanic ethnicity for men than for women because the latter generally take a new name at marriage. It should be noted that these rates were higher than those found in a study by Fernandez (1975) where a comparison of classifications by self-reports and surnames were used. Using the 1971 CPS data, he found that self-report and surname data overlapped only by about two-thirds.

While the reliance on third party or surname information to classify the race/ethnicity of decedents and newborns is unavoidable, for school-age students, both third-party observer reports and direct self-reports are

options. This is an important consideration because it is not clear whether a self-identification procedure is valid when the respondents are young children or adolescents, as occurs in the NAEP assessments.

Although research has not been conducted on large samples, several studies have shown that children as young as six or seven possess racial/ethnic awareness (Brand, Ruiz, & Padilla, 1974; LeVine & Ruiz, 1978; Rice, Ruiz & Padilla, 1974; Weiland & Coughlin, 1979). Ethnic constancy, or the awareness that racial/ethnic characteristics are permanent, unlike religion or nationality, has been found to be present at about the age of eight (Aboud, 1984). However, direct self-report in a questionnaire format was not used for these studies. Rather, children were required to select photographs or dolls which they felt to be most like them or to match the likenesses presented in certain criterion statements.

Only a few studies have compared survey-questionnaire self-reports of race/ethnicity with observer reports. Massey (1980) found that observers correctly identified the race of adult respondents self-classified as White and Black 99.6% and 94.1% of the time, respectively. However, the agreement between the two methods was lower for the Asian/Pacific-Islander (67.7%) and Alaskan-Native/American-Indian (30.1%) groups. (A separate Hispanic American category was not used.) In another study of observed and self-reported race/ethnicity with adults, Schneider (1980) obtained very different results from Massey's for the Asian/Pacific Islander and Alaskan-Native/American-Indian groups, 86.3% and 92.3% respectively. It is unclear how he defined "match rates," but we surmise that these percentages have the number observed in a category as the denominator, unlike Massey who used the self-report count in the denominator. For the "Spanish surnamed", "Negro," and non-minority groups, Schneider found match rates of 89.5%, 93.9%, and 94.7%, respectively.

Using a sample of 201 high school seniors in Puerto Rico, Ginorio and Berry (1972) also found a high correspondence between self-reported race (Black/White) and observers' ratings of individuals' race. Unlike Massey's or Schneider's classification, their scale was continuous ranging from "mas blanco" (more White) to "mas negro" (more Black). They found a correlation of .97 between self-reported and observed racial classifications.

Besides the Ginorio and Berry report, we were unable to find other studies that compared observed and questionnaire-reported race or ethnicity in school-aged samples. Published evaluations of the use of a questionnaire format for reporting race/ethnicity with children are apparently nonexistent. The research presented here may help to fill this gap.

Research Questions and Rationale

Specifically, the questions addressed by this study include: How much agreement exists between self-reported and visually identified race/ethnicity for students Age 17 and/or Grade 11 assessed for the NAEP? Does agreement vary by racial or ethnic subgroups? What implications do these findings have for the validity of statistics reported by race/ethnic group in pre-1983 assessments (when observer reports were used for classification)? Is self-reported race/ethnicity as valid for the younger cohorts (Ages 9 and 13, Grade 4 or 8) as it is for 17-year-olds, or are observer reports less flawed for these groups?

Based on the research reviewed above, the perspective taken here is that the validity of self-reported race/ethnicity can be regarded as well established for students Age 17 and/or in Grade 11. Accordingly, observer reports in the past NAEP assessments are evaluated against this standard. For the younger groups, none of the available measures of race/ethnicity can be considered a well-tested standard. Nevertheless, a comparison of the findings

of students of different ages may shed some light on the accuracy of self-reports for younger participants and the relative merits of visual vs. self-identification.

Although all subgroups identified in the NAEP are considered, there is a greater emphasis on the Hispanic group, for which identification is known to have been problematical in many surveys. More information is on hand for the Hispanic subgroup because the language background data can serve as an additional indicator of Hispanicity. It can function only as a partial and less than perfect verifier, because many persons of Hispanic descent, especially those of later generations, do not have Spanish-language background.

Since there are multiple data sets over time in the NAEP, it is possible to replicate analyses and provide confirmation of the findings. This article will extend research on the agreement between observer reports and ethnic self-identification to a greater variety of ages and racial/ethnic subgroups than that previously studied. It is the first formal evaluation of classification procedures in a large survey where young children responded to a multiple-choice written question on race/ethnicity.

Method

Sample Selection

The NAEP project, funded since 1969 by the Office of Education and the National Institute of Education, was designed to measure the educational progress of in-school youth and young adults. A stratified, multi-stage sampling plan is used in the NAEP. In the first stage, the primary sampling units are counties selected by geographical region and community type; at the

second stage, the sampling units are defined by school type; and at the third stage, the sampling units are students within a school (Moore, Chromy, & Rogers, 1974, pp. 16-19; Messick, Beaton & Lord, 1983, p. 22; Beaton, 1986, Chapter 13).

The original architects of the NAEP planned the survey so that analyses would be done and findings reported at the item level. Prior to 1983, sampling was done by age only, to avoid linking any findings to specific grades. Beginning with the 1983-84 assessment (hereafter called 1984), several revisions were made to the sampling procedures. The most notable component is the sampling of students both by age and modal grade. That is, students are sampled within a school if they are aged 9, 13, or 17 or if they are in Grades 4, 8, or 11. In the 1984 assessment, the corresponding age and grade samples overlapped by about two-thirds. For example, of the total students sampled (28,405) in Grade 8 and/or who were age thirteen, 21,850 were in the modal Grade 8, 21,070 were thirteen years old, and these two groups had 14,515 students in common.

The seven data sets for this study were derived from four NAEP assessments for which both self- and visual-identification data of students' race/ethnicity were available. Table 1 provides an overview of all the ages/grades and years included. Henceforth, the data sets will be referred to by the year of the second half of the academic year in which an assessment was conducted (e.g., 1976 for 1975-76).

Insert Table 1 about here.

In years where more than one content area was assessed, the data sets for the same age and/or grade were aggregated across content areas for purposes of this study. An exception was made for 1984, where students assessed in the area of writing were excluded because the data were not ready to be analyzed; thus, the results reported here for 1984 are based only on the sample of students who had reading achievement scores as of May 1985.

Observer Reports and Race/Ethnicity Questionnaire Items

Traditionally, racial/ethnic information has been collected for the NAEP using observer-report. The observed or visual procedure consists in having an assessment administrator note a student's racial/ethnic background. In cases of uncertainty, the administrators are advised to note the student's name and/or to listen to him or her speak (NAEP, 1981, p. 18).

Self-reported racial/ethnic data was first collected for 17-year-olds in 1976 using the race/ethnicity questions from the 1972 National Longitudinal Survey. Similar questions were asked of 17-year-olds in all subsequent assessments. In 1982, the collection of self-classification data was extended to 13-year-olds. The 1984 assessment was the first one in which students of all ages and grades self-reported race/ethnicity.

Although observed and self-reported race/ethnicity data were available for some age groups from 1969 to 1982, only the visual-identification data were used solely to classify students and to report student achievement; thus, any self-reported race/ethnicity information collected was not utilized. In 1984, however, the decision was made to rely primarily on the self-report rather than the visual-identification measure to classify students. Observer reports and language background were used to fill in missing data (see Beaton, 1986, Table 12(3)). With the exception of 1976, students were classified into six

racial/ethnic categories: White (W), Black (B), Hispanic (H), American Indian (I), Asian American (A), and Other (O). In 1976, the observers did not have the option to use the American Indian and Asian American categories; students observed to be members of these groups were classified as Other.

Over the years, a variety of questions were used to obtain racial/ethnic information. In 1976, one question was used to elicit self-reported race/ethnicity from students--"How do you describe yourself"? The potential responses were: "American Indian or Native American"; "Black or Afro-American or Negro"; Mexican-American or Chicano"; "Puerto Rican"; "Other Latin-American origin"; "Oriental or Asian-American"; "White or Caucasian", and "Other".

Race and ethnicity were distinguished by two questions in 1980. The first was, "What is your racial background?" to which the answer choices were: "American Indian or Alaskan Native"; "Asian or Pacific Islander"; "Black"; "White"; "Other (Please specify)". The second question asked was, "Is your ethnic heritage Hispanic (such as Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin)?" The response options for the question were "yes" or "no".

For the subsequent assessment in 1982, the race question was the same as in 1980. The ethnicity question was changed to: "Is your ethnic heritage Hispanic?" for which the possible responses were: "No (not Hispanic)"; "Yes (Mexican, Mexican-American, Chicano)"; "Yes (Puerto Rican)"; "Yes (Cuban)"; "Yes, other Spanish/Hispanic". In 1984, students responded to the following two questions. The first, which related to race, was--"Are you": "American Indian or Alaskan Native"; "Asian or Pacific Islander"; "Black"; "White"; "Other (What?)". The second, which focused on eliciting information related to Hispanic ethnicity, was: "Are you Hispanic?" for which the responses were "No" or "Yes, Mexican, Mexican American, or Chicano"; "Yes, Puerto Rican"; "Yes, Cuban"; "Yes, other Spanish/Hispanic (What?)".

Although the changes in questions over the years are subtle, they could have differential effects on the composition of the self-identified groups. However, the questions and responses are sufficiently comparable to be considered partial replications for this study.

Language Questionnaire Items

Two questions were used to assess language use in the home in the 1976, 1980, and 1982 assessments. They were: "Is English the language most often spoken in your home"? for which response options were: "Yes" or "No" and "Is a language other than English spoken in your home?" to which students could respond "Often", "Sometimes", or "Never". In 1984, the questions were changed to: "What language do you speak most often in your home?" for which the answer choices were "English", "Spanish", "Another language (What is it?)"; and, "What language do most people in your home speak?" with the potential responses being "English", "Spanish", "Another language (What is it?)".

For the language analysis, the decision was made to use only the 1984 data because they provide specific information about Spanish language use. Using the 1984 language question on the language most often spoken by others in the home, students were classified into two categories, Spanish background and non-Spanish background, according to the language reported.

Research Questions

Analyses were conducted to answer three groups of issues. (1) In order to determine the degree of correspondence or concordance rates for each racial/ethnic group and across data sets, the following questions were addressed: How much agreement is there between self and visual-identification for students who self-identified as White, Black, Hispanic, American Indian, Asian, or Other? When there is disagreement between observed and self-reported race/ethnicity, what are the more common discrepancies?

Are there some groups for which there is more agreement than for others? How much do concordance rates for each group vary across different assessment years? Are there differences in concordance rates among 9-year-olds, 13-year-olds, and 17-year-olds? (2) In order to assess the consistency among indicators of Hispanicity, we considered the following questions: What is the relationship between the use of Spanish in the home, visual racial/ethnic identification, and racial/ethnic self-identification? And, (3) In order to evaluate the effectiveness of self-report for younger students, we examined: Is the relationship between Spanish use and self-identification the same or different for 9-year-olds and 13-year-olds as compared to 17-year-olds?

Analyses

Using both the observed and self-report data for each age and grade group in each assessment year considered for the study, contingency tables were generated that crossed the observed by self-reported racial/ethnic categories. The percent agreement or concordance rate between visual and self-identification was calculated for each racial/ethnic group by dividing the number of cases that were both visually and self-identified in that category by the number of persons self-identified in the category.

We defined concordance rates in this way and rejected the idea of using a measure of inter-rater agreement such as Cohen's kappa to define concordance because we are taking the perspective that, for 7-year-olds, self-reported race/ethnicity is the "true" standard with relatively little measurement error (akin to a judgment, for example, that a cause of death was suicide and not an accident or murder.) Massey (1980) defined concordance in the same way. This view of self-report has wide support among sociologists and demographers as discussed in the introduction. We consider observer reports to be potentially far more fallible for 17-year-olds and thus evaluate them by how well they

reproduce the self-report standard. Although this perspective is less tenable for students aged 9 and 13, we defined concordance in the same way for these ages in order to be able to compare the identical index across ages.

The disparity among the sample sizes of the race/ethnic groups is another reason for examining concordance rates for each group individually. Because the White group dwarfs all other groups in size, an overall index would have too little sensitivity to the accuracy with which observers classified members of the smaller groups. The bias introduced by errors of classification in a reported mean can be greater for non-White than Whites because the latter group is so much larger. For example, the exclusion of 1000 cases who are Hispanic students from the Hispanic category would have a much larger impact on the Hispanic mean than if the same number of cases were overlooked in the White category because this number represents a larger portion of the total Hispanic count. Therefore, the focus is on the accuracy of classification for each of the smaller groups because of its effect on the reported means for these groups. Furthermore, defining concordance rates in this way yields proportions for individual race/ethnic groups in each year that are essentially statistically independent (except for cluster sampling effects). These proportions can then be compared to each other with t-tests to see if some categories are more accurately identified by observers than other categories.

For all the calculations, data were weighted by corresponding NAEP sample weights (see NAEP Technical Report). Comparisons were made among concordance rates for racial/ethnic groups, and for different ages, grades, and assessment years.

The statistical significance of differences in concordance rates was evaluated using the t-test for two proportions, taking into account the nonrandom sampling design in the estimation of standard errors. In surveys with complex sampling designs such as this one, the estimates of standard errors will be too small if one computes them using classical formulas that assume simple random sampling. (Beaton, 1986, Chapter 13, 14). The actual sampling variance of a statistic in complex designs is larger because the clustering produces correlations among observations. The ratio of the empirically estimated sampling variance to the sampling variance calculated from classical formulas is called a design effect.

There are two common ways to adjust statistical tests to take into account departures from simple random sampling in the estimation of standard errors. One is to scale down the sample sizes by the average design effect (Kish, 1965), (which makes estimates of the standard error larger); another is to estimate individual standard errors for each contrast empirically using jackknifing (Tukey & Mosteller, 1977). The first method can be considered an approximation, since it is based on what prior studies have found is the average design effect for a given measure in a particular data set. The second method, which is more laborious, can be considered to be more exact because jackknifed standard errors are derived empirically for the specific variables in one's study. Of course, these jackknifed estimates are still random variables subject to sampling error. The specific type of jackknifing used in the NAEP is described in the Technical Report (Beaton, 1986, Chapter 13).

Survey statisticians have not developed clear guideline yet as to which method is better in any given situation. In this study, both estimating methods were used as described below.

To answer the first group of research questions, two types of significance testing were conducted. First, to determine how much agreement there was between self and visual identification by ethnic group and age, contrasts in concordance rates were made among samples of the same ethnic group from different data sets and age cohorts. Twenty-one contrasts were calculated within each group among the independent samples of various ages and assessment years. For the White, Black, and Hispanic groups there were seven data sets and 21 contrasts. For Asian Americans, American Indians, and the Other group, only fifteen contrasts were calculated because visual racial/ethnic identification information was not collected for the 1976 assessment. To keep the overall Type I error rate within each family of contrasts to approximately .05, individual pairwise comparisons were considered significant only if $p < .0024$.

Second, to see if there were some groups for which there was more agreement than others, concordance rates for different ethnic groups were compared within the same data set. Fifteen pairwise comparisons were made among the six groups within a given age in a particular year. For 1976, contrasts were limited to only the White, Black, and Hispanic categories because the other three racial/ethnic classifications were collapsed and thus were not comparable with other assessment years. The per-comparison significance level was set to .0033 to keep the overall Type I error rate to approximately .05 for each family of 15 contrasts within a given year.

For the above analyses, standard errors calculated from the usual formulas that assume random sampling were adjusted using the design effect approach. While NAEP statisticians have generally used a composite design effect of two, for this study the results were evaluating using two values, two and four. This decision was based on the large fluctuations in design effects for Hispanic groups and the results of a study by Folsom (1977), where it was found that two was the average but four was usually the largest value obtained. For the 1969 and 1972 data sets, 94%, and 100% of design effects, respectively were four or smaller. The 1983-84 design effects for reading reported in the NAEP Technical Report (Beaton, 1986, Table 14.2) rarely exceeded values of 2.5 although a few exceeded a value of six. In the present study, we found that the choice of design effect, two or four, made little difference in the findings. The significance levels reported are the more conservative ones, using a design effect of four. The position taken was that overestimating the sampling variability was a lesser error than underestimating it.

To answer the second and third group of research questions on the relationship between home use of Spanish and Hispanicity, the percent of students reporting Spanish-language use in their home in Hispanic subgroups of visual by self-classification were compared within a given age cohort. Also, the numbers of Spanish-language background students identified by self vs. visual identification were contrasted.

To test the statistical significance of these differences, estimates of standard errors of the differences in proportions and frequencies were computed through jackknifed sampling because design effects fluctuated greatly in these samples and did not appear to converge on any one number. In the analyses of concordance rates, approximate standard errors were adequate

because the large sample sizes made the results of statistical tests robust; thus, if a test was significant assuming a design effect of two, it was usually also significant assuming a design effect of four. However, for the language analyses among Hispanic subgroups, a few decimal places in the size of the design effect could affect the outcome for some groups because of the small sample sizes. As before, the t-test for proportions was used to make the contrasts, setting the overall type I error to .05.

Results

Correspondence between visual and self-identification

The cross-classification of self-reported by observed race/ethnicity for each age and grade group is displayed in Table 2. The range of percentages shown for each cell in Table 2 indicates the lowest and highest percent of cases falling into that cell across different assessment years. These ranges are based on values for four assessments for the 17-year-olds--1976, 1980, 1982, 1984; two assessments for the 13-year-olds--1982, 1984; and one assessment for the 9 year-olds--1984. For more detail on the specific values for all assessment years, see Table 3 which shows observed race/ethnicity of self-identified Hispanic students for each data set. Tables A.1 to A.5 in the Appendix contain the analogous information for the other groups.

Insert Tables 2 and 3 about here.

It should be noted that the White and Black samples were so large that very small differences in concordance rates were found to be significantly different. For example, when comparing the samples of various years and ages in the White group, trivial differences of about 1% were statistically

significant at the .0024 (two-tailed) level. Also, as expected, the results for grade-level data in 1984 were very close to those for the corresponding ages, owing to the substantial overlap in the age and grade samples. To avoid redundancy, only age results will be discussed for each ethnic group.

White Group. The group self-identifying as White constituted about 71% of the sample, on the average, in any given year. For these students, observed race/ethnicity across age and grade were nearly perfectly concordant with self-report, coinciding 96.4% to 99.4% of the time. The lowest value occurred for children Age 9 in 1984 and the highest were found for 17-year-olds in all years and 13-year-olds in 1982.

The value for 9-year-olds was significantly lower than those of the older groups in all years, although the differences were less than 3%. In three out of four contrasts, the values for 17-year-olds were higher than those for 13-year-olds by about 1%; this exceeded chance level. The two estimates for 13-year-olds differed significantly, but the estimates of 17-year-olds over the four years varied at no more than chance level. Hence, there is a very slight but significant trend toward lower concordance rates in the younger groups who self-identified as White.

When these White students were classified into a non-White category by observers, the most common classification was Hispanic (0.3% for 17-year-olds in 1976 and 1982 to 2.3% for Age 9 students in 1984). The next most likely non-White category was Black (0.1% for 17-year-olds in 1980 to 0.4% for Age 9 students in 1984).

Black Group. For students who self-identified as Black, representing about 13% of the samples, the concordance rates were also very high. They ranged from a low of 94.8% for 17-year-olds and 13-year-olds in 1984 to a high of 97.8% for 13-year-olds in 1982. Unlike the rates for the White Group, the

concordance rates for the youngest Black groups were sometimes higher than for the older group; very few contrasts across the ages and years were actually significantly different.

When observers classified self-reported Black students into a non-Black category, the most likely classification was White (from 1.9% for Age 13 in 1982 to 4.5% for Age 17 in 1984). The next most frequent non-Black category was Hispanic (from 0.1% to 1.5% for Age 17 in 1982 and 1980, respectively).

Hispanic Group. For the group self-identified as Hispanic, which constituted, on the average, about 9% of the samples, the concordance rates were quite variable and considerably lower than they were for self-reported Blacks and Whites, ranging from 45.8% for 9-year-olds to 74.5% for 17-year-olds in 1984. The large fluctuations in concordance rates are shown in more detail in Table 3.

As was found for the White Group, the younger groups tended to have the lowest concordance rates. However, in one year, 1980, the rate for 17-year-olds was so low--52.2%--that it was not significantly different from that of 9-year-olds or 13-year-olds in 1984; also, it was actually significantly lower than that of 13-year-olds in 1982. Two thirds of the contrasts among concordance rates for 17-year-olds for various assessment years exceeded 10% and were highly significant ($p < .0001$, two-tailed).

When observers classified self-reported Hispanic students into a non-Hispanic category, the most common was White (from 18.8% for 17-year-olds in 1984 to 41.5% for 9-year-olds in 1984). The next most frequent classification was Black, ranging from 2.6% for 17-year-olds in 1976 to 14.1% for 13-year-olds in 1982. About 2% of the cases were classified as Asian--from 1.2% for 17-year-olds in 1980 to 2.6% for 17-year-olds in 1982.

American Indians. Students who self-identified as American Indian represented, on the average, less than 2% of the sample. As can be seen in Table 2, for this group there was extremely low concordance with observed race/ethnicity. It ranged from 7.0% for Age 9 in 1984 to 29.3% for Age 17 in 1982. The lowest concordance rates were found for Age 9 (7.0%) in 1984, although this value was not significantly different from that of three out of four of the 17-year-old samples. There were only two out of the 15 contrasts across years and ages in which the difference exceeded chance; concordance rates were remarkably consistent in having low values across all ages and years.

Most of the self-identified American Indian students were classified as White by observers (from 59.9% for 17-year-olds in 1982 to 91.4% for 17-year-olds in 1976). The next most frequent non-Indian classifications were Black (ranging from 2.5% for Age 17 in 1976 to 10.5% for Age 13, 1982) and Hispanic (ranging from 1.6% for Age 17 in 1982 to 10% for Age 9 in 1984). The percent classified as Asian by observers varied from 0.2% for Age 17 in 1980 to 2.4% for Age 9 in 1984.

Asian Group. The concordance rates for self-identified Asians, who constituted, on the average, less than 2% of the samples, were most similar to those of the Hispanic group. They ranged from 64.7% for Age 17 in 1980 to 85.9% for Age 17 in 1982. The concordance rates for the younger groups were all above 70%; they were slightly higher than, but not significantly different from those for 17-year-olds in all years except 1982. Although there were large fluctuations in concordance rates for different years, only three out of fifteen contrasts across the various ages and years within Asians were statistically significant ($p < .0033$) because the sample sizes were relatively small.

Among the observed categories that were discrepant from the self-report, again the most frequent one was White (ranging from 9.8% for Age 17 in 1983 to 21.7% for Age 17 in 1979). The next most frequent non-Asian category was Hispanic, ranging from 1.3% to 11.2% for Age 17 in 1982 and 1984 respectively. Other. Among those students who self-identified as Other, usually less than 3% of the samples, the large majority were classified as White. The rate was from 62.4% for Age 9 in 1984 to 82.4% for Age 17 in 1976. Only 0.8% for Age 9 in 1984 to 6.7% for Age 17 in 1982 were actually classified by observers as Other. (A higher value of 9.9% for 1976 reflected the fact that the American Indian, Asian American, and Other classifications were collapsed into one group). Excluding the 1976 assessment from the contrasts, the concordance rates across ages and years for this category were consistently low; only one out of 15 contrasts exceeded chance level.

Comparison across ethnic groups. The concordance rates are summarized in Table 4 where the self-identified racial/ethnic groups are ordered according to the size of the percent agreement grouped in intervals of five percentage points. The results for self-identified Hispanics of different ages and assessment years are ranked on the left and those for other groups on the right. Subgroup categories are abbreviated by an initial--W, B, H, I, and A. This arrangement shows that all but two groups of the same self-reported race/ethnicity tended to cluster around a value characteristic for that group. The exceptions were the Asian and Hispanic groups which had large fluctuations.

Insert Table 4 About Here

The highest percentages of agreement, which clustered in the mid to high 90's, were found for Whites and Blacks. The concordance rates for these categories were consistently high for all three ages. The percent-agreement figures for Black students were significantly lower ($p < .0033$, two-tailed) than for White students of corresponding ages and years, except for Grade 4 or 9-year-olds in 1984. Though statistically significant, these differences of less than 2% were trivially small. The poorest concordance rates (below 30%) were found for cases classified as Other or American Indian; the percent agreement was consistently low for all ages and assessment years. Needless to say, there was a highly significant difference ($p < .0001$) between these values and corresponding ones for Black, Hispanic, and Asian groups within the same year or age. In three of six contrasts, the rate for American Indians was significantly higher ($p < .0033$) than for the Other category.

For Asian, American, and Hispanic students, the percent concordance was generally above 50%, but still considerably lower than for Black and White students. In every year and age, the concordance rates for Asian and Hispanic students were significantly lower than for corresponding White and Black students ($p < .0001$). However, in three out of six contrasts between Asian, American, and Hispanic students, there were no significant differences. When the difference exceeded chance level, the value for Asian students was higher than for Hispanic students.

For two groups--White and Hispanic--the percent agreement tended to be slightly, but significantly, lower for the younger groups. However, the concordance rates for the various groups of 9-year-olds and 13-year-olds were, generally within $\pm$ 5% of the smallest value found for the corresponding groups of 17-year-olds in the various years. Unfortunately, there is only one year in which both visual and self-identification information exists for

9-year-olds, and, thus, we cannot observe the variability of the results from sample to sample for this age.

Since the Hispanic group constitute the largest group for which visual and self-identification are substantially discrepant, a more detailed look at this data is presented in Table 5.

This table summarizes the results for the Hispanic Group from a different perspective.

Insert Table 5 about here

It illustrates the self-identified breakdown among those who were visually identified as Hispanic. Among this group, from 68.7% for Age 9 in 1984 to 93.6% for Age 17 in 1982, did, in fact, self-identify as Hispanic. The largest source of disagreement was the inclusion in the Hispanic category of some students who self-identified as White. This percentage ranged from 3.2% for Age 17 in 1982 to 21.1% for Age 9 in 1984. From this perspective, it can be seen that visual identification had relatively few "false positives"; at least two-thirds of those visually identified as Hispanic self-identified as such. It should be noted that when observed race/ethnicity was used, fewer students were classified as Hispanic (about 9% of the total sample by self-report and 7% by visual identification).

Use of Spanish in the Home

To investigate which method of identifying Hispanicity -- visual vs. self-identification -- had the strongest relationship to Spanish-language use, we take two approaches. We ask (1) what percent of students classified by each method report that most people in their home speak Spanish and (2) among those who report predominant use of Spanish in the home, what percent are

observed or self-identified as Hispanic? Answers to the first and second question can be seen in Tables 6a-c and Table 7, respectively.

Insert Tables 6a-c and 7 about here

In answer to the first question, we see from Tables 6a-c that for all three age groups, the groups classified as Hispanic by both methods (H/H) had the largest frequencies (566, 487, and 521 for 17-, 13-, and 9-year olds, respectively) of students who reported predominant Spanish use in the home. These frequencies represent 51%, 50%, and 60% of the students within the H/H group for the three ages, respectively. From the perspective of the second question shown in Table 7, these frequences represent 85%, 74%, and 58% of the students who report predominant Spanish use in the home for each age group, respectively. That is, the majority of the H/H group report predominant Spanish use in the home, and most of the students from homes in which Spanish is spoken predominantly are classified as Hispanic by both methods.

However, to identify which classification is more consistent with language background, it is necessary to examine the cases in which the two classifications were discrepant. Because some of the groups for which this occurs were very small in number, several categories were collapsed. Respondents who self-identified as Hispanic but were observed to be White, Black, American Indian, Asian, or Other were combined into one group--self-identified Hispanic but visually identified as not Hispanic (H/NH). Those who were visually identified as Hispanic but self-identified as something else were analogously collapsed into the category NH/H.

Looking at the issue from the perspective of what percent of the respondents reported predominant Spanish use in the home within each discordant category, we found results that were opposite for 9- and 13-year-olds to those of 17-year-olds. In Tables 6a-c, it can be seen that 17-year-old, self-identified members of the Hispanic Group who were not visually identified as such (i.e., H/NH) reported predominant Spanish-language use in the home much more frequently than visually-identified members of the Hispanic Group who were not self-identified as such, (i.e., NH/H)--19% $\pm$ 2% (with a standard error of 2%) vs. 7% $\pm$ 5%--and this difference in percentages was highly significant ($p < .0001$, two-tailed). In contrast, 9-year-old, self-reported Hispanics (H/NH) indicated a far lower rate of Spanish language use than visually identified Hispanics (NH/H)(13% $\pm$ 2% vs. 34% $\pm$ 6%). A similar pattern was also found for 13-year-olds (14% $\pm$ 2% vs. 25% $\pm$ 8%). These differences were also significant (respectively, $p < .0001$ and $p < .01$, two-tailed) but in the opposite direction from the findings for 17-year-olds.

From the other perspective, we can ask what percent or frequency of all students from predominantly Spanish-language homes occur in each discordant self/visual category. Among 17-year-olds, we can see from the figures in Table 7 that proportionately more of the students from Spanish-language homes occur in the H/NH group than in the NH/H group (11% vs. less than 2%).

When the frequencies frequencies were contrasted using jackknifed standard errors, the number of homes where Spanish was spoken among respondents in the H/NH was significantly higher ($p < .001$, two-tailed) than the number in the NH/H group. For the 13-year-olds, the percentage of Spanish-language-home students included in the H/NH group was also higher than in the NH/H group (11% vs. 8%), but the frequencies in the two categories were not significantly different. For the 9-year-olds, the breakdown among students with

Spanish-language homes into the H/NH vs. NH/H groups was almost exactly equal(14.8% vs. 15.2%), and the frequencies in the two categories were not significantly different.

We can also examine the results at the marginals, i.e., the overall group H/* self-classified as Hispanic (regardless of visual identification), and the overall group */H visually identified as Hispanic (regardless of self-report). After adding the frequencies for the H/H and H/NH groups, there were 639 17-year-olds in the H/* group from Spanish-speaking homes; as shown in Table 6a, this frequency represented 43% of those self-identified as Hispanic and 96% (see Table 7) of all students from Spanish language homes. The group visually identified as Hispanic */H = (H/H + NH/H) comprised 576 students from Spanish-speaking homes. This frequency represented 46% (Table 6a) of the group visually identified as Hispanic and 87% (Table 7) of all students from Spanish-speaking homes. That is, 9% more of the students who came from homes in which Spanish was spoken predominantly were identified by the self-report procedure than by the observer reports.

For the 13-year-olds, in the H/* group there were 559 who self-reported they were Hispanic and came from Spanish-speaking homes. This frequency represents 38% (see Table 6b) of all those self-reported to be Hispanic and 85% (see Table 7) of all the students from Spanish-speaking homes. In comparison, among those visually identified as Hispanic, there were 541 from Spanish-speaking homes, a frequency that represents 45% (see Table 6b) of respondents visually classified as Hispanic and 83% (see Table 7) of all students from Spanish speaking homes. For the 9-year-olds self-reported to be Hispanic, there were 654 with Spanish-language background. This frequency represents 35% (see Table 6c) of those self-reported to be Hispanic and 73% (see Table 7) of all students with Spanish-language background. Among those visually identified as Hispanic, the corresponding frequencies and percentages were 658, 52%, (Table 6c) and 73% (Table 7).

Thus, for the 17-year-olds, the results show that the self-report procedure identified significantly more of the students with Spanish-language background than did the visual procedure, although the Spanish-language-home students represented less than half of the group self-reported to be Hispanic. A slightly greater proportion of the visually identified group had a predominantly Spanish-language background, because the self-reported group was larger, and there were many students self-reported to be Hispanic whose relatives spoke English most of the time. When the visual and self-classifications were discrepant, the self-report procedure for 17-year-olds gave results more consistent with language background. For the two younger groups, both procedures identified about the same number of students with Spanish-language homes, unlike the 17-year-olds. However, like the 17-year-olds, the students from Spanish-speaking homes represented a smaller proportion of the overall groups self-reported to be Hispanic than the overall groups visually identified as Hispanic because the visually identified groups were smaller. For these younger cohorts, the results differed depending on how one asks the question. On the one hand, the frequency of Spanish-language homes were about the same for the two procedures; on the other hand, these frequencies in relation to the total number in the category were proportionately larger within the groups visually identified as Hispanic than among the self-reported Hispanics Group, because there were fewer students observed to be Hispanic.

Discussion

As indicated earlier, past research has corroborated the validity of self-reported race ethnicity for persons over 14 years of age for major subgroups; therefore, with some confidence, we can consider self-identification by 17-year-olds and students in Grade 11 as the "truth,"

and evaluate observer reports against this standard. The present results on language support this position for the Hispanic group because significantly more of the students who reported predominant Spanish use in the home were self-identified rather than observed to be Hispanic when the two categories were discrepant.

For this age and grade group, the findings indicate that observer reports are highly accurate only for classifications of Black and White students. For these two groups, observer reports and self-reports agree more than 95%. The high percentages confirm Ginorio and Berry's (1972) and Massey's (1980) results on observer reports for differentiations along the Black/White continuum. However, as in the Massey (1980) study, concordance rates between self- and visual-identification for the other groups are low enough to raise questions about the validity of the visual procedure. Concordance rates below 30% were found for American Indians and Others. The Asian and Hispanic rates were somewhat higher but quite variable. Since the Asian group was quite small, the fluctuation in concordance rates may have been the result of sampling error. However, most of the variations in these rates for the Hispanic group were statistically significant.

The implications of these findings for the interpretation of achievement data on the Hispanic group are more serious than for the other groups because the Hispanic subsample is more numerous and achievement averages are reported separately for this group in NAEP publications. Individual group means are not reported for American Indians, Asian-Americans, and Others, since they each represent on the average less than 2% of the sample.

In general, observers fail to classify as Hispanic about 25% to 40% of 17-year-olds who self-identify as Hispanic, leading to an undercounting of the Hispanic group by roughly 1% to 6% in the total sample. Members of the

Hispanic group who are missed by observers are usually classified as White. The fluctuations in the degree of agreement between self- and visual-identification from year to year tend to be quite large, and seem unrelated to changes in the race/ethnicity questionnaire items. For example, the most substantive change in the race/ethnicity question occurred between 1976 and 1980, when a second ethnicity item was added. Yet, the concordance rates were not significantly different for these two years. These results suggest that the variability among data sets for 17-year-olds is in the observer procedures, not the self-reports. Since the Hispanic group is racially heterogenous, the observers' accuracy may depend more on their ability to recognize Spanish surnames and detect accented English, rather than a student's appearance. Differences in this training may lead to inconsistent classifications.

Therefore, there is a real possibility that the mean reading achievement reported for the Hispanic group in pre-1983 assessments (where only observer reports were used for grouping students) may be biased by classification errors, particularly in 1980 when the concordance was found to be only 52%.

These results also cast serious doubt on the continuity of the NAEP data sets for the Hispanic group over time; the large fluctuations in concordance rates from 1976-1986 may indicate that the observer procedures were not consistent from year to year. The present results suggest that the longitudinal trend analyses should be recomputed for at least 17-year-old Hispanics for whom self-reported race/ethnicity is available. The effects on reading achievement are addressed in a separate report by Pennock-Roman and Rivera, which is forthcoming.

To evaluate the adequacy of the two classification procedures at Ages 9 and 13, it would be desirable to have a direct validity check such as parental reports, but these are not feasible for the NAEP surveys, unfortunately. Instead, one needs to take into consideration a variety of factors, such as the literature on young children's awareness of their own race/ethnicity; the amount of overlap between language background and each classification procedure for the Hispanic group; differences in concordance rates between the two methods at various ages; the agreement between the proportion classified as Hispanic by each method in the sample and demographic figures for the population; and the desirability of maintaining continuity in procedures with other national surveys of educational achievement.

Research on race/ethnicity identification in young children has clearly shown that most children are aware of their own race/ethnicity before the age of nine, but some doubt remains as to how accurately they can reply to a written survey question. One can expect more self-report errors compared to older students, possibly because of lower reading proficiency, less experience with multiple-choice forms, and less familiarity with terms such as "Hispanic" or "Asian" rather than the names of individual subgroups, e.g., Cuban or Japanese. The extent to which self-classifications are affected by format have not been studied, although a variety of non-verbal procedures with dolls, photographs, pictures, and other methods have worked well. Despite these gaps in research on questionnaire assessment of race/ethnicity in children, on the whole, past research suggests that self-reported race/ethnicity in the younger ages sampled by the NAEP may be acceptable because the students do have sufficient awareness of their race/ethnicity.

In weighing the evidence from the language background, there are several major limitations to keep in mind. One is that observers sometimes rely on

the presence or absence of a foreign accent in making their classifications because they may ask students to speak when they are not sure of a student's race/ethnicity. Hence, to the extent that observers rely on speech, the language question is a validity check that is not completely independent from observers' classifications, which also measure language background, at least in some cases.

Thus, we can expect that observer's reports are favored by the use of language background as a criteria (as opposed to some other checks, such as parents' immigration history, which are not feasible to ask in the NAEP).

Second, we have to consider that responses to the language question may also be subject to more errors for the younger students for the same reasons that race/ethnicity self-reports are less accurate.

Third, the question used here is a very narrow and simplified measure of language background because it asks for the language that most people speak in the home. Students are classified into two categories by their responses: the SPAND group, in which Spanish is the dominant language in the home, and the ENGOLD group, in which English or a language other than Spanish predominates in the home. The ENGOLD classification does not differentiate among three distinct subgroups of Hispanic homes. (1) The English-dominant bilingual homes where there is at least one grandparent or relative who speaks Spanish although English is spoken most of the time. (2) The English-monolingual homes, where Spanish is not spoken at all and the residents are usually persons of Hispanic ancestry from later-born generations. And, (3) the other-language-dominant home, which is very rare among Hispanics, but could include homes in which an Indian language such as Quechua predominates.

The majority of Hispanic American students fall into the first subgroup, English-dominant bilingual. Thus, the inclusion of students in the ENGOLD category does not mean that Spanish is never spoken in their home. Students from English-Spanish bilingual homes where English predominates are not identified by the question as having a Spanish-language home; they are classified as ENGOLD and not SPAND. Among the 17-year-olds in the NAEP, the ENGOLD category represented as many as 57% of the students self-reported to be Hispanic (i.e., 57% did not indicate that most persons spoke Spanish in their home).

Another limitation of language use as a criterion is that it can identify as Hispanic individuals of non-Hispanic ancestry who have learned to speak Spanish because they were raised in a Spanish-speaking country or received formal instruction in the language. This error is less likely in the present situation. While some persons from this group might report being able to speak Spanish, few would be expected to say that it is spoken by most persons in their home. Among the 17-year-olds, Table 7 shows that only 4% of the students from SPAND homes self-identified as non-Hispanic.

In general, students in the U.S. with SPAND homes are probably Hispanic; on the other hand, Hispanic students do not necessarily come from SPAND homes. In fact, the majority come from ENGOLD homes. Specifically, among, 17-year-olds in the NAEP, 96% of the SPAND group was Hispanic, whereas 57% of the Hispanic group originated in ENGOLD homes. The ethnicity of students from ENGOLD homes can be Hispanic or non-Hispanic.

Despite these limitations and ambiguities, the responses to the language question are compared to the identification reports because there are no other indicators of Hispanicity available in the NAEP. The comparison is useful in two ways. One, these data give a basis for contrasting the accuracy of

observer reports or self-reports in the younger groups with that of 17-year-olds. Two, these comparisons allow us to evaluate which identification procedure has a higher degree of agreement with the language background in the younger groups.

If we assume that the proportion of Hispanic students from Spanish-language homes is the same at all three ages, we should find the same relationship to language background in younger vs. older students provided that self-reports and observer reports were equally valid for all ages. Large departures from the pattern evident for 17-year-olds would be indicative of greater error rates in the younger groups.

The results demonstrate some differences in the relationship between language background and self-reports among the three ages. Tables 6a-c show that the percentages of students with ENGOLD homes among those who identified as Hispanic (the H/* group) for 9 and 13-year-olds are slightly larger (65% and 63%, respectively) than the percentage for 17-year-olds (57%), which suggests overidentification (i.e. non-Hispanics self-identifying as Hispanics). There is also evidence of underidentification in Table 7, which shows that the proportion of SPAND students who failed to self-identify as Hispanic (false negatives) was higher among 9 (27%) and 13-year-olds (15%) relative to that of 17-year-olds (4%). Therefore, these comparisons with statistics on 17-year-olds suggest a slightly higher number of self-classification errors among 9- and 13-year-olds, possibly in the direction of both over- and under-identification of Hispanic students.

Although one would expect observer classifications to be equally accurate with younger and older students, there is also evidence of modestly higher rates of classification errors at Ages 9 and 13. Using the figures in Table 6 for the observed categories, one can derive corresponding proportions for the

/H group (visually identified as Hispanic, regardless of self classification) not directly shown on the table. The proportions of ENGOLD students in the group observed to be Hispanic (/H) for Ages 17 and 13 are about the same (54% and 55%, respectively) but this proportion is lower for 9-year-olds (48%), which suggests underidentification at Age 9. Table 7 shows that observers missed classifying the SPAND group students as Hispanic (false negatives) more at Ages 9 (27%) and 13 (17%) than at Age 17 (11%), which is also evidence of underidentification.

These conclusions must be considered tentative. The differences among the three ages in the relationship between language in the home and self-report procedures could also reflect more errors in responses to the language question in the younger respondents or differences in language-use distribution between younger and older Hispanic students.

Regardless of the method used, it appears that there are more classification errors for the younger age groups. But which method can be said to be the lesser of two evils? Moreover, which way of measuring agreement should be given more weight? Should we consider (1) what percent of the students identified as Hispanic by each method fall into the SPAND group, or (2) what percent of the SPAND group is identified as Hispanic by each method? These measures of agreement can give us different answers. We argue that measure #1 should be given less weight because it depends on the size of the ENGOLD group and membership in this group tells us little about ethnicity directly.

The dependence of measure #1 on the size of the ENGOLD groups is more evident when it is expressed algrebraically. If we represent the number of SPAND members among those identified as Hispanic by a given method as S and the number of ENGOLD members in that classification as E, the total number of

persons classified as Hispanic by that method is H · S + E. Then the proportion of SPAND members in that group is equal to (S/H - S/(S + E)). Thus, the size of measure #1 contains a term in the denominator that depends on the size of the ENGOLD group. In contrast, measure #2, the proportion of SPAND members identified as Hispanic by a given method, can be represented as S/T, where T is the total number of persons in the SPAND group. Thus, we fin that measure #2 has no terms related to the size of the ENGOLD group, which therefore gives us less ambiguous results.

Though preferable to measure #1, measure #2 has the drawback of overestimating the accuracy of observer reports in the overall group that comprises all language categories. Our results in Table 6a show that observers are more accurate in identifying 17-year-old self-reported as Hispanic in the SPAND group (566/639 = 89%) than self-reported as Hispanic in the ENGOLD group (539/845 = 64%), as predicted. Therefore, if we assume we can extrapolate from these findings to the younger groups, we can expect that observers will be more accurate in the SPAND group than in the ENGOLD group, and that measure #2 will give us a maximum estimate of their accuracy in general for students Age 9 and 17. In contrast, there is no reason to believe that the accuracy of self-reports is affected by language background. Thus, measure #2 is probably a good estimate of the accuracy of self-reports overall and <u>not</u> a maximum.

The results in Table 6 show that with measure #1 there are proportionately more students in the SPAND group among those observed to be Hispanic than among those self-reported to be Hispanic at all three ages. But with measure #2, shown in Table 7, the less ambiguous method, both classifications identify the same proportion of SPAND students as Hispanic at ages 9 and 13. If only observer reports are used to classify students rather than self-reports, 17%

of the 13-year-olds and 27% of the 9-year-olds who came from Spanish-language homes would be left out of the Hispanic category. If only self-reports are used to classify students, 15% of the 13-year-olds and 27% of the 9-year-olds with Spanish-language homes are missed. Either way, a substantial number of the students who are most likely Hispanic are left out of the category.

In sum, measures #1 and #2 give us different answers to the question concerning which method more closely agrees with language background for Ages 9 and 13. We place more weight on measure #2 because it is less ambiguous, since it doesn't depend on the size of the ENGOLD group. The results show that the two methods are equal in accuracy using measure #2, i.e., in identifying students from SPAND homes. While there is no reason to believe that measure #2 is a biased estimate of the accuracy of self-reports in general, the results on 17-year-olds suggest that measure #2 represents a maximum estimate of the observers' accuracy in general. If an external criterion for race/ethnicity were known for students in the ENGOLD group, we would expect to find that the accuracy for self reports in this group was no lower than measure #2 (in the SPAND group). But the accuracy of observers would most likely decrease in the ENGOLD group, to a level below that of self-reports, as it does in 17-year-olds. Therefore, we infer that self-reports are probably somewhat more accurate than observer reports overall for the younger group because the estimated accuracy of self-reports equals the maximum estimate of accuracy for observer reports.

Because of the absence of a good criterion for Hispanicity among students with predominantly English-language homes, this conclusion partly rests on inferences and assumptions about similarities between 17-year-olds and the younger groups that may not hold. We are also ignoring possible inaccuracies in the language responses. Thus, this conclusion should be considered

tentative until better validity checks for race/ethnicity identification and language background are included in future studies.

The analyses of concordance rates provide another perspective. There is a trend toward lower concordance rates between self- and visual-identification for the younger students who self-identified as White, Hispanic, or American Indian, although the results for the latter group were not statistically significant because of the small sample sizes. These findings are equivocal because lower concordance rates could arise from either more observer errors or more self-report errors among younger students.

More self-report errors seem more likely. There is less reason to believe observers would be systematically less accurate with younger students. Although the literature on ethnic identification in children suggests that students of these ages are aware of their ethnicity, there could be increases in errors of self-classification because of the problems children have in responding to questionnaires. Nevertheless, rise in self-report errors appeared to be small. Usually, the differences in concordance rates between the younger cohorts and 17-year-olds were less than five percentage points.

An examination of the overall percentages of students classified as Hispanic in the total sample in Tables 3 and 5 shows that proportionately more 9-year-olds (11.2%) and 13-year-olds (8.5%) self-identified as Hispanic in 1984 than did 17-year-olds in 1984 (7.8%). They differed from the percentage visually identified as Hispanic by 3.7% and 1.7%, respectively, in contrast to 1.3% for the 17-year-olds. These self-report figures are more consistent than observer reports with the high drop-out rates among Hispanic students in junior and senior high school, which lead to the presence of proportionately more Hispanic students in the earlier grades. However, it is also possible that there may be a slight increase in overidentification errors, especially for 9-year-olds.

Finally, in terms of continuity with other data sets, self-report methods offer a definite advantage over observer reports, other things being equal, because the use of questionnaire responses to classify students by race/ethnicity is almost universal. Of course, most of these other data sets include high-school-age students only. It would be desirable to keep identification procedures for the different age groups as standard as possible to facilitate possible future follow-ups at age 17 for the students who in 1984 were nine and 13 years old.

To summarize, let us return to our two questions about the younger cohorts. Are self-reports and observer reports less accurate with students Age 9 and 13 than with 17-year-olds? Yes, both methods appear to be slightly less accurate, judging from the language data and concordance rates, but the differences in errors are modest.

Which method is the lesser of two evils? The evidence from the language data, which is not clearcut, weakly favors classification by self-reports. The ambiquity arises from the absence of a good criterion measure of Hispanicity for young students from homes in which English predominates (ENGOLD). The accuracy of each method cannot be directly measured in this group against a criterion. But based on the results for 17-year-olds, we can infer that, in the ENGOLD group, the accuracy of observer reports is probably lower than the accuracy for self-reports. In contrast, membership in the SPAND-home category does serve as a good criterion of Hispanicity. However, in this group, we can infer that observers are likely to be maximally accurate in recognizing Hispanics; that is, they are probably less accurate in the group from ENGOLD homes. This maximum accuracy rate for observers was found to be no higher than the rate for self-reports. In the SPAND group, both identification procedures had about the same rate of accuracy. In sum,

self-reports are probably superior in the ENGOLD group, which is the largest, and they have an accuracy matching the maximum of observer reports in the smaller SPAND group. Thus, these results suggest that, overall, classifications of younger students are probably more accurate when self-reports are used rather than observer reports.

The self-report procedure is supported to some extent by other considerations: the literature on children's awareness of race/ethnicity; concordance rates for the younger cohorts that were usually only five percentage points below the lowest value for 17-year-olds; agreement with demographic trends; and standardization of classification methods which enhances the continuity of NAEP surveys among all ages and with other surveys.

Taken together, the weight of these considerations favors continued use of self-report procedures for classification purposes for the younger age groups, with observer reports as a substitute in the small percentage who do not respond to the race/ethnicity questions.

Conclusions and Recommendations

There is little doubt that the use of self-reported race/ethnicity for classifying 17-year-olds in the most recent NAEP assessment represents a marked improvement over pre-1984 surveys for the classification of Hispanic students, and, possibly, for American Indians, Asian Americans, and Others. This change makes relatively little difference for White and Black students, for whom there is a consistently high level of agreement between the different methods. Thus, the NAEP data sets appear to have adequate continuity over time for all age groups in the White and Black subsamples.

However, the 1984 and pre-1984 data sets for Hispanics and other groups are markedly discontinuous because there are large discrepancies between visual and self-identification for Hispanics, American Indians, Asian Americans, and Others.

The concordance rates give indirect evidence that the validity of self-reports for younger students varies by ethnic group. In the Black subsamples, the high concordance rates suggest that self-reports of students age 9, 13, and 17 appear to be equally valid. For Asians and Others, there is also no evidence of lower concordance in the younger cohorts. The younger American Indian subsamples had lower concordance rates but not significantly so. However, the use of self-reports for 9 and 13-year-old Hispanic and White students appears to be slightly less valid than for 17-year-olds. Differences between the language-background distributions of Hispanic students aged 9 and 13 and that of 17-year-olds also suggest slightly more self-classification errors among the younger cohorts.

Nevertheless, considering a variety of factors, the advantages of self-reports for the younger groups appear to outweigh their limitations. There is evidence of racial/ethnic awareness in young children; concordance rates are only slightly lower in the younger ages; in the 17-year-olds, fluctuations in concordance rates over the years indicate low reliability for observer reports for several groups; demographic figures suggest greater agreement with self-reports; and self-identification leads to more continuity with other data sets.

The language background results, though ambiguous, suggest that the self-report procedure is likely to give more accurate classifications than visual identification when the Hispanic students originate from predominantly English-language homes, as the majority do. Observers were probably more

accurate in classifying students from predominantly Spanish-language homes than Hispanic students with other kinds of language background. In the group from Spanish-language-dominant homes, there were no significant differences in the accuracy of the two classification procedures. It appears that the accuracy of self-identification at least equals and is probably superior to that of visual identification in the overall Hispanic group. This conclusion should be verified in future studies using parental reports of racial/ethnic identification and language background to serve as criterion measures.

Because of the difficulties in accurately classifying very young students, the feasibility of adding more race/ethnicity indicators in the NAEP should be researched. For example, a pilot study could investigate whether school records on race/ethnicity could be released and whether they are accurate as evaluated against parental interviews. Other supplementary indicators might include a standard mechanical procedure for coding surnames, if NAEP policies were changed to allow names to be included for coding purposes only.

Furthermore, errors in reporting race/ethnicity could perhaps be reduced by having administrators read the questions out loud, give examples, and clarify the meaning of terms for the younger students. The use of the joint terms containing "American" (e.g., Mexican American, or American Indian) may be particularly confusing for children who are less experienced readers. To many students, the terms "Hispanic" or "Asian" are probably less familiar than the names of the individual subgroups. "Focus group interviews" as described by Scherr (1980) could be carried out with the children to shed light on problems in wording.

The low concordance rates between visual- and self-reported identifications suggest a need to re-examine the effect of classification procedures on achievement data. Such an analysis would help to determine

whether the use of observer reports for classification purposes introduced biases in reported means for Hispanics in past NAEP assessments. (Although the racial/ethnic identification methods have equally low or lower agreement for the Asian American, American Indian, and other groups, achievement means have not been reported in NAEP publications separately for these groups which represent a relatively small percentage of the total samples.) If biases are found, then analyses of trends over time should be recomputed for 17-year-olds (for whom self-reports are unquestionably valid) because the amount of overlap between visual and self-identification shows such large fluctuations from year to year, indicating substantial discontinuity for Hispanic students. Another study in progress will deal with these issues (Pennock-Roman and Rivera, forthcoming).

References

Aboud, Frances E. 1984. "Social and cognitive bases of ethnic identity constancy." Journal of Genetic Psychology (December):217-230.

Beaton, A.E. 1986. National Assessment of Educational Progress 1983-84: A technical report. Princeton, NJ: Educational Tetsing Service.

Brand, Elaine S., Ruiz, Rene A., and Padilla, Amado M. 1974. "Ethnic identification and preference: A review." Psychological Bulletin (November):860-890.

Buriel, J. R., 1987. "Ethnic labeling and identity among Mexican Americans," in Phinney, J. S. & Rotheram, M. J. (Eds.), Children's ethnic socialization: Pluralism and development. Sage Publications, Beverly, CA.

Fernandez, Edward W. 1975. "Comparison of Spanish surname and persons of Spanish origin in the United States." Technical paper No. 38, Washington, DC: Government Printing Office.

Folsum, Ralph E., Jr. 1977. "National assessment approach to sampling error estimation." Sampling Error Monograph 250-796-5. Research Triangle Institute, Research Triangle Park, North Carolina.

Garcia, John A. 1982. "Ethnicity and Chicanos: Measurement of ethnic identification, identity, and consciousness." Hispanic Journal of Behavioral Sciences (July):295-314.

Ginorio, Angela B. and Berry, Paul C. 1972. "Measuring Puerto Rican's perceptions of racial characteristics." Proceedings, 80th Annual Convention, American Psychological Association.

Hernandez, Jose, Estrada, Leo, and Alvirez, David. 1973. "Census data and the problem of conceptually defining the Mexican American population. Social Science Quarterly (March):671-864.

Johnson, Charles E., Jr. 1974. "Consistency of reporting of ethnic origin in the Current Population Survey." Technical Paper No. 31 (February), U. S. Bureau of the Census.

Kish, Leslie. 1965. Survey Sampling. New York Wiley.

Lampe Philip E. 1984. "Mexican Americans Labeling and mislabeling." Hispanic Journal of Behavioral Sciences (March):77-85.

Levin, Michael J., McKenney, Nampeo R., and Berman, Patricia. 1984. "Uses and interpretation of racial and ethnic data from the U.S. Census." Social Statistic Section, Proceedings of the American Statistical Association, pp. 88-95.

LeVine, Elaine S. and Ruiz, Rene A. 1978. "An exploration of multicorrelates of ethnic group choice." Journal of Cross-Cultural Psychology (June):179-190.

Lowry, Ira S. 1980. "The science and politics of ethnic enumeration." Paper, Annual Meeting of the American Association for the Advancement of Science, California (January).

Massey, James T. 1980. "A comparison of interviewer observed race and respondent reported race in the National Health Interview Survey." Social Statistics Section, Proceedings of the American Statistical Association, pp. 425-428.

McKenney, Nampeo R., Farley, Reynolds and Levin, Michael J. 1983. "Direct and indirect measures of ethnicity." Social Statistics Section, Proceedings of American Statistical Association, pp. 123-130.

Messick, Samuel J., Beaton, Albert E., and Lord, Frederic. 1983. "A new design for a new era." NAEP Report 85-1. National Assessment of Educational Progress, Educational Testing Service, Princeton, NJ (March).

Miller, Michael N. 1976. "Mexican Americans, Chicanos, and others: Ethnic self-identification and selected social attributes of rural Texas youth." Rural Sociology (Summer): 234-247.

Minutes and Report of Committee Recommendations, Census Advisory. 1979. Committee on Population Statistics, U.S. Bureau of the Census (April 6).

Moore, R. Paul, Chromy, James R., and Rogers, W. Todd. 1974. "The National Assessment approach to sampling." Monograph, National Assessment of Educational Progress, Education Commission for the States, Denver, CO.

National Assessment of Educational Progress. 1981. "Procedural handbook: 1979-80 reading assessment." NAEP Report No. 11-RL-40. National Assessment of Educational Progress, Education Commission of the States, Denver, CO. (September)

Panel on Decennial Census Plans, Committee on National Statistics, Assembly of Behavioral and Social Sciences, National Research Council. 1978. "Counting the people in 1980: An appraisal of Census plans." National Academy of Sciences, Washington, D.C.

Pennock-Roman, Maria and Rivera, Charlene. "Issues in Ethnic Identification in the National Assessment of Educational Progress, Part II: Effect on Mean Reading Achievement Reported for Hispanics." (Forthcoming). Research Report, Educational Testing Service.

Rice, Audrey S., Ruiz, Rene A., and Padilla, and Amado M. 1974. "Person perceptions, self-identity, and ethnic group preference in Anglo, Black, and Chicano preschool and third-grade children." Journal of Cross-Cultural Psychology (March):100-108.

Scherr, Marvin G. 1980. "The use of focus group interviews to improve the design of an administrative form: A case study at the Social Security Administration." Survey Research Methods Section, Proceedings of the American Statistical Association, pp. 347-352.

Schneider, Phillip A.A. 1980. "Race/Ethnic data collection in the Federal Civil Service." Survey Research Methods Section, Proceedings of the American Statistical Association, pp. 364-367.

Smith, Tom W. 1980. "Ethnic measurement and identification." Ethnicity, (March):78-95.

Smith, Tom W. 1981. "The subjectivity of ethnicity," unpublished paper, NORC.

Sullivan, Teresa A., Gillespie, Francis P., Hout, Michael, and Greeley, Andrew W. 1983. "Surname versus self-identification in the analysis of Hispanic data." Social Statistics Section, Proceedings of the American Statistical Association, pp 117-122.

Tienda, Marta and Ortiz, Vilma. 1986. "'Hispanicity' and the 1980 Census." Social Science Quarterly, in press.

Tukey, John W., and Mosteller F. 1977. Data Analysis and Regression. Addison-Wesley, Reading, MA.

Weiland, Andrea and Coughlin, Ronald. 1979. "Self-identification of White and Mexican-American first and third graders." Journal of Cross-Cultural Psychology (September):356-365.

Table 1

Data Sets Used in this Study

Year	Age	Grade	Weighted N	Unweighted N	Content Areas Assessed
1983-84[a]	17		2,495,425	18,881	Reading
		11	2,550,137	18,976	
1981-82	17		2,965,293	31,044	Science, Math, Citizenship, and Social Studies
1979-80	17		4,193,016	38,025	Reading and Literature
1975-76	17		3,311,938	35,631	Math, Social Studies and Citizenship
1983-84[a]	13		2,498,791	15,859	Reading
		8	2,495,046	16,412	
1981-82	13		3,070,931	30,899	cience, Math, Citizenship, and Social Studies
1983-84[a]	9		2,691,609	16,670	Reading
		4	2,823,205	17,747	

[a]The age and corresponding grade samples for 1983-84 overlap by about two-thirds. There were 16,787 students (unweighted N) who were age 17 and in Grade 11; 14,515 students (unweighted N) who were age 13 and in Grade 8; 12,953 students (unweighted N) who were age 9 and in Grade 4.

Table 2

Percentage Distribution of Visual Identification Within Each Self-Classification Across Different Assessment Years

Self-Classification and Age	Visual Identification					
	White	Black	Hispanic	Indian	Asian	Other
White						
Age 17	99.0-99.4	0.1-0.3	0.3-0.5	0.0-0.0	0.0-0.0	0.0-0.0
Age 13	98.3-99.4	0.2-0.2	0.4-1.0	0.0-0.0	0.1-0.1	0.0-0.1
Age 9	96.4	0.4	2.3	0.0	0.1	0.0
Black						
Age 17	2.0-4.5	94.8-97.7	0.1-1.5	0.0-0.2	0.0-0.1	0.0-0.0
Age 13	1.9-4.0	95.1-97.8	0.3-0.5	0.0-0.0	0.0-0.0	0.1-0.1
Age 9	3.2	94.9	1.4	0.1	0.3	0.0
Hispanic						
Age 17	18.8-33.4	2.6-13.2	52.2-74.5	0.1-0.8	1.2-2.6	0.0-0.9
Age 13	23.9-35.1	6.5-14.1	48.4-66.2	0.1-0.4	1.6-1.9	0.?-1.5
Age 9	41.5	10.3	45.8	0.3	1.8	0.2
Indian						
Age 17	59.9-91.4	2.5-7.7	1.6-2.3	12.1-29.3	0.2-1.9	0.0-0.0
Age 13	73.8-74.2	6.4-10.5	1.7-6.7	11.6-12.9	0.4-0.6	0.3-1.0
Age 9	70.5	9.1	10.0	7.0	2.4	0.9
Asian						
Age 17	9.8-21.7	0.9-6.7	1.3-11.2	0.0-3.2	64.7-85.9	0.5-4.7
Age 13	11.6-20.1	3.3-3.9	3.5-9.9	0.3-1.0	70.8-72.3	0.8-2.5
Age 9	13.9	4.9	6.9	0.7	70.2	3.4
Other						
Age 17	66.6-82.4	4.8-14.8	3.7-8.2	0.3-0.9	2.8-10.4	0.9-6.7
Age 13	71.1-75.0	11.. 2	5.2-5.9	0.3-1.1	3.8-10.6	1.1-1.7
Age 9	62.4	18.3	12.2	0.4	5.8	0.8

Note. 1976 data are excluded for the Other observed column because, in that year, observers did not have the option of using the Indian or Asian classification. Students in these groups were classified as Other.

Table 3

Percentage Breakdown of Visual Identification Among Students

Self-Identified a Hispanic: Comparison Across Assesaments

			Visual Identification								
Year	Age	Grade	White	Black	Hispanic	Indian	Asian	Other	Missing	Total[c]	%Sample
1984[a]	17		18.8	4.0	74.5	0.1	1.7	0.9	0.0	100.0	7.8
		11	19.5	4.6	72.3	0.2	2.6	0.8	0.0	100.0	7.9
1982	17		25.2	13.2	58.3	0.8	2.6	0.1	0.0	100.0	9.6
1980	17		33.4	12.7	52.2	0.5	1.2	0.0	0.0	100.0	9.2
1976	17		31.4	2.6	64.1	__[b]	__[b]	1.9[b]	0.0	100.0	4.7
1984[a]	13		23.9	6.5	66.2	0.1	1.6	1.5	0.2	100.0	8.5
		8	21.8	4.3	70.6	0.2	1.7	1.3	0.1	100.0	8.6
1982	13		35.1	14.1	48.4	0.4	1.9	0.2	0.0	100.0	13.7
1984[a]	9		41.5	10.3	45.8	0.3	1.8	0.2	0.1	100.0	11.2
		4	37.3	9.8	50.4	0.3	1.9	0.2	0.1	100.0	11.0

[a]The age and corresponding grade samples for 1983 are about two-thirds overlapping.

[b]In 1975, observers did not have the option to use the Indian and Asian classifications. Students in these categories were classified as Other.

[c]Due to rounding, all percentages do not add up to 100%.

Table 4

Groups Ranked by Percent Concordance between Self and Visual Identification in Various NAEP Assessments: Ages 9, 13, and 17

Hispanics	Percent Concordance	Non-Hispanics
	1-5	O (Age 17, 1976, 1980, 1984; Age 13, 1982, 1984; Age 9, 1984)
		I (Age 17, 1975)
	6-10	O (Age 17, 1976, 1982); I (Age 9, 1984)
	11-15	I (Age 17, 1980, 1984); I (Age 13, 1982; 1984)
	...	
	26-30	I (Age 17, 1982)
	...	
H (Age 13, 1982, Age 9, 1984)	46-50	
H (Age 17, 1980)	51-55	
H (Age 17, 1982)	56-60	
H (Age 17, 1976)	61-65	A (Age 17, 1980)
H (Age 13, 1984)	66-70	A (Age 17, 1984); A (Age 9, 1984)
H (Age 17, 1984)	71-75	A (Age 17, 1976); A (Age 13, 1982; 1984)
	...	
	86-90	A (Age 17, 1982); B (Age 17, 1976)
	90-95	B (Age 9, 1984); B (Age 17, 1984)
	95-100	W (Age 17, 1976, 1980, 1982, 1984; Age 13, 1982, 1984; Age 9, 1984)
		B (Age 17, 1980, 1982; Age 13, 1982, 1984)

W = White; B = Black; H = Hispanic; I = American Indian; A = Asian; O = Other. Concordance rates are computed by using the number of cases both visually and self-classified into the category by the number of cases self-identified in the same category. In 1975, there were no separate categories for visual identification of American Indians and Asian groups, in which case, self-identification of Asian or American Indian was considered concordant with visual identification if the visual category was Other.

Table 5

Percentage Breakdown of Self-Identification Among Students

Visually Identified as Hispanic: Comparison Across Assessments

			Self-Identification								
Year	Age	Grade	White	Black	Hispanic	Indian	Asian	Other	Missing	Total[b]	%Sample
1984[a]	17		5.9	0.7	88.9	0.2	3.1	1.1	0.0	100.0	6.5
		11	4.9	0.9	89.4	0.3	3.7	0.9	0.0	100.0	6.3
1982	17		3.2	0.3	93.6	0.6	0.3	1.1	0.9	100.0	6.0
1980	17		6.1	2.9	84.4	0.7	0.4	3.4	2.1	100.0	5.7
1976	17		5.0	1.1	75.8	2.6	1.8	5.1	8.6	100.0	4.0
1984[a]	13		10.6	1.0	82.8	1.3	2.2	2.1	0.0	100.0	6.8
		8	10.2	0.9	84.1	1.0	2.0	1.8	0.0	100.0	7.2
1982	13		3.3	0.5	90.7	0.7	0.9	2.5	1.5	100.0	7.3
1984[a]	9		21.1	2.6	68.7	2.0	1.3	4.3	0.0	100.0	7.5
		4	20.7	2.3	70.5	1.4	1.2	3.8	0.0	100.0	7.9

[a]The age and corresponding grade samples for 1984 are about two-thirds overlapping.

[b]Due to rounding, all percentages do not add up to 100%.

Table 6a

Dominant Home Language by Self/Visual Ethnic Categories: Age 17

Self/Visual Group		Spanish	English/Other	Total in Category
H/H	f	566	539	1,105
	R%	51.2%	48.8%	100.0%
H/NH	f	73	306	379
	R%	19.3%	80.7%	100.0%
NH/H	f	10	131	141
	R%	7.1%	92.9%	100.0%
NH/NH	f	15	17,344	17,359
	R%	0.1%	99.9%	100.0%
H/*	f	639	845	1,484
	R%	43.1%	56.9%	100.0%
NH/*	f	25	17,475	17,500
	R%	0.1%	99.9%	100.0%
*/H	f	576	670	1,246
	R%	46.2%	53.8%	100.0%
*/NH	f	88	17,650	17,738
	R%	0.5%	99.5%	100.0%
TOTAL SAMPLE	f	664	18,320	18,984
	R%	3.5%	96.5%	100.0%

Note. The weighted frequencies (f) were scaled to be commensurate with the raw Ns by dividing them by the ratio [(total weighted N)]/(total raw N)] corresponding to each sample. The sample sizes shown here are somewhat larger than the other 1984 analyses because reading achievement scores had been computed for more students when these analyses were run.

Table 6b

Dominant Home Language by Self/Visual Ethnic Categories: Age 13

Self/Visual Group		Spanish	English/Other	Total in Category
H/H	f	487	496	983
	R%	49.5%	50.5%	100.0%
H/NH	f	72	434	506
	R%	14.2%	85.8%	100.0%
NH/H	f	54	161	215
	R%	25.1%	74.9%	100.0%
NH/NH	f	42	15,789	15,831
	R%	0.3%	99.7%	100.0%
H/*	f	559	930	1,489
	R%	37.5%	62.5%	100.0%
NH/*	f	96	15,950	16,046
	R%	0.6%	99.4%	100.0%
*/H	f	541	657	1,198
	R%	45.2%	54.8%	100.0%
*/NH	f	114	16,223	16,337
	R%	0.7%	99.3%	100.0%
TOTAL SAMPLE	f	655	16,880	17,535
	R%	3.7%	96.3%	100.0%

NOTE: The weighted frequencies (f) were scaled to be commensurate with the raw Ns by dividing them by the ratio [(total weighted N)/(total raw N)] corresponding to each sample. The sample sizes shown here are somewhat larger than the other 1984 analyses because reading achievement scores had been computed for more students when these analyses were run.

Table 6c

Dominant Home Language by Self/Visual Ethnic Categories: Age 9

Self/Visual Group		Spanish	English/Other	Total in Category
H/H	f	521	344	865
	R%	60.2%	39.8%	100.0%
H/NH	f	133	890	1,023
	R%	13.0%	87.0%	100.0%
NH/H	f	137	265	402
	R%	34.1%	65.9%	100.0%
NH/NH	f	108	14,401	14,509
	R%	0.7%	99.3%	100.0%
H/*	f	654	1,234	1,888
	R%	34.6%	65.4%	100.0%
NH/*	f	245	14,666	14,911
	R%	1.6%	98.4%	100.0%
*/H	f	658	609	1,267
	R%	51.9%	48.1	100.0%
*/NH	f	241	15,291	15,532
	R%	1.6%	98.4	100.%
TOTAL SAMPLE	f	899	15,900	16,799
	R%	5.4%	94.6%	100.0%

Note. The weighted frequencies (f) were scaled to be commensurate with the raw Ns by dividing them by the ratio [(total weighted N)] /(total raw N)] corresponding to each sample. The sample sizes shown here are somew at larger than the other 1984 analyses because reading achievement scores had been computed for more students when these analyses were run.

Table 7

Frequencies and Percentages in Self/Visual Categories among Students with Predominantly Spanish-Language Homes

Age	Self-Identification		Visual Identification: H	NH	T	
Age 17	H	f	566	73	639	= H/*
		(%)	(85)	(11)	(96)	
	NH	f	10	15	25	= NH/*
		(%)	(2)	(2)	(4)	
	T	f	576	88	664	
		(%)	(87)	(13)	(100)	
			= */H	= */NH		
Age 13	H	f	487	72	559	= H/*
		(%)	(74)	(11)	(85)	
	NH	f	54	42	96	= NH/*
		(%)	(8)	(6)	(15)	
	T	f	541	114	655	
		(%)	(83)	(17)	(100)	
			= */H	= */NH		
Age 9	H	f	521	133	654	= H/*
		(%)	(58)	(15)	(73)	
	NH	f	137	108	245	= NH/*
		(%)	(15)	(12)	(27)	
	T	f	658	241	899	
		(%)	(73)	(27)	(100)	
			= */H	= */NH		

Note. All percentages are found by dividing the cell or marginal frequency by the total number from Spanish language homes. The marginal frequencies H/*, NH/*, */H, and */NH in this case are summed within the Spanish-dominant category; they do not represent sample totals in this case.

Appendix

Breakdown of Visual Identification among

Self-identified Whites, Blacks, Americans

Indians, Asians, and Others

Table A.1

Percentage Breakdown of Visual Identification Among Students Self-Identified as White: Comparison Across Assessments

			Visual Identification								
Year	Age	Grade	White	Black	Hispanic	Indian	Asian	Other	Missing	Total[c]	% Sample
1984[a]	Age 17		99.0	0.3	0.5	0.0	0.1	0.1	0.0	100.0	74.1
		Gr. 11	99.1	0.3	0.4	0.0	0.1	0.1	0.0	100.0	73.2
1982	Age 17		99.4	0.3	0.3	0.0	0.0	0.0	0.0	100.0	69.3
1980	Age 17		99.3	0.1	0.5	0.0	0.0	0.0	0.0	100.0	73.5
1976	Age 17		99.2	0.3	0.3	—[b]	—[b]	0.2[b]	0.0	100.0	71.2
1984[a]	Age 13		98.3	0.2	1.0	0.0	0.1	0.1	0.3	100.0	72.3
		Gr.8	98.2	0.3	1.0	0.0	0.2	0.1	0.2	100.0	72.3
1982	Age13		99.4	0.2	0.4	0.0	0.1	0.0	0.0	100.0	63.3
1984[a]	Age 9		96.4	0.4	2.3	0.0	0.8	0.0	0.1	100.0	69.4
		Gr. 4	96.3	0.4	2.4	0.0	0.8	0.0	0.1	100.0	69.2

[a]The age and corresponding grade samples for 1984 are about two-thirds overlapping.

[b]In 1976, observers did not use the Indian and Asian classifications. Students in these categories were classified as Other.

[c]Due to rounding, all percentages do not add up to 100%.

Table A.2

Percentage Breakdown of Visual Identification Among Students Self-Identified as Black: Comparison Across Assessments

			Visual Identification								
Year	Age	Grade(s)	White	Black	Hispanic	Indian	Asian	Other	Missing	Total[c]	% Sample
1984[a]	Age 17		4.5	94.8	0.3	0.0	0.2	0.1	0.0	100.0	13.9
		Gr. 11	3.9	95.5	0.4	0.0	0.2	0.0	0.0	100.0	14.7
1982	Age 17		2.1	97.7	0.1	0.0	0.0	0.0	0.0	100.0	14.7
1980	Age 17		2.0	96.5	1.5	0.0	0.0	0.0	0.0	100.0	10.8
1976	Age 17		4.2	95.2	0.5	0.1	—[b]	—[b]	0.0[b]	100.0	8.8
1984[a]	Age 13		4.0	95.1	0.5	0.0	0.0	0.1	0.4	100.0	13.9
		Gr.8	4.0	94.8	0.5	0.0	0.0	0.1	0.6	100.0	14.0
1982	Age 13		1.9	97.8	0.3	0.0	0.0	0.1	0.0	100.0	11.6
1984[a]	Age 9		3.2	94.9	1.4	0.1	0.3	0.0	0.0	100.0	13.9
	Gr.4		2.9	95.4	1.3	0.0	0.4	0.0	0.1	100.0	14.6

[a]The age and corresponding grade samples for 1984 are about two-thirds overlapping.

[b]In 1976, observers did not have the option to use the Indian and Asian classifications. Students in these categories were classified as Other.

[c]Due to rounding, all percentages do not add up to 100%.

Table A.3

Percentage Breakdown of Visual Identification Among Students Self-Identified as Indian: Comparison Across Assessments

			Visual Identification								
Year	Age	Grade	White	Black	Hispanic	Indian	Asian	Other	Missing	Total[c]	% Sample
1984[a]	Age 17		75.8	7.3	1.6	12.1	1.9	1.3	0.0	100.0	0.9
		Gr. 11	81.3	3.4	2.2	10.7	0.8	1.6	0.0	100.0	0.8
1982	Age 17		59.9	7.7	2.1	29.3	1.0	0.0	0.0	100.0	1.6
1980	Age 17		75.8	6.5	2.3	15.2	0.2	0.0	0.0	100.0	1.6
1976	Age 17		91.4	2.5	2.3	—[b]	—[b]	3.8[b]	0.0	100.0	4.6
1984[a]	Age 13		73.8	6.4	6.7	11.6	0.6	1.0	0.0	100.0	1.3
		Gr. 8	74.6	5.4	6.4	11.1	0.6	1.9	0.0	100.0	1.1
1982	Age 13		74.2	10.5	1.7	12.9	0.4	0.3	0.0	100.0	2.9
1984[a]	Age 9		70.5	9.1	10.0	7.0	2.4	0.9	0.0	100.0	1.5
		Gr. 4	70.9	8.7	8.8	7.9	2.7	1.0	0.0	100.0	1.3

[a]The age and corresponding grade samples for 1984 are about two-thirds overlapping.

[b]In 1976, observers did not have the option of using the Indian and Asian classifications. Students in these categories were classified as Other.

[c]Due to rounding, all percentages do not add up to 100%.

Table A.4

Percentage Breakdown of Visual Identification Among Students Self-Identified as Asian: Comparison Across Assessments

Year	Age	Grade	Visual Identification: White	Black	Hispanic	Indian	Asian	Other	Missing	Total[c]	% Sample
1984[a]	Age 17		9.8	4.2	11.2	2.0	68.0	4.7	0.0	100.0	1.8
		Gr. 11	9.9	4.0	12.0	1.9	67.9	4.2	0.0	100.0	1.9
1982	Age 17		9.9	2.6	1.3	0.0	85.9	0.5	0.0	100.0	1.7
1980	Age 17		21.7	6.7	3.1	3.2	64.7	0.6	0.0	100.0	0.8
1976	Age 17		20.5	0.9	5.9	—[b]	—[b]	72.7[b]	0.0	100.0	1.2
1984[a]	Age 13		11.6	3.3	9.9	0.3	72.3	2.5	0.2	100.0	1.5
		Gr. 8	12.0	3.3	9.5	0.0	72.5	2.6	0.1	100.0	1.6
1982	Age 13		20.1	3.9	3.5	1.0	70.8	0.8	0.0	100.0	1.9
1984	Age 9		13.9	4.9	6.9	0.7	70.2	3.4	0.0	100.0	1.4
		Gr. 4	12.3	4.0	6.6	1.3	72.7	3.1	0.0	100.0	1.5

[a]The age and corresponding grade samples for 1984 are about two-thirds overlapping.

[b]In 1976, observers did not have the option to use the Indian and Asian classifications. Students in these categories were classified as Other.

[c]Due to rounding, all percentages do not add up to 100%.