

## **DEVELOPMENT OF CONCEPTS AND METHODS**

**1-1 Initial Planning of Surveys**

**1-2 Publication and Product Planning**

**1-3 Computation of Response Rates**

**1-4 Codes and Abbreviations**

**1-5 Defining Race and Ethnicity Data**

## **SUBJECT: INITIAL PLANNING OF SURVEYS**

### **NCES STANDARD: 1-1**

**PURPOSE:** To provide an initial planning document that includes the information required for a decision on whether or not to proceed with the preliminary design and implementation plans of a specific survey or survey system. This information combined with the information described in Standard 2-1 provides the foundation for a statement of work.

**KEY TERMS:** assessment, design effect, effect size, effective sample size, key variables, minimum substantively significant effect (MSSE), planning document, power, privacy, response rate, survey, and survey system.

**STANDARD:** The initial plan for developing a survey or survey system must include the justification for the study, must describe the survey methodology, and must identify and define key variables. Prior to an OMB fiscal year budget request for data collection, the initial planning document must be presented to the OC/ODC for review and a decision on whether to proceed with the design phase. The initial planning document must include the following:

1. A justification for the survey, including: the rationale for the survey, the goals and objectives, and related hypotheses to be tested. This justification must include evidence that consultations with potential users have occurred.
2. A review of related studies, surveys, and reports of federal and non-federal sources to ensure that part or all of the data are not available from an existing source, or could not be more appropriately obtained by adding questions to existing surveys sponsored by NCES or other agencies. The goal here is to minimize respondent burden. If a new survey is needed, efforts should be made in the development of the questionnaire and any assessment items to minimize the burden to individual respondents.
3. Surveys that involve interviewing students in elementary and secondary schools must adhere to the requirements of the Protection of Pupil Rights Act and related amendments (see 20 U.S.C. 1232h and amendments included in Section 1061 of the No Child Left Behind Act of 2001). Specifically, students may voluntarily participate in surveys; but without written consent from a student's parent, students cannot be *required* to answer questions about the following:
  - a. Political affiliations or beliefs of the student or the student's parent;
  - b. Mental and psychological problems potentially embarrassing to the student and his/her family;
  - c. Sex behavior and attitudes;
  - d. Illegal, antisocial, self-incriminating, and demeaning behaviors;
  - e. Critical appraisals of other individuals with whom the respondents have close family relationships;
  - f. Legally recognized privileged or analogous relationships, such as those of lawyers, physicians, and ministers;

- g. Religious practices, affiliations, or beliefs of the student or the student's parents, or
- h. Income (other than that required by law to determine eligibility for participation or for receiving financial assistance under such a program).

In addition, the confidentiality and privacy provisions of the Privacy Act, the Education Sciences Reform Act of 2002 (ESRA 2002), and the E-Government Act of 2002, Title V, Subtitle A, Confidential Information Protection (CIP 2002) must be taken into account in designing any studies that will collect individually identifiable data from any survey participants (see Standard 4-2).

- 4. A preliminary survey design that discusses the proposed target population, response rate goals (see Standard 2-3), sample design, sample size determination based on power analyses for the MSSEs for key items, data collection methods, and methodological issues.
- 5. A preliminary analysis plan that identifies analysis issues, objectives, key variables (including definitions), minimum substantively significant effect sizes (e.g., the smallest differences the survey is intended to measure in the total population and for key reporting variables by reporting domains), and proposed statistical techniques.
- 6. A list of data items that will be maintained over time as part of an NCES data series, including the justification for each item.
- 7. A preliminary time schedule that accounts for the complete survey cycle from planning to data release.
- 8. A preliminary publication and dissemination plan that identifies proposed major publications and their target audiences (see Standard 1-2).
- 9. A preliminary survey evaluation plan that identifies the proposed analyses necessary for data users to understand the quality and limitations of the survey (see Standard 4-3).
- 10. An internal cost estimate that reflects all of the above items.

## **SUBJECT: PUBLICATION AND PRODUCT PLANNING**

### **NCES STANDARD: 1-2**

**PURPOSE:** To ensure that all proposed NCES products are included in an annual NCES publication plan that will assist with the coordination of publications across divisions, in an effort to avoid duplication and to maximize collaboration. The publication plan will make explicit the status of all anticipated publications for the next year; provide target dates for all mandatory and required publications; and assure that appropriate attention is given to all necessary aspects of the planning process.

**STANDARD 1-2-1:** All NCES publications and data products must be included in the annual NCES publication plan. This includes mandatory, required, and planned publications.

1. Mandatory publications include a limited number of high profile publications that the agency is committed to releasing in a specific month.
2. Required publications are those that are scheduled for release within the fiscal year. It includes most first releases from NCES data collections, including data files, CD-ROMs, and electronic codebooks.
3. Planned publications include all other descriptive, analytic, research and development reports, and other planned reports.

(See List 1-1-A for a description of NCES product types, Standard 7-2 for a description of content requirements by product type, and Standard 6-1 for the type of review required by product type.)

**GUIDELINE 1-2-1A:** A publication should be added to the Review Tracking System (RTS) during the development of the NCES publication plan.

**GUIDELINE 1-2-1B:** A request for printing should be included in the RTS.

**GUIDELINE 1-2-1C:** Program Directors should update changes in the RTS on an as needed basis.

**STANDARD 1-2-2:** Any publications and data products proposed after the annual publication planning process must receive Program Director, Associate Commissioner, and Commissioner approval before inclusion in the RTS.

**STANDARD 1-2-3:** All publications and data products must have a target delivery date for the first deliverable and a target delivery date to the Chief Statistician for NCES review.

**GUIDELINE 1-2-3A:** For Web release only publications, the release date is three weeks after final sign-off by the IES Director or the NCES Commissioner. (For additional information about web publishing, contact: [NCESWebmaster@ed.gov](mailto:NCESWebmaster@ed.gov).)

**GUIDELINE 1-2-3B:** Requests for printing must be approved by the Commissioner. If approved for printing, the date for printed release is approximately 3 to 5 weeks after the final sign-off by the IES Director or the NCES Commissioner.

**STANDARD 1-2-4:** All analytic, descriptive, and research and development publications must have a written analysis plan approved by the Program Director prior to beginning an analysis.

**GUIDELINE:** The analysis plan should be developed in consultation with the Associate Commissioner and the Chief Statistician.

## LIST 1-1-A. — NCES PUBLICATION TYPES

**Brochure/Pamphlets** present an overview of NCES programs or surveys.

**CD-ROMs** present NCES data and related documentation. Products include micro-data files, documentation for micro-data files, data embedded in data analysis systems, and data in electronic tabulations.

**Compendia** are comprehensive resource publications that summarize major education statistics on the status and progress of education at one or more levels of education from preprimary through graduate education, adult education, and lifelong learning.

**Conference Reports** are compilations of papers presented at NCES-sponsored conferences and workshops.

**Data Files** present NCES data and related documentation. Products include micro-data files and documentation for micro-data files.

**Directories** typically present listings of educational institutions and agencies.

**First Look Reports** are a collection of up to 15 pages of tables, presented with a brief introduction, one page of findings, and necessary technical notes that briefly describe the sample design and data collection, per Standard 7-2-A Checklist.

**Guides** provide descriptions of data collection programs and manuals of procedures which describe how to complete the activity.

**Handbooks** provide descriptions of procedures and recommendations for best practices.

**Questionnaires/Glossaries** are copies of questionnaires and glossaries from selected NCES data collections.

**Research and Development (R&D) Reports** are detailed reports of emerging issues, state-of-the-art analytic approaches, and new software applications. The findings reported in developmental work are subject to revision as the work continues and additional data become available.

**Statistical Analysis Reports** present an overview of results from one survey, or from one topic based on analysis across several surveys. The data and findings are presented with commentary to identify substantively and statistically significant results, and their relationship to educational research.

**LIST 1-1-A. — NCES PUBLICATION TYPES (continued)**

**Statistics in Brief** present descriptive data in tabular formats to provide useful information to a broad audience, including members of the general public. They address simple and topical issues and questions. They do not investigate more complex hypotheses, account for inter-relationships among variables, or support causal inferences.

**Technical/Methodological Reports** provide in-depth analyses of analytic methods, survey design, survey procedures, or data quality issues.

**User's Manuals/Data File Documentation** present information on NCES data and related documentation.

**Web Tables** are topical compilations of tabulations from individual data collections. Web tables may be presented as standalone tables in an online table library. Alternatively, they may be released as sets of related tables that are accompanied by introductory text that describes the data source and the topics included in the tables, but no findings.

**Videotapes** are video files of survey findings, case studies or best practices.

**Working Papers** may be used to release the results of survey field tests. These papers have not undergone a rigorous review for consistency with NCES standards.

## **SUBJECT: COMPUTATION AND REPORTING OF RESPONSE RATES**

### **NCES STANDARD: 1-3**

**PURPOSE:** To ensure that the computation of response rates is consistent across NCES surveys.

**KEY TERMS:** base weight, cross-sectional, estimation, frame, item nonresponse, longitudinal, overall unit nonresponse, probability of selection, required response items, response rate, stage of data collection, strata, substitution, survey, total nonresponse, unit nonresponse, and wave.

**STANDARD 1-3-1:** All response rates must be calculated using the sample base weights (inverse of the probability of selection) when weighting is employed. Report the weighted unit response rates for each stage of data collection (e.g., schools, students, teachers, administrators), and for overall unit response rates. Report the range of total response rates for items included in each publication. Also, report item and total response rates for each item with a response rate that falls below 70 percent (see Standard 2-2 and Standard 2-1 for response rate and design issues, see Standard 3-2 on methods of achieving acceptable response rates, and see Standard 7-2 for response rate reporting requirements).

**GUIDELINE 1-3-1A:** Unweighted response rates may be used for monitoring field operations (see Standard 1-3-3).

**STANDARD 1-3-2:** Unit response rates (RRU) are calculated as the ratio of the weighted number of completed interviews (I) to the weighted number of in-scope sample cases (AAPOR 2000). There are a number of different categories of cases that comprise the total number of in-scope cases:

I = weighted number of completed interviews;

R = weighted number of refused interview cases;

O = weighted number of eligible sample units not responding for reasons other than refusal;

NC = weighted number of noncontacted sample units known to be eligible;

U = weighted number of sample units of unknown eligibility, with no interview; and

e = estimated proportion of sample units of unknown eligibility that are eligible.

The unit response rate represents a composite of these components:

$$RRU = \frac{I}{I + R + O + NC + e(U)}$$



*EXAMPLE:* In a school-based survey, the numerator of the unit response rate is the number of responding schools. The denominator includes the number of responding schools plus the summation of the number of schools that refused to participate, the number of eligible schools that were nonrespondents for reasons other than refusal, and an estimate of the number of eligible schools from those with unknown eligibility. Note that in this example, there are no cases reported in the category for the number of eligible schools that were not successfully contacted. This can occur when the only way of determining whether a respondent is eligible is by contacting the respondent.

**STANDARD 1-3-3:** Overall unit response rates for cross-sectional analysis ( $RRO^C$ ) are calculated as the product of two or more unit level response rates when a survey has multiple stages

$$RRO^C = \prod_{i=1}^K RRU_i$$

Where K = the number of stages and C denotes cross-sectional.

There may be instances where fully accurate, current-year frame data are available for all cases at each stage of a survey; in that case, the estimation of overall response rates could be improved. However, in the absence of current-year frame data (as is usually the case), such improvements are not possible and the above formula should be used.

**STANDARD 1-3-4:** Special procedures are needed for longitudinal surveys where previous nonrespondents are eligible for inclusion in subsequent waves. The overall unit response rate used in longitudinal analysis ( $RRO^L$ ) reflects the proportion of all eligible in the sample who participated in all waves in the analysis, multiplied by the product of the response rates for all but the last stage of data collection used in the analysis. In some longitudinal surveys, some of the stages surveyed for the first wave are not resurveyed in subsequent waves, but the unit response rates for the earlier stages are components of the overall unit response rates for subsequent waves.

$$RRO^L = \frac{I^L}{(I^L + R + O + NC + e(U) + W)_{JK}} * \prod_{i=1}^{K-1} RRU_i$$

Where K = the last stage of data collection used in the analysis;

J = the last wave in the analysis;

$I^L$  = the weighted number of responding cases common to all waves in the analysis;

W = respondents to the last wave in the analysis who were nonrespondents in at least one of the preceding waves in the analysis; and

$\prod RRU_i$  = the product of the unit response rates for all but the last stage of data collection.

*EXAMPLE:* For an example in which the respondent in one stage is not resurveyed in subsequent waves, consider a teacher survey where states must be contacted to get a list of schools. This results in a first stage unit response rate for the school listing activity ( $RRU_1$ ). The schools must then be contacted to obtain a list of teachers. This results in a second stage unit response rate for the teacher listing activity ( $RRU_2$ ). Then, once a teacher sample is drawn from the lists, the teacher component of the survey has a third stage unit response rate for the responding teachers ( $RRU_3$ ). The product of the first, second, and third stage unit response rates is the overall response rate for teachers in the first wave of the data collection. To examine changes in job status, teachers are followed up in the second wave in the next school year ( $RRU_4$ ) and in the third wave the following year ( $RRU_5$ ). In an analysis that looks only at the results from the first and third waves, the response rate for teachers is the product of the response rate for the school listing function ( $RRU_1$ ), the response rate for the teacher listing function ( $RRU_2$ ), and the response rate for teachers eligible in both waves of the survey (i.e. the intersection of  $RRU_3$  and  $RRU_5$ ).

**GUIDELINE 1-3-4A:** The product of the unit response rate across all stages and waves used in an analysis is approximately equal to the equation for  $RRO^L$ .

**STANDARD 1-3-5:** Item response rates (RRI) are calculated as the ratio of the number of respondents for whom an in-scope response was obtained ( $I^x$  for item x) to the number of respondents who are asked to answer that item. The number asked to answer an item is the number of unit level respondents (I) minus the number of respondents with a valid skip for item x ( $V^x$ ). When an abbreviated questionnaire is used to convert refusals, the eliminated questions are treated as item nonresponse.

$$RRI^x = \frac{I^x}{I - V^x}$$

In longitudinal analyses, the numerator of an item response rate includes cases that have data available for all waves included in the analysis and the denominator includes the number of respondents eligible to respond in all waves included in the analysis.

In the case of constructed variables, the numerator includes cases that have available data for the full set of items required to construct the variable, and the denominator includes all respondents eligible to respond to all items in the constructed variable.

*EXAMPLE:* In a survey of postsecondary faculty, while all respondents are asked to report the number of hours spent teaching classes per week, only those who report actually teaching classes are asked about the number of hours spent

teaching remedial classes ( $I^x$ ). In this case, the denominator of the item response rate excludes faculty who do not teach classes ( $I - V^x$ ).

In the case of a longitudinal analysis, when all faculty are followed up in the next year to monitor time spent on teaching remedial classes, the numerator of the item response rate for this variable is the number of faculty who responded to this variable in both years. The denominator includes all who were asked in both years.

Faculty job satisfaction is measured using a constructed variable that is the average of 3 separate items—satisfaction with professional development, satisfaction with administration, and satisfaction with teaching assignment. Only full-time faculty members are eligible to answer the satisfaction items. The numerator includes all full-time faculty who answered all 3 satisfaction items and the denominator includes all full-time faculty who completed a faculty questionnaire.

**STANDARD 1-3-6:** Total response rates ( $RRT^x$ ) for specific items are calculated as the product of the overall unit response rate ( $RRO$ ) and the item response rate for item  $x$  ( $RRI^x$ ).

$$RRT^x = RRO * RRI^x$$

*EXAMPLE:* The product of the overall response rate from a faculty survey ( $RRO$ ) and the item response rate for income ( $RRI^x$ ) is the item-specific total response rate for faculty income.

**STANDARD 1-3-7:** To supplement a sample when too few cases are obtained, one or more independent random samples of the population or sampling strata can be drawn and released. When this is done, the released samples must be used in their entirety. In this case, reported response rates must be based on the original and the added sample cases.

*EXAMPLE:* In the event a random supplemental sample is fielded, all cases are included in the response rate—both the original and supplemental cases. Assume that six schools were sampled from a stratum, each with a base weight of 10. Four are respondents and two are nonrespondents. In addition, a supplemental sample of two schools was sampled from the stratum and was fielded in an attempt to compensate for the low initial rate of response. Both of the cases from the supplemental sample are respondents. Taking the combined sample into account, each fielded school has a base weight of 7.5. The response rate then is:

$$((7.5+7.5+7.5+7.5+7.5+7.5) / (7.5+7.5+7.5+7.5+7.5+7.5+7.5+7.5)) \times 100=75\%$$

**STANDARD 1-3-8:** Substitutions may only be done using matched pairs that are selected as part of the initial sample selection. If substitutions are used to supplement a sample, unit response rates must be calculated without the substituted cases included (i.e., only the original cases are used).

*EXAMPLE:* As an example of the case where substitutes are used, but not included in the response rate, assume that two schools were sampled from a stratum. One has a base weight of 20 and the other has a base weight of 10. The first school is a respondent, while the school with a base weight of 10 does not respond. However, a matched pair methodology was used to select two substitutes for each case in the original sample. After fielding the substitutes for the nonrespondent, the first substitute also did not respond, but the second substitute responded. Since we must ignore the substitutes, this response rate is:

$$((20) / (20+10)) * 100 = 66.67 \text{ percent.}$$

In multi-stage designs, where substitution occurs only at the first stage, the first-stage response rate must be computed ignoring the substitutions. Response rates for other sampling stages are then computed as though no substitution occurred (i.e., in subsequent stages, cases from the substituted units are included in the computations). If multiple-stage sample designs use substitution at more than one stage, then the substitutions must be ignored in the computation of response rates at each stage where substitution is used.

#### **REFERENCE**

The American Association for Public Opinion Research. (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Ann Arbor, MI: AAPOR.

## **SUBJECT: CODES AND ABBREVIATIONS**

### **NCES STANDARD: 1-4**

**PURPOSE:** To provide uniform codes, abbreviations, and acronyms for use in NCES data collection and processing that will facilitate the exchange of information and to ensure uniformity in NCES data releases.

**KEY TERMS:** Combined Statistical Area (CSA), Core Based Statistical Area (CBSA), New England City and Town Area (NECTA), Metropolitan Statistical Area, Metropolitan Division, Micropolitan Statistical Area, NCES locale codes, principal city, urban cluster, and urbanized area.

**STANDARD 1-4-1:** The American National Standards Institute (ANSI) Inter-National Committee for Information Technology Standards (INCITS), the Office of Management and Budget, and individual statistical agencies maintain a variety of abbreviations and coding systems. The following standard abbreviations and coding systems, or more current updates, must be used in all NCES data releases:

BSR INCITS 38-200x, Information Technology - Codes for the Identification of the States and Equivalent Areas within the United States, Puerto Rico, and the Insular Areas

These codes establish a structure for the assignment of identifying codes to states and state equivalents of the United States and its insular areas, for the purpose of information interchange among data processing systems.

BSR INCITS 31-200x, Information Technology - Codes for the Identification of Counties and Equivalent Areas of the United States, Puerto Rico, and the Insular Areas

These codes establish a structure for the assignment of identifying data codes to counties and county equivalents of the United States and its insular and associated areas, for the purpose of information interchange among data processing systems.

BSR INCITS 454-200x, Information Technology - Codes for the Identification of Metropolitan and Micropolitan Statistical Areas and Related Statistical Areas of the United States and Puerto Rico

These codes establish a structure for the assignment of data codes to uniquely identify metropolitan and micropolitan statistical areas (generically referred to as “core based statistical areas”) and related statistical areas (i.e., metropolitan divisions, combined statistical areas, New England city and town areas (NECTAs), NECTA divisions, and combined NECTAs of the United States and Puerto Rico, for the purpose of information interchange among data processing systems.

BSR INCITS 455-200x, Information Technology - Codes for the Identification of Congressional Districts and Equivalent Areas of the United States, Puerto Rico, and the Insular Areas

These codes establish a structure for the assignment of identifying data codes to congressional districts of the United States and its insular and associated areas, for the purpose of information interchange among data processing systems

Standard Occupational Codes (SOC) (Bureau of Labor Statistics at: <http://www.bls.gov/soc/>)

**STANDARD 1-4-2:** The North American Industry Classification System (NAICS) was developed jointly by the United States, Canada, and Mexico to provide new comparability in statistics about business activity across North America. NAICS coding has replaced the U.S. Standard Industrial Classification (SIC) system, previously released as FIPS Publication 66. Current NAICS codes may be obtained from the U.S. Census Bureau at [www.census.gov/epcd/www/naics.html](http://www.census.gov/epcd/www/naics.html).

**STANDARD 1-4-3:** The following IES-sponsored coding systems must be used, where applicable:

1. The Classification of Instructional Programs (CIP), which is the accepted federal government statistical standard on instructional program classifications at the post-secondary level. (See Classification of Instructional Programs 2010 at <http://nces.ed.gov/ipeds/cipcode/Default.aspx?y=55>)
2. The College Course Map (CCM), which is a classification scheme for college courses offered in the United States. (See Adelman, C. 1995. *The New College Course Map and Transcript Files*. Washington, DC: U.S. Department of Education, National Institute on Postsecondary Education, Libraries, and Lifelong Learning.) (<http://www.ed.gov/rschstat/research/pubs/empircurr/cmmtax2003.pdf>)
3. The Secondary School Taxonomy, which is a classification scheme for high school courses offered in the U.S. (See Bradby, D. and Hoachlander, G. (1999). *1998 Revision of the Secondary School Taxonomy* (NCES 1999-06). U.S. Department of Education. Washington, D.C.: National Center for Education Statistics Working Paper <http://nces.ed.gov/pubs99/199906.pdf>; Bradby, D. (2007). *The 2007 Revision of the Career/Technical Education Portion of the Secondary School Taxonomy* (NCES 2008-030). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2008030>); and Nord, C., Hicks, L., Hoover, K., Jones, M., Lin, A., Lyons, M., Perkins, R., Roey, S., Rust, K., and Sickles, D. (2011). *The 2009 High School Transcript Study User's Guide* (NCES 2011-465). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. *Appendix I. 2009 High School*

*Transcript Study Classification of Secondary School Courses*  
<http://nces.ed.gov/nationsreportcard/pdf/studies/2011465.pdf>

4. NCES Locale Codes, which are a means of identifying the urbanicity or, conversely, rural status of specific geographic areas, include City Large, City Midsize, City Small, Suburb Large, Suburb Midsize, Suburb Small, Town, Fringe, Town Distant, Town remote, Rural Fringe, Rural Distant, Rural Remote (National Center for Education Statistics at: [http://nces.ed.gov/ccd/rural\\_locales.asp](http://nces.ed.gov/ccd/rural_locales.asp)). These designations are based on population sizes and distance from principal cities, urbanized areas, and urban clusters.

**STANDARD 1-4-4:** Where appropriate, the IES *Publication Guide* must be utilized (<http://nces.ed.gov/statprog/styleguide/pdf/styleguide.pdf>), along with the *United States Government Printing Office Style Manual (GPO Style Manual)* ([www.access.gpo.gov](http://www.access.gpo.gov)). <http://www.gpo.gov/fdsys/pkg/GPO-STYLEMANUAL-2008/pdf/GPO-STYLEMANUAL-2008.pdf> Official national, State and international abbreviations are listed in the *Style Manual* of the U.S. Government Printing Office (GPO). These abbreviations must be used where appropriate in NCES publications.

**STANDARD 1-4-5:** When setting up a manual coding process to convert text to codes, create a quality assurance process that verifies at least a sample of the coding to determine if a specific level of quality is being maintained.

## **SUBJECT: DEFINING RACE AND ETHNICITY DATA**

### **NCES STANDARD: 1-5**

**PURPOSE:** To provide common language to promote uniformity and comparability for the collection and reporting of data on race and ethnicity. This standard is in compliance with the definitions and procedures included in the 1997 revision of the OMB Statistical Policy Directive No. 15 and the U.S. Department of Education Final Guidance on Maintaining, Collecting, and Reporting Racial and Ethnic Data to the Department of Education (Federal Register, Vol. 72, No. 202, 10/19/2007).

**KEY TERMS:** American Indian or Alaska Native, Asian, Black or African American, confidentiality, edit, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, public-use data file, White, and survey.

**STANDARD 1-5-1:** NCES will follow OMB and Department of Education guidelines for definitions of ethnicity and race.

Ethnicity is based on the following categorization:

Hispanic or Latino: A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term "Spanish origin" can be used in addition to "Hispanic or Latino."

Race is based in the following five categorizations:

American Indian or Alaska Native: A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

Asian: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

Black or African American: A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American."

Native Hawaiian or Other Pacific Islander: A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.



**STANDARD 1-5-2:** NCES will follow OMB and Department of Education guidelines on the use of a two-question format to collect data on ethnicity and race. The ethnicity question must come first, followed by the question on race. The race question must ask respondents to select one or more of the five racial groups.

The ethnicity question is:

What is this person's ethnicity?

Hispanic or Latino

Not Hispanic or Latino

The race question is:

What is this person's race? Mark one or more races to indicate what this person considers himself/herself to be.

American Indian or Alaska Native

Asian

Black or African American

Native Hawaiian or Other Pacific Islander

White

**GUIDELINE 1-5-2A:** Generally, data collections will only collect the categories that are listed above in the sample questions. These produce seven reporting categories: Hispanic/Latino, American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, and Two or more races. These seven are the minimum categories established by OMB, and are the standard adopted by the Department of Education. However, in the case of data collected by NCES from individual respondents, where there is a research need, and where sample size is sufficient, NCES may elect to expand the ethnicity question to a format to add subcategories, as long as they can be recombined into the basic seven categories.

*EXAMPLE 1:*

Is this person Hispanic or Latino?

No, not Hispanic/Latino

Yes, Mexican, Mexican American, Chicano

Yes, Puerto Rican

Yes, Cuban

Yes, other Spanish/Latino (specify \_\_\_\_\_)

*EXAMPLE 2:*

Is this person of Asian descent?

No

Yes, Japanese

Yes, Chinese

Yes, Korean

Yes, Pakistani

Yes, Indian

Yes, other (specify \_\_\_\_\_)

In all such cases, the greater detail must be collapsible to the minimum standard.

**GUIDELINE 1-5-2A:** In the case of data collected by NCES from individual respondents, where sample size is sufficient, the “Two or more” races category should be expanded to represent specific race combinations.

**STANDARD 1-5-3:** Aggregated data reported by education institutions about students, faculty, or staff will follow the Department of Education October 2007 final guidance on ethnicity and race data (<http://www.gpo.gov/fdsys/pkg/FR-2007-10-19/pdf/E7-20613.pdf>), with implementation scheduled for data collections for the 2010-11 school year. Education institutions were advised to collect information from individuals using the two-question format (see Standard 1-5-2). They were advised to keep full detail on what individuals reported about their own race and ethnicity in their record systems for the length of time specified in the instructions for the data collection. Per the Department guidance, the institutions will report aggregate data on ethnicity and race of their students, faculty, or staff to the Department of Education using the following categories:

- Hispanic or Latino, of any race
- American Indian or Alaska Native, not Hispanic or Latino
- Asian, not Hispanic or Latino
- Black, not Hispanic or Latino
- Native Hawaiian or Other Pacific Islander, not Hispanic or Latino
- White, not Hispanic or Latino
- Two or more races, not Hispanic or Latino

Postsecondary institutions may use two additional categories in their reporting of aggregated race and ethnicity data:

- Nonresident alien
- Not reported (race and ethnicity unknown)

**STANDARD 1-5-4:** When reporting data on race and ethnicity in government publications, every effort must be made to use at least the minimal reporting categories, described below, whenever possible. In the case of data collected by NCES from individual respondents, more categories of race combinations or detailed subgroups may be used when there are enough cases to support finer detail and there is justification for more categories (see Guideline 1-5-2A and Guideline 1-5-2B). However, if there are not enough cases in any individual category of race or Hispanic ethnicity, the data for that category and for the next smallest category must be included in the total but not shown separately, and footnoted as such. Alternatively, if several categories cannot be shown, the combined categories must be reported as an "other" category, and footnoted to describe the exact components.

The following are the desired and minimal reporting categories for race and ethnicity in government publications. The decision rules for each combination of race and ethnicity are shown in italics:

Hispanic/Latino

*(This category includes individuals of any race who identify as Hispanic or Latino)*

American Indian or Alaska Native

*(This category includes only persons who reported American Indian or Alaska Native as their sole race and did not report Hispanic/Latino ethnicity.)*

Asian

*(This category includes only persons who reported Asian as their sole race and did not report Hispanic/Latino ethnicity.)*

Black or African American

*(This category includes only persons who reported Black or African American as their sole race and did not report Hispanic/Latino ethnicity.)*

Native Hawaiian or Other Pacific Islander

*(This category includes only persons who reported Native Hawaiian or Other Pacific Islander as their sole race and did not report Hispanic/Latino ethnicity.)*

White

*(This category includes only persons who reported White as their sole race and did not report Hispanic/Latino ethnicity.)*

Two or more races

*(This category includes any combination of two or more races and not Hispanic/Latino ethnicity.)*

**GUIDELINE 1-5-4A:** In the text, the names for the groups should be capitalized, per the U.S. Government Printing Office (e.g., White, Black, Asian, etc.).

**GUIDELINE 1-5-4B:** When the publication contains substantial text, the category names may be abbreviated after the first presentation of the categories. The authors should introduce the shortened version of the category label by saying that the two are used interchangeably in the text.

The following abbreviated names are suggested for use in text or in tables and figures:

American Indian (instead of American Indian or Alaska Native)

Black (instead of Black or African American)

Pacific Islander (instead of Native Hawaiian or Other Pacific Islander)

Hispanic (instead of Hispanic or Latino)

A footnote is needed to describe these "abbreviations" as follows:

American Indian includes Alaska Native, Black includes African American, Pacific Islander includes Native Hawaiian, and Hispanic includes Latino.

**STANDARD 1-5-5:** Full detail on race and ethnicity as reported to NCES in data collected from individual respondents must be maintained on restricted-access data files and on public-use data files, within the constraints imposed by relevant confidentiality laws and administrative policies (see Standard 4-2).

**GUIDELINE 1-5-5A:** Survey documentation should describe how race and ethnicity questions were asked. If other than the OMB standard two question format and seven data categories were used, documentation must include collection format, how imputation and edits were accomplished, and what decisions were made to create aggregation categories.

**STANDARD 1-5-6:** The changes resulting from the implementation of the 1997 revision of the OMB Statistical Policy Directive No. 15 ([http://www.whitehouse.gov/omb/fedreg\\_1997standards/](http://www.whitehouse.gov/omb/fedreg_1997standards/)) and the 2007 U.S. Department of Education Final Guidance on Maintaining, Collecting, and Reporting Racial and Ethnic Data to the Department of Education involve key variables and survey procedures for this data series. Consistent with NCES Standard 2-5, a plan must be developed that describes adjustment methods, such as crosswalks and bridge studies that will be used to preserve trend analyses.