DRAFT

NCES STATISTICAL STANDARDS

May 2002

DRAFT

TABLE OF CONTENTS

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
1-6 Discretionary Grant Descriptions

PLANNING AND DESIGN OF SURVEYS
2-1 Design of Surveys
2-2 Survey Response Rate Parameters
2-3 Developing RFPs for Surveys
2-4 Field Testing Survey Systems
2-5 Maintaining Data Series
2-6 Educational Testing

COLLECTION OF DATA
3-1 Coverage for Frames and Samples
3-2 Achieving Acceptable Response Rates
3-3 Monitoring and Documenting Survey Contracts
3-4 Documenting a Survey System

PROCESSING AND EDITING OF DATA
4-1 Imputation of Item Nonresponse
4-2 Maintaining Confidentiality
4-3 Evaluation of Surveys
4-4 Nonresponse Bias Analysis

ANALYSIS OF DATA/PRODUCTION OF ESTIMATES OR PROJECTIONS
5-1 Statistical Analysis, Inference, and Comparisons
5-2 Variance Estimation
5-3 Rounding
5-4 Tabular and Graphic Presentations of Data

ESTABLISHMENT OF REVIEW PROCEDURES
6-1 Review of Reports and Data Products

DISSEMINATION OF DATA
7-1 Machine Readable Products
7-2 Survey Documentation in Reports
7-3 Release and Dissemination of Reports and Data Products

APPENDIX A
Race and Ethnicity

APPENDIX B
Imputation

APPENDIX C
Tabular Guidelines

APPENDIX D
Survey Titles

INTRODUCTION

Purpose of Statistical Standards

This Web site contains the draft revised statistical standards and guidelines for the National Center for Education Statistics (NCES). It is intended for NCES staff and contractors to guide them in their data collection, analysis, and dissemination activities. These standards and guidelines are also intended to present a clear statement for data users regarding how data should be collected in NCES surveys, and the limits of acceptable applications and use.

Background of Statistical Standards

NCES first adopted written statistical standards in the spring of 1987. These standards were the result of a multi-year evaluation and planning process, that included a recommendation for the development of statistical standards from the Committee on National Statistics at the National Academy of Science. With that recommendation, a statistical standards program was initiated at NCES in 1985. Using the Energy Information Administration's Standards Manual and the Census Bureau's technical paper on "Standards for Discussion and Presentation of Errors in Survey and Census Data," NCES staff, in consultation with outside experts developed the 1987 version of NCES statistical standards.

With the adoption of this first set of standards, the Agency Director called for a formal evaluation to start the following fall, to insure that the standards were fully implemented and to identify any difficulties with the standards. In 1989, the Center undertook a full-scale revision of the 1987 standards. The revisions were developed by NCES staff, and reflected their first-hand experiences in using the 1987 standards. After multiple reviews of interim drafts by NCES staff and the NCES Advisory Council of Education Statistics, NCES Senior Staff accepted the revised standards in the spring of 1992.

At the June 1992 release of the *NCES Statistical Standards* report, the Acting Commissioner summarized the standards in the following statement:

> They: (1) codify how we expect to behave professionally, (2) indicate the basis on which we expect to be judged by our peers in the statistical community, (3) represent the quality we expect in any of our efforts or those of our contractors and grantees, (4) provide a means to assure consistency among the studies the Center conducts, and (5) document for users, the methods and principles the Center employs in the collection of data.

The Acting Commissioner also reiterated the Center's commitment to periodic evaluations of the implementation of the standards and to a periodic review of the standards' operational feasibility.

The current revision process began in the summer of 1999 with a review of existing standards from a number of national and international statistical policy agencies and committees and from other international and national statistical agencies. At the same time the 1992 *NCES Statistical Standards* were made available on the Web, and NCES staff were given a 30-day period to submit comments concerning potential revisions and additions to the NCES standards. Following these activities an agency-wide Steering Committee was formed to work on the standards revision process. The Steering Committee formed 15 Working Groups that comprised more than one-half of the NCES staff to work on the set of topics identified in the 1999 reviews.

Each Working Group drafted their assigned standards; each of which underwent a multi-step review process. Following a 30-day NCES staff comment period, the working group members made revisions, the Steering Committee reviewed the drafts and submitted them to Senior Staff. The drafts were then reviewed by Senior staff, modified as necessary, and then shared with a group of 40 to 50 representatives of the contractors who work with NCES on data collection, analysis, and dissemination. Additional revisions were incorporated following the input from this broad group. The draft standards on this Web site are the result of the efforts of the many persons who participated in this multi-stage review process.

During the recent NCES standards revision, the Office of Management and Budget (OMB) issued government-wide guidelines for ensuring and maximizing the quality of information disseminated by Federal agencies. The OMB guidelines direct all agencies covered by the Paperwork Reduction Act (44 U.S.C. chapter 35) to develop and implement procedures for reviewing and substantiating the quality of information disseminated by the agency. In order to meet these goals, each agency is required to develop and promulgate quality guidelines. Quality includes the utility, objectivity and integrity of information.

In response to the OMB guidelines, the federal statistical agencies collaborated to identify a set of activities that are essential to maintaining the quality and credibility of statistical data. The NCES draft revised standards are organized around the shared framework for federal statistical agencies. NCES remains committed to the principles outlined by the 1992 NCES Acting Commissioner; what is more, these principles are reaffirmed in the OMB call for data quality guidelines.

We welcome your comments on the NCES draft statistical standards. Please submit them to the following address by May 30, 2002: Marilyn.McMillen@ed.gov.

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

1-6 Discretionary Grant Descriptions

SUBJECT: INITIAL PLANNING OF SURVEYS

NCES STANDARD: 1-1-02

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.

DEFINITIONS: An individual **survey** is driven by one data collection form, such as the Private School Survey or the Academic Library Survey.
A **survey system** is a set of individual surveys that are interrelated components of a data collection, such as the Schools and Staffing Survey or the Integrated Postsecondary Education Data System.
At a minimum, **key items** include items for which aggregate estimates are commonly published by NCES. They include, but are not restricted to, variables most commonly used in table row stubs. Key items include important analytic composites and other policy-relevant variables.
The **minimum substantively significant effect (MSSE)** is the smallest difference considered to be important for the analysis of key variables. The minimum substantively significant effect is determined during the design phase. For example, the planning document should provide the minimum change in key variables that the survey should be able to detect for a specified population domain, or subdomain of analytic interest. The MSSE should be based on a broad knowledge of the field, related theories, and supporting literature.
Effect size refers to the standardized magnitude of an effect or the degree of departure from the null hypothesis. For example, the effect size may measure the amount of change over time, or the size of the difference between two populations, divided by the appropriate population standard deviation. Some programs will only report results which are substantively significant.
The **power** of a test is defined as the probability of rejecting the null hypothesis when it is actually false ($1- \beta$). In the design phase, given the MSSE, design effect, an alpha of 0.05 and a desired power of 0.80, this information can be used to determine the appropriate actual sample size to meet these specifications.
In the design phase, the **effective sample size** is the sample size under a simple random sample design that is equivalent to the actual sample under the complex sample design. In the case of complex sample designs, the actual sample size is determined by multiplying the effective sample size by the anticipated design effect.
The **design effect** is the ratio of the true variance of a statistic (taking the complex sample design into account) to the variance of the statistic for a simple random sample with the same number of cases. Design effects differ for different subgroups and different statistics; no single design effect is universally applicable to any given survey or analysis.

STANDARD: The initial plan for developing a survey or survey system must include the justification for the study and must describe the survey methodology. The initial planning document will be presented to the OC/ODC for review and a decision on whether to proceed with the design phase, prior to an OMB fiscal year budget request. 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 should 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.
3. A preliminary survey design that discusses the proposed target population, response rate goals (see Standard 2-3-02), sample design, sample size determination based on power analyses for key items, data collection methods, and methodological issues.
4. A preliminary analysis plan that identifies analysis issues, objectives, key items, minimum substantively significant effect sizes, and proposed statistical techniques.
5. A list of data items that will be maintained over time as part of an NCES data series, including the justification for each item.
6. A preliminary time schedule that accounts for the complete survey cycle from planning to data release.
7. A preliminary publication and dissemination plan that identifies proposed major publications and their target audiences (see Standard 1-2-02).
8. A preliminary survey evaluation plan that identifies the proposed analyses necessary for data users to understand the quality and limitations of the survey.
9. An internal cost estimate that reflects all of the above items.

SUBJECT: PUBLICATION AND PRODUCT PLANNING

NCES STANDARD: 1-2-02

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

DEFINITIONS: NCES uses the following three categories for publication planning:

Mandatory (Priority 1): This category includes a limited number of high profile publications that the agency is committed to release in a specific month.
Required (Priority 2): This category includes publications 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.
Projected (Priority 3): This category includes publications that may be completed but for which there is no predetermined expectation about a release date. Publications over which the agency has less control over timing and staff-initiated in-depth reports are included.

STANDARD 1: All NCES publications and data products must be included in the annual NCES publication plan. (See list A for a description of NCES product types.)

> **GUIDELINE A:** A publication should be added to the publication plan by the time it is signed off for Division review.
>
> **GUIDELINE B:** Project Directors should update changes in the NCES publication plan on an as needed basis.

STANDARD 2: All proposed publications and data products must receive Program Director and Associate Commissioner approval before inclusion in the NCES publication plan.

> **GUIDELINE:** Bimonthly meetings between Office of the Commissioner (OC) staff and the Associate Commissioners and their division staff should be held to review progress on the publication plan.

STANDARD 3: All mandatory and required publications must have firm target delivery dates to the OC Publication Database Coordinator.

GUIDELINE A: The date for printed release is approximately 6 to 8 weeks after the final post-adjudication sign-off by the Chief Statistician.

GUIDELINE B: For Web release only publications, the release date can be simultaneous with the post-adjudication sign-off, but should occur within 1 week.

GUIDELINE C: For early Web release publications, the pdf file will be posted when the publication is sent to the U.S. Government Printing Office (GPO) to ensure that the Media and Information Service (MIS) edits are included in the Web release.

STANDARD 4: For printed release publications, the reports will not be sent to GPO until the pdf file and the Web publishing form are submitted to the Webmaster.

STANDARD 5: 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 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.

E.D. TABs are a collection of tables, presented with minimal analyses. The purpose of an E.D. TAB is to make tabular data available quickly.

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.

Issue Briefs/NAEPfacts are a two-to-four page summary of a particular topic. A limited number of tables and charts are presented with descriptive text intended to provide a quick view of a current topic.

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 A.— NCES PUBLICATION TYPES (continued)

Statistics in Brief are a short, focused analysis of a specific topic. Generally 4-to-15 pages in length, these reports are designed to provide data on policy-relevant topics.

Technical/Methodological Reports are an in-depth analysis 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.

Videotapes are VHS formatted tapes of survey findings, case studies or best practices.

Working Papers provide preliminary analysis of substantive, technical, and methodological issues. They are works in progress that are presented to promote the sharing of valuable work experience and knowledge. These papers have not undergone a rigorous review for consistency with NCES standards.

SUBJECT: COMPUTATION AND REPORTING OF RESPONSE RATES

NCES STANDARD: 1-3-02

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

DEFINITIONS: Unit nonresponse occurs when a respondent fails to respond to all critical data items(i.e., fill out or return a data collection instrument).
Critical data items include the minimum set of items required for a case to be considered a respondent.
Overall unit nonresponse reflects a combination of unit nonresponse across two or more levels of data collection, where participation at the second stage of data collection is conditional upon participation in the first stage of data collection.
Item nonresponse occurs when a respondent fails to respond to one or more relevant item(s) on a survey.
Total nonresponse reflects a combination of the overall unit nonresponse and item nonresponse.
Stage of data collection includes any stage or step in the sample identification and data collection process in which data are collected from the identified sample unit. This includes information required to proceed to the next stage of sample selection or data collection (e.g. school district permission for schools to participate or schools providing lists of teachers for sample selection of teachers).
A **wave** is a round of data collection in a longitudinal survey (e.g., the base year and each successive follow-up are each waves of data collection).

STANDARD 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 overall unit response rates (see Standard 7-2-02). Report the range of total response rates for items included in each publication. Also, report item and total response rates when the item response rates falls below 70 percent.

> **GUIDELINE:** Unweighted response rates may be used for monitoring field operations.

STANDARD 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:

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

As an 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 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.

STANDARD 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. The unit response rates for the earlier stages are components of the overall unit response rates for subsequent waves. Note in longitudinal analyses, at least the last stage of a data collection has multiple 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.

For example, for stages 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 and in the third wave the following year. 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.

> **GUIDELINE:** 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 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 in the constructed variable, and the denominator includes all respondents eligible to respond to all items in the constructed variable.

For 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 three satisfaction items and the denominator includes all full-time faculty who completed a faculty questionnaire.

STANDARD 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).

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

As an 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 7: Substitutions occur when cases are used to replace nonrespondents or ineligibles. If the substitutions are done, they must use matched pairs that are selected as part of the initial sample selection. In this case, unit response rates must be calculated both with and without the substituted cases included. When the substituted cases are included, the rate is calculated using both the initial and substituted cases and the base weight is split between the initial nonresponding case and the substitute case.

In multiple stage sample designs, where substitution occurs only at the first stage, the first stage response rate must be computed using the appropriate substitution option. Response rates for other sampling stages must be computed as though no substitution has occurred. If multiple stage sample designs use substitution at more than one stage, then the appropriate response rate must be used at each stage where substitution is used.

To avoid substitutions, an independent random sample of the population or sampling strata can be released and used in its entirety. In this case, reported response rates must be based on the original and the added sample cases.

As an example of the case where substitutes are not included in the response rate, assume that two schools were sampled from a stratum. One has a basic weight of 20 and the other has a basic 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 are ignoring the substitutes in this response rate, it is:

((20)/(20+10) x100= 66.67 percent.

Following the same example, with the matched pair substitutes included in the response rate, recall that the original respondent has a basic weight of 20. In this case, the basic weight of 10 for the original nonrespondent is equally split between the original case and the two fielded substitutes (i.e., each case has a basic weight of 3.33). The response rate then is:

((20+3.33)/(20+3.33+3.33+3.33))x100= 77.77 percent.

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 basic 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))x100= 75 percent.

REFERENCE

The American Association for Public Opinion Research. (2000). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Ann Arbor, MI: AAPOR.

SUBJECT: CODES AND ABBREVIATIONS

NCES STANDARD: 1-4-02

PURPOSE: To provide uniform codes, abbreviations and acronyms for use in NCES data collection and processing in order to facilitate the exchange of information and to ensure uniformity in NCES data releases. This standard requires adherence to the procedures described below.

Definitions: An area qualifies for recognition as a **Metropolitan Statistical Area** (**MSA)** in one of two ways: (1) if it includes a city of at least 50,000 population, or (2) if it includes a Census Bureau-defined urbanized area (of at least 50,000 population) with a total metropolitan population of at least 100,000 (75,000 in New England). In addition to the county(ies) containing the main city or urbanized area, an MSA may include additional counties that have strong economic and social ties to the central county(ies) and meet specified requirements of metropolitan character. The ties are determined chiefly by census data on commuting to work. A metropolitan statistical area may contain more than one city of 50,000 population and may cross State lines.
An area that meets theses requirements for recognition as an MSA but also has a total population of one million or more may be recognized as a **Consolidated Metropolitan Statistical Area** (**CMSA)** if: (1) separate component areas can be identified within the entire area by meeting specified statistical criteria, and (2) local opinion indicates there in support for the component areas. If recognized, the component areas are designated **Primary Metropolitan Statistical Areas (PMSAs)**, and the entire areas becomes a CMSA. If no PMSAs are recognized, the entire are is designated an MSA
New England County Metropolitan Areas (NECMSs) are county-based alternatives to the city- and town-based metropolitan areas in New England. The NECMA for an MA or CMSA includes: (1) the county containing the first-named city in that MSA/CMSA title (this county may include the first-named cities of other MAs/CMSAs), and (2) each additional county having at least half its population in the MA/CMSA(s) whose first-named cities are in the county identified in step 1. NECMAs are not defined for individual PMSAs.

STANDARD 1: The National Institute of Standards maintains a variety of abbreviations under the Federal Information Processing Guidelines (FIPS PUBS). (See www.itl.nist.gov/fipspubs/index.htm for the most recent versions of these standards.) The following FIPS standards, or more current updates, must be used in all NCES data releases:

FIPS PUB NUMBERS

5-2 States and Outlying Areas of the U.S.
6-4 County and County Equivalent of the States of the U.S. and D.C.
8-6 Metropolitan Areas, including Metropolitan Statistical Areas (MSAs), Consolidated Metropolitan Statistical Areas (CMSAs), Primary

Metropolitan Statistical Areas (PMSAs), and related units called New England County Metropolitan Areas (NECMAs)

9-1 Congressional Districts of the U.S.

92 Standard Occupational Codes (SOC)

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. NAICS codes must now be used instead of SIC codes for industry coding. (See Standard 2-5-02 for guidance on maintaining comparability when adopting NAICS coding for existing data series.) Current NAICS codes may be obtained from the U.S. Census Bureau at www.census.gov/epcd/www/naics.html.

STANDARD 2: In addition, the following OERI-sponsored coding systems must be used, where applicable:

- 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 (CIP-2000 Edition)*. 2002. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics, NCES-2002-165. See nces.ed.gov/ipeds/pdf/webBase2000/cipman.pdf for an electronic version of this publication.
- 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.)
- 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.* Washington, D.C.: U.S. Department of Education, National Center for Education Statistics Working Paper, NCES-1999-06.)

STANDARD 3: Where appropriate, the OERI Publication Guide must be utilized. Official national, State and international abbreviations are listed on pages 147-170 of the *Style Manual*, 2000 edition, of the U.S. Government Printing Office (GPO). These abbreviations must be used where appropriate in NCES publications. The current version of the *Style Manual* may be obtained at the GPO website: www.access.gpo.gov.

SUBJECT: DEFINING RACE AND ETHNICITY DATA

NCES STANDARD: 1-5-02

PURPOSE: To provide common language to promote uniformity and comparability for 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.

DEFINITIONS: 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."
Hispanic or Latino: A person of Cuban, Mexican, Puerto Rican, Cuban, 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."
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: Pending further government-wide research on the best practices for collecting information about race and ethnicity, NCES will follow OMB guidelines on the use of a two-question format—except under rare circumstances in which a one-question format is justified on the basis of research or other documentation.

With the two-question format, the ethnicity question must come first, followed by the question on race. The race question must allow respondents to choose one or more of the listed categories. Taken together, the Hispanic/Latino category from the ethnicity question and the 5 race categories result in 64 possible combinations of race and Hispanic ethnicity.[1]

[1] See appendix A for a full list of the 64 categories.

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.

White
Black or African American
Asian
American Indian or Alaska Native
Native Hawaiian or Other Pacific Islander[2]

GUIDELINE: Generally, data collections will only collect the minimum information (the categories listed above). For example, in the case of the ethnicity question, it will be a simple two-category question. However, there are surveys that NCES conducts for which there may be interest and large enough sample size to expand the ethnicity question to a format similar to the 2000 Decennial Census question to ask about specific Hispanic or Latino ethnicities. For example:

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 _____)

Similarly, if there is a need for more detail and the sample size can support it, an expanded list of races may be used. However, the greater detail must be collapsible to the minimum standard.

STANDARD 2: Data that are aggregated from institutional records must be reported in the following way:

Waiting for new categories for aggregation of administrative record data, to be implemented in Fall 2004.

Not Hispanic or Latino:

White (only)
Black or African American (only)
Asian (only)

[2] The categories are presented in order of numerical frequency in the population, rather than alphabetically. Previous research studies have found that following alphabetical order in the question categories. That is, first category "American Indian or Alaska Native" results in substantial over reporting of this category.

Native Hawaiian or Other Pacific Islander (only)
American Indian or Alaska Native (only)
White/Black or African American (only these two races reported)
White/Asian (only these two races reported)
White/American Indian (only these two races reported)
Black/American Indian (only these two races reported)

Other combinations of race (any other combinations of two or more races, including combinations that include the two-race combinations listed above along with one or more other races)

No race specified
Hispanic or Latino—
White
Black or African American
Asian
Native Hawaiian or Other Pacific Islander
American Indian or Alaska Native
More than one race
No race specified

STANDARD 3: Full detail on race and ethnicity as reported by individuals or from administrative data must be maintained on restricted-access data files and on public-use data files, within the constraints imposed by relevant confidentiality laws.

GUIDELINE: Survey documentation should describe how race and ethnicity questions were asked, how imputation and edits were accomplished, and what decisions were made to create aggregation categories.

STANDARD 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. More categories should be used when there are enough cases to support finer detail. 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 minimal reporting categories for race and ethnicity in government publications. The decision rules for each combination of race and ethnicity are shown in italics:

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

Asian, not Hispanic or Latino
(This category includes only persons who reported Asian as their sole race, but did not report Hispanic ethnicity.)

Black, not Hispanic or Latino
(This category includes only persons who reported Black as their sole race, but did not report Hispanic ethnicity.)

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

White, not Hispanic or Latino
(This category includes only persons who reported White as their sole race, but did not report Hispanic ethnicity.)

More than one race, not Hispanic or Latino
(This category includes any combination of more than one race and not Hispanic or Latino ethnicity or Hispanic or Latino ethnicity not reported.)

Hispanic or Latino, regardless of race
(This category includes Hispanic or Latino ethnicity and any combination of race.)

GUIDELINE A: 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 B: 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. Race categories exclude Hispanic origin unless specified.

SUBJECT: DISCRETIONARY GRANT DESCRIPTIONS

NCES STANDARD: 1-6-02

PURPOSE: To assist NCES staff in the preparation of high quality discretionary grant descriptions. The description should include the information required to allow an applicant to submit a proposal that demonstrates technical and managerial competence sufficient to successfully complete a project. Each grant description should also include the selection criteria to be used in accordance with federal and Department of Education regulations.

STANDARD 1: Grant descriptions must be written in compliance with guidelines established in the Education Department General Administrative Regulations (EDGAR).

> **GUIDELINE:** The Grants Policy and Oversight Staff (GPOS) in the Office of the Chief Financial Officer can provide expertise and guidance in the development of the grant description and application process.

STANDARD 2: The team leader for the grant is responsible for (1) monitoring technical performance of the grantee, and (2) providing technical advice and recommendations to both the grantee and the grant team.

> **GUIDELINE:** Within NCES, the staff member who develops the application package and related documents should be designated as grant team leader. The individual who develops the application package should have completed required courses for administering the grants process. Minimally, the grant team leader should be included in the development process, and should be familiar with the grant requirements and expectations.

STANDARD 3: The grant process must include the following four activities:

1. Submit the Application Notice for publication in the *Federal Register*. This invites applications for a competition, gives basic program and fiscal information, and informs potential applicants when and where they may obtain applications.
2. Prepare the Grant Application Package, which must include the standard information for all discretionary grant programs to comply with the policies and regulations of the Department and the Office of Management and Budget (OMB). In addition, a clear, precise, and accurate description of the problem to be addressed and the expected activities, services, or products, and level of effort to be delivered under the grant. This includes technical, statistical, managerial, and product objectives.

3. Provide Guidance for Completing Applications, which describes the required elements of a grant application package, including cover sheet, narrative of proposed activities and budget for these activities, and assurances of compliance with requirements imposed by the U.S. Secretary of Education.
4. Develop an Application Technical Review Plan that describes how applications for funding should be evaluated. This plan should include procedures for evaluating applications, including review panels, criteria for selecting reviewers, technical review forms, method for ranking applications for funding, and basis for recommending applications for funding.

GUIDELINE: The application package should provide the applicant with a statement of statistical, temporal, and reporting guidelines for design, implementation, and analysis, as appropriate. Managerial guidelines should delineate those to be performed by the grantee and those to be performed by NCES. The products (e.g., analysis plans, final reports) should be termed "deliverables" and guidelines for due dates should be provided.

PLANNING AND DESIGN OF SURVEYS

2-1 Design of Surveys

2-2 Survey Response Rate Parameters

2-3 Developing a Request for Proposal (RFP) for Surveys

2-4 Field Testing Survey Systems

2-5 Maintaining Data Series

2-6 Educational Testing

SUBJECT: DESIGN OF SURVEYS

NCES STANDARD: 2-1-02

PURPOSE: To identify the necessary survey design components for conducting a data collection.

STANDARD: A technical document that delineates the basic design of a survey or survey system must be developed prior to the initiation of a data collection. The document must address: the objectives of the survey as indicated in the initial planning document; how the survey will be designed and conducted; and the personnel resources, funds, and time needed to achieve high data quality. To meet this standard, the survey design plan must include the following:

1. A detailed discussion of the goals and objectives of the survey or survey system, including the information needs that will be met, content areas included, target population(s), and analytic goals.
2. A discussion of the sample design that describes how it will yield the data required to meet the objectives of the survey. The discussion must include the following: identification of the sampling frame and the adequacy of the frame (see Standard 3-1-02); sampling strata; power analyses to determine sample sizes for key variables by reporting domains, sample size by stratum; the known probability of selection; expected yield by stratum; estimated efficiency of sample design; weighting plan; variance estimation techniques appropriate to the survey design; and expected accuracy of key estimates.
3. A listing of all survey data items, including time series data items, how each item can best be measured (e.g., through questionnaires, tests), and reasonable evidence that these items are valid and can be measured both accurately and reliably.
4. An analysis plan providing evidence that the basic information needs which justify the study can be met through the proposed data collection. The plan must demonstrate how the proposed sample, the survey items, and the measurement methods are related to the objectives of the survey.
5. The anticipated data collection procedures including: timing of data collection; primary mode of collection; and methods for achieving acceptable response rates (see Standard 3-2-02).
6. A plan for preserving the confidentiality of the data during collection, processing, and analysis, if individually identifiable data will be collected. An analysis plan for disclosure risk control is also required to prepare a public use data file (see Standard 4-2-02).
7. An outline of a plan for quality assurance during each phase of the survey process that will permit monitoring and assessing the performance during implementation. The plan must include contingencies to modify the survey procedures, if design

parameters appear unlikely to meet expectations (for example, low response rates). A plan for field testing the survey or survey system (see Standard 2-4-02).

8. An outline of the general parameters for evaluating survey procedures and results (see Standard 4-3-02).
9. General specifications for an internal project management system that identifies critical activities and key milestones of the survey that will be monitored, and the time relationships among them.
10. An Independent Government Cost Estimate (IGCE) for the entire study, including, for example, the pilot test, the main study, file preparation and documentation, disclosure risk analysis, the survey evaluation, and analysis and reporting.

SUBJECT: SURVEY RESPONSE RATE PARAMETERS

NCES STANDARD: 2-2-02

PURPOSE: High survey response rates help to ensure that survey results are representative of the target population. Surveys conducted by or for NCES must be designed and executed to meet high rates of response.

DEFINITIONS: At a minimum, **key items** include items for which aggregate estimates are commonly published by NCES. They include, but are not restricted to, variables most commonly used in table row stubs. Key items include important analytic composites and other policy-relevant variables.

Nonresponse bias occurs when the expected observed value deviates from the population parameter due to differences between respondents and nonrespondents. Nonresponse bias is likely to occur as a result of not obtaining 100 percent response from the selected cases.

The **potential magnitude of nonresponse bias** is estimated as the product of the nonresponse rate and the difference in values of a characteristic between respondents and nonrespondents.

Stage of data collection includes any stage or step in the sample identification and data collection process in which data are collected from the identified sample unit. This includes information obtained that is required to proceed to the next stage of sample selection or data collection (e.g., school district permission for schools to participate or schools providing lists of teachers for sample selection of teachers).

STANDARD 1: NCES universe data collections must be designed to meet a target unit response rate of at least 95 percent.

> **GUIDELINE:** A unit level nonresponse bias analysis is *recommended* if the universe survey unit response rate is less than 90 percent. (See Standard 4-4-02 for a discussion of nonresponse bias analysis.)

STANDARD 2: NCES sample survey unit response rates must be calculated both with and without substitutions. NCES sample survey data collections must be designed to meet unit level response rate parameters that are at least consistent with historical response rates from surveys conducted with best practices.

> **GUIDELINE:** The following parameters summarize current NCES historical experiences:

1. For longitudinal sample surveys, the target school level unit response rate should be at least 70 percent. In the base year and each follow-up, the target unit response rates at each additional stage should be at least 90 percent.

2. For cross-sectional samples, the target unit response rate should be at least 85 percent at each stage of data collection.
3. For random-digit dial sample surveys, the target unit response rate should be at least 70 percent for the screener and at least 90 percent for each component.
4. For household sample surveys, the target response rates should be at least ?80? percent for the screener, and at least ?90? percent for the respondents.
5. For assessments, the target response rate should be at least ?80? percent for schools and at least ??? percent for students.

STANDARD 3: NCES sample survey data collections must be designed to meet a target item response rate for each key item of at least 90 percent.

STANDARD 4: A nonresponse bias analysis is *required* at any stage with a unit response rate less than 85 percent. If the item response rate is below 85 percent for any items used in a report, a nonresponse bias analysis is also *required* for each of those items (this does not include individual test items). The extent of the analysis must reflect the magnitude of the nonresponse (see Standard 4-4-02).

- In longitudinal sample surveys, item nonresponse bias analyses need only be done once for any individual item, unless there is a substantial deterioration in the item response rate.

STANDARD 5: In cases where prior experience suggests the potential for an overall unit response rate of less than 50 percent, the decision to proceed with data collection must be made in consultation with the Associate Commissioner, Chief Statistician, and Commissioner.

SUBJECT: DEVELOPING A REQUEST FOR PROPOSAL (RFP) FOR SURVEYS

NCES STANDARD: 2-3-02

PURPOSE: To assist NCES staff in the preparation of high quality RFPs. Each RFP should include the information required to allow an offeror to submit a proposal that demonstrates technical and managerial competence sufficient to complete successfully all phases of surveys. Each RFP should include evaluation criteria to assist the government in selecting the best offeror to conduct the work. The RFP should provide a clear, precise, and accurate description of the requirement for the work and the expected activities, services, products, and level of effort to be delivered under the contract.

DEFINITION: An **Award Incentive Plan** links all or some of the contract deliverables to performance incentive payments beyond the fixed fee of the contract. There are minimum performance-based requirements that must be specified in order for a contract to be considered as an Award Incentive performance-based contract.

STANDARD 1: RFPs must be written in compliance with guidelines established in the Federal Acquisition Regulations (FAR) and in other departmental administrative procedures and guidelines.

> **GUIDELINE A:** The contracting office of the Department of Education is responsible for the acquisition process for NCES and can provide expertise and guidance in the development of the RFP.
>
> **GUIDELINE B:** Within NCES, the staff member who develops a Statement of Work (SOW) and related documents should also be designated Contracting Officer's Representative (COR). The individual who develops the SOW should have completed courses required for COR certification. Minimally, the individual designated as COR should be included in the development process, to provide familiarity with the contractual requirements and expectations.

STANDARD 2: The Statement of Work (SOW) must contain technical, managerial, and deliverable specifications (see Standards 1-1-02 and 2-2-02).

> **GUIDELINE:** The technical specifications for all phases of design, implementation, and analysis include: statistical, timeline, resource, analysis, and data file parameters. Managerial specifications should be written as specific activities and tasks. Those to be performed by the contractor and those to be performed by NCES should be clearly delineated. There should be a schedule for all deliverables (e.g., analysis plans, final reports).

STANDARD 3: The COR must be fully certified and must maintain COR certification. COR certification requires courses on contracting overview, independent government cost estimates, preparing performance-based statements of work, and contract administration. To maintain COR certification, the COR must complete an advanced contract administration course every 2years as well as periodic required courses, such as courses on the Department of Education's financial management system, EDCAPS, and the Contracts and Purchasing Software System (CPSS).

STANDARD 4: The COR must develop an Independent Government Cost Estimate (IGCE) that includes estimates of the cost of the project for all phases and elements of the survey system in terms of the contractor's manpower commitment by labor categories and other related costs. Automated Data Processing (ADP) cost, or Information Technology (IT) costs, must be estimated within each of the budget categories, to yield an estimate of total ADP costs within the total budget. Total estimated cost is to equal NCES budget amounts.

> **GUIDELINE:** For further information, consult previous comparable project estimates.

STANDARD 5: To obtain funding commitment, the COR must initiate the authorization and have it approved by the Division's Associate Commissioner. The COR must confirm the survey's fiscal year scheduled activity and obtain all accounting information with the budget contact source in the Office of the Deputy Commissioner (ODC). The ODC will commit the survey funds in the Department's financial system and electronically submit the authorization to the Contracting Officer (CO).

STANDARD 6: The Proposal Evaluation Plan specifies the membership of the Technical Evaluation Panel (TEP), who act as advisers to the Contracting Officer (CO). The plan also provides the criteria on which the COR and the TEP assess the proposals. The COR, in collaboration with the CO, assigns the factors and weights associated with each criterion. Only criteria and weights stated in the RFP may be used to evaluate submitted proposals (see Standard 1-1-02 and 2-2-02).

> **GUIDELINE:** The criteria may include such factors as technical competence, analysis plan, familiarity with data files, management plan.

STANDARD 7: The Proposal Preparation Instructions inform the offeror as to the substantive, format, and organizational requirements for completing their proposal. The offeror must submit two separate proposals: (1) technical and (2) business. They are evaluated separately.

STANDARD 8: The COR must prepare the required clearances and approvals for the planned survey activity. The standard clearances for all new RFPs are currently the Information Technology (IT) Resources clearance, Impact Determination clearance, and the Administrative Test for Characterizing Particular Services as "Personal" or "Nonpersonal" clearance.

GUIDELINE: Each RFP survey may have its own applicable/special clearances depending on the type of procurement required. (The ACS Departmental Directive, C: GPA: 2-105, Acquisition Planning, dated June 10, 1992 or later should be referenced to explain the standard clearances noted above and possible other clearances or approvals that might be required.)

STANDARD 9: The Award Incentive Plan for a performance-based contract must include a description of deliverables, schedules, and other evaluation criteria. It must also provide definitions of quality for each criterion and the associated incentive award fee or penalty. The evaluation criteria must include, but are not limited to, the definition of the work in measurable terms and/or mission-related terms.

GUIDELINE A: This plan tells the contractor what activity or product is required to be considered for an award incentive, above and beyond the acceptable standards for the contract. It also tells the contractor when penalties may be applied. In addition to a specified set of activities or products, NCES may include an option to pre-select at random additional deliverables for award or penalty.

GUIDELINE B: Award incentives criteria frequently include such criteria as quantity, timeliness, or quality. Other criteria that are sometimes used include commercial or industry-wide standards that are used to measure performance.

GUIDELINE C: An award fee incentive can be applied as a specified amount for a specific deliverable or the award fee can be applied in increments related to quality of the deliverable. Award incentive fees are based on the Contracting Officer's Representatives (COR) evaluation and ranking of the deliverables. The amount of the award incentive fee is determined by negotiations involving the COR, NCES senior management, and the Contracting Officer prior to awarding the contract.

GUIDELINE D: The following documents offer specific guidance on how to develop a performance-based solicitation:

- "Information on Best Practices for Performance-Based Service Contracting," October 1998, Published by the Office of Federal Procurement Policy at OMB.
- "Federal Acquisition Circular 97-1."
- "Federal Acquisition Regulation Subpart 37.6."

These documents are accessible through the *Acquisition Reform Network* (http://www.arnet.gov).

SUBJECT: FIELD TESTING SURVEY SYSTEMS

NCES STANDARD: 2-4-02

PURPOSE: To ensure that all components of a survey system will function as intended when implemented in the full-scale survey.

STANDARD: Components of a survey system need to be successfully demonstrated through previous work, or they must be pretested prior to implementation of the full-scale survey. The design of a pretest must reflect conditions likely to pose difficulties for the survey. Documentation of the pretest (e.g., materials for technical review panels, working papers, technical reports) must include the design of the pretest; a description of the procedures followed; analysis of the extent to which the survey components met the pre-established criteria; discussion of other potential problems uncovered during the pretest; and recommendations for changes in the design to solve the problems.

GUIDELINE A: Elements to be tested and measured may include alternative approaches to accomplishing a particular task. Elements to be tested may include: frame development; sample selection; questionnaire design; data collection; response rates; data processing (e.g., entry, editing, imputation); estimation (e,. weighting, variance computation); file creation; and tabulations.

GUIDELINE B: For an ongoing survey, new elements or content should be pretested, along with elements being changed as a result of the evaluation of the survey (see Standard 4-3-02).

GUIDELINE C: The evaluation criteria for a successful pretest should be developed before the pretest begins. Key evaluation criteria are established during the design stage. If the criteria are not met, that survey component should not be implemented without pretesting a redesigned component.

GUIDELINE D: The results of a pretest should be available and analyzed for internal use prior to making a decision to implement the full-scale survey.

GUIDELINE E: Survey design and instrumentation should be revised to reflect modifications suggested by the results of the pretest. A revised budget should be developed to reflect both changes in design and knowledge gained during the pretest about resource requirements.

SUBJECT: MAINTAINING DATA SERIES

NCES STANDARD: 2-5-02

PURPOSE: To maintain NCES data series that are consistent over time, and to establish and maintain a regular schedule for the release of NCES time series data.

DEFINITIONS: A **crosswalk study** delineates how categories from one classification system are related to categories in a second classification system.
A **bridge study** continues an existing methodology concurrent with a new methodology for the purpose of defining the relationship between the new and old estimates.
A **consistent data series** maintains comparability over time by keeping an item fixed, or by incorporating appropriate adjustment methods in the event an item is changed.

STANDARD 1: NCES must maintain and report on a consistent set of data series that may be analyzed over time. Ongoing data collections must maintain and report on a consistent set of data items and data collection procedures.

> **GUIDELINE A:** Identify the basic data items to be assessed on a regular basis to address policy issues and other information needs.
>
> **GUIDELINE B:** Provide estimates of both change and level for time series data in reports. For survey reports, consider publishing 3 or more years of the time series data along with the current year to highlight the time series.
>
> **GUIDELINE C:** Provide a list of other publications containing the data for previous years in the appendix of a survey report.

STANDARD 2: If changes are needed in data items or survey procedures for data series, a plan must be developed that provides justification or rationale for the changes in terms of their usefulness for policy-makers, conducting analyses, and addressing information needs. The plan must also describe adjustment methods, such as crosswalks and bridge studies that will be used to preserve trend analyses.

SUBJECT: EDUCATIONAL TESTING

NCES STANDARD: 2-6-02

PURPOSE: To ensure that instruments used in NCES surveys for measuring and making inferences about education-related domains are technically sound and fair.

DEFINITIONS: Instrument refers to an evaluative device that includes tests, scales, and inventories to measure a domain using standardized procedures.
Domain refers to a defined universe of knowledge, skills, abilities, attitudes, interests, or other human characteristics.
An **accommodation** is a change in how a test is presented, in how a test is administered, or in how the testtaker is allowed to respond. This term generally refers to changes that do not substantially alter what the test measures. The proper use of accommodations does not substantially change academic level or performance criteria. Appropriate accommodations are made to provide equal opportunity to demonstrate knowledge.
An a**ssessment** is any systematic procedure for obtaining information from tests and other sources that can be used to draw inferences about characteristics of people, objects, or programs.
Classical test theory postulates that a test score can be decomposed into two parts—a true score and an error component; that the error component is random with a mean of zero and is uncorrelated with true scores; and that observed scores are linearly related to true scores and error components.
A **cut score** is a specified point on a score scale such that scores at or above that point are interpreted or acted upon differently from scores below that point.
A **derived score** is a raw score converted by numerical transformation into a new score providing a more meaningful and/or different measure (e.g., conversion of raw scores to percentile ranks, standard scores, or grade equivalence).
Differential Item Functioning (DIF) exists when examinees of equal ability differ on an item solely because of their membership in a particular group.
Disability is a physical or mental impairment that substantially limits one or more of the major life activities (42 U.S.C. 12102).
Equating of two tests is established when examinees of every ability level and from every population group can be indifferent about which of two tests they take. Not only should they have the same expected mean score on each test, but they should also have the same errors of measurement.
An **Individualized Education Program (IEP)** refers to a written statement for each individual with a disability that is developed, reviewed, and revised in accordance with Title 42 U.S.C. Section 1414(d).
Item Response Theory postulates that the probability of correct responses to a set of test questions is a function of true proficiency and of one or more parameters specific to each test question.
Linkage results from placing two or more tests on the same scale, so that scores can be used interchangeably.

Reliability is the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and repeatable for an individual test taker.
Scaling refers to the process of assigning a scale score based on the pattern of responses.
Scoring/rating is the process of evaluating the quality of the examinee's responses to individual cognitive questions.
Section 504 of the Rehabilitation Act of 1973, as amended (Title 29 U.S.C. 794 Section 504), prohibits discrimination on the basis of handicap in federally assisted programs and activities.
Validity is the extent to which a test or set of operations measures what it is supposed to measure. Validity refers to the appropriateness of inferences from test scores or other forms of assessment.

STANDARD 1: Instrument Development—All testing instruments used in NCES surveys must be developed following an explicit set of specifications for the instruments. All evidence bearing on the development of the instrument must be compiled and documented so that it can be replicated. The instrument documentation must include the following:

1. The purpose(s) of the instrument;
2. The domain/construct that will be measured;
3. The framework of the instrument in terms of items, tasks, or questions, the response formats, modes of responding;
4. The number of items and time to administer;
5. The context in which the instrument will be used;
6. The characteristics of the intended participants;
7. The desired psychometric properties of the items, and the instrument as a whole;
8. The conditions and procedures of administering the instrument;
9. The procedures of scoring; and
10. The reporting of the obtained scores.

GUIDELINE A: Relevant experts should review the definition of the domain and the instrument specifications. The qualifications of the experts, the process by which the review is conducted, and the results of the review should be documented.

GUIDELINE B: All items should be reviewed before and after pilot and field tests. The pilot and field tests should be conducted on subjects with characteristics similar to intended participants. The procedures used to select the samples for pilot and field tests and the characteristics of the samples should be documented.

GUIDELINE C: The field test sample should include an adequate number of cases with the characteristics necessary to determine the psychometric properties of items.

GUIDELINE D: The item review process should include empirical analysis. The model (e.g., Classical and/or Item Response Theory) used to evaluate the psychometric properties of the items should be documented.

GUIDELINE E: When a time limit is set for performance, the extent to which the scores include a speed component and the appropriateness of this component to the defined domain should be documented.

GUIDELINE F: Permissible variations in conditions for administration should be identified, and a rationale for permitting the different conditions should be documented when the conditions of administration are permitted to vary from one group or participant to another.

GUIDELINE G: The directions for test administrations should be described with sufficient clarity for others to replicate.

GUIDELINE H: When a shortened or altered form of an instrument is used, the differences from the original instrument and the implications of those differences for the interpretations of scores should be documented.

STANDARD 2: Validity—All testing instruments used in NCES surveys must meet the purpose(s) stated in the instrument specification. All intended interpretations and proposed uses of raw scores, scale scores, cut scores, equated scores, and derived scores must be supported by evidence and theory. The degree of support must be appropriate for the intended interpretations and uses of the scores, including composite scores, sub-scores, score differences, and profiles.

GUIDELINE A: The rationale for each intended use of the testing instruments and proposed interpretations of obtained scores should be explicitly stated.

GUIDELINE B: Evidence of validity should be based on an analysis of the content, response process, internal structure of the instrument, or the relationship of scores to a criterion.

GUIDELINE C: When judgments occur in the validation process, the selection process for the judges (experts/observers/raters) and the criteria for judgments should be described.

GUIDELINE D: Potential users should be cautioned of unsupported interpretations; that is, interpretations of scores that have not been investigated, or interpretations of scores inconsistent with available evidence.

STANDARD 3: Reliability—The scores obtained by a testing instrument must be free from the effects of random variations due to factors such as conditions of administration and/or differences between scorers. The reliability of the scores must be adequate for the intended interpretations and uses of the scores.

GUIDELINE A: The reliability should be reported as a standard error of measurement or as an appropriate reliability coefficient (e.g., alternate form coefficient, test-retest/stability coefficient, internal consistency coefficient, generalizability coefficient).

GUIDELINE B: All relevant sources of measurement errors and summary statistics of the size of the errors from these sources should be reported.

GUIDELINE C: When average scores for participating groups are used, standard errors of group averages should be reported.

GUIDELINE D: The method (including selection of sample, sample sizes, sample characteristics) of quantifying the reliability of both raw and scale scores should be fully described.

GUIDELINE E: Scorer reliability, rater to rater, and rater-year reliability should be reported when the scoring process involves judgment.

GUIDELINE F: The reliability information on scores for each group should be reported when an instrument is used to measure different groups (e.g., race/ethnicity, gender, age, or special populations).

GUIDELINE G: Reliability information should be reported for each version of an instrument when original and altered versions of an instrument are available.

GUIDELINE H: Separate reliability analyses should be performed when major variations of the administration procedure are permitted to accommodate disabilities.

STANDARD 4: Fairness—Testing instruments used in NCES surveys must be designed, developed, and administered in ways that treat participants equally and fairly regardless of differences in personal characteristics such as race, ethnicity, gender, age, socioeconomic status, or disability that are not relevant to the intended uses of the instrument.

GUIDELINE A: Language, symbols, words, phrases, and content that are generally regarded as offensive by members of different groups should be eliminated, except when judged to be necessary for adequate representation of the domain.

GUIDELINE B: Meaningful differences between groups in test results should be investigated to make sure that they are not caused by construct-irrelevant factors. However, differences in the subgroups' performance do not necessarily indicate that a measurement instrument is unfair.

GUIDELINE C: When research shows that Differential Item Functioning (DIF) exists, studies should be conducted to detect and eliminate aspects of test design, content, and format that might bias test scores for a particular subgroup.

GUIDELINE D: In testing applications where the level of linguistic or reading ability is not a purpose of the assessment, the linguistic or reading demands of the test should be kept to a minimum.

GUIDELINE E: The testing or assessment process should be carried out so that test takers receive comparable and equitable treatment during all phases of the testing process.

STANDARD 5: Testing individuals with disabilities or limited English proficiency—Whenever possible, scores derived from testing instruments used in NCES surveys must validly, reliably, and fairly reflect the performance of all participants, including individuals with disabilities and individuals of diverse linguistic backgrounds. Appropriate and reasonable accommodations in accordance with applicable federal nondiscrimination laws for special populations must be incorporated. Differences in performance must reflect the construct under measurement rather than any irrelevant factors such as disabilities and/or language differences.

GUIDELINE A: Permitted accommodations and/or modifications for special populations and the rationale for each accommodation should be described in detail.

GUIDELINE B: Measurement standards of validity and reliability of data with accommodations should be provided.

GUIDELINE C: Specific information about accommodations should be included in the data file so that users can properly interpret the test scores.

For individuals with disabilities:

GUIDELINE D: The evidence of empirical procedures used to review items to ensure fairness, to evaluate whether DIF exists, and to determine accommodations for students/individuals with disabilities should be documented.

GUIDELINE E: Decisions about accommodations for individuals with disabilities should be made by individuals who are knowledgeable of existing research on the effects of the disabilities in question on test performance.

GUIDELINE F: The participant's Individualized Education Program (IEP) or Section 504 plan must be consulted prior to making determinations of whether a participant with a disability will participate in the assessment, and what accommodations, if any, are appropriate.

For individuals of diverse linguistic backgrounds:

GUIDELINE G: The evidence of empirical procedures used to review items to ensure appropriateness of materials for participants with various backgrounds and characteristics (e.g., nativity, experience in U.S. schools) should be documented to evaluate whether DIF exists, and to evaluate the linguistic or reading demands to ensure that they are no greater than required.

GUIDELINE H: Translation, evaluation procedures, and the comparability of the translated instrument to the original version should be documented when a translation of the instrument is made.

STANDARD 6: Administration—Administration of all testing instruments used within each NCES survey must be standardized. The administration must follow the procedures prescribed in the test administration manual. The administration manual must include the following:

1. A brief description of the purpose of the survey and the population to be tested;
2. A description of the required qualifications of those administering the instrument;
3. A description of the required identifying information of the participant;
4. A description of the materials, aids, or tools that are required, optional, or prohibited;
5. A description of the allowable instructions to the participants and procedures for timing the testing;
6. A description of assignment of participants to groups, or special seating arrangements, and preparation of participants as relevant;
7. A description of allowable accommodations;
8. A description of desired testing conditions/environment; and
9. A description of the procedures to maintain security of the materials as applicable, and actions to take when irregularities are observed.

GUIDELINE A: Administration procedures should be field tested. The approved procedures should be described clearly so they can be easily followed.

GUIDELINE B: Survey staff administering the instrument should be trained according to the procedures prescribed in the administration manual.

GUIDELINE C: Modifications or disruptions to the approved procedures should be documented so the impact of such departures can be studied.

GUIDELINE D: Instructions presented to participants should include sufficient detail so the participants can respond to the task in the manner intended by the instrument developer.

GUIDELINE E: Samples of administration sites should be monitored to ensure that the instrument is administered as specified.

STANDARD 7: Scoring and Scaling—Test scoring must be standardized and scales must be stable if used over time.

GUIDELINE A: Machine scoring procedures should be checked for accuracy. The procedure should be documented, as well as the nature and extent of scoring errors.

GUIDELINE B: Hand scoring procedures should be documented, including rules governing scoring decisions, training procedures used to teach the rules to the coding staff, quality monitoring system used, and quantitative measures of the reliability of the resulting ratings. The criteria for evaluating the quality of individual responses should not be changed during the course of the scoring process.

GUIDELINE C: All systematic sources of errors during the scoring process should be corrected and documented.

GUIDELINE D: Consistency among scorers and potential drift over time in scoring/rating should be evaluated and documented.

GUIDELINE E: The meanings, interpretations, limitations, rationales, and processes of establishing the scores that are reported should be clearly described in the technical report.

GUIDELINE F: Stability of the scale should be checked periodically when a scale is maintained over time.

GUIDELINE G: The procedures for scoring—raw scores, scale scores—should be documented. The documentation should also include a description of the populations used for their development.

GUIDELINE H: Procedures for deriving the weights should be described when weights are used to develop the scale scores.

GUIDELINE I: The population norms to which the summary statistics refer should clearly be defined when group performance is summarized using norm scores.

GUIDELINE J: The rationale and procedures for establishing the cut scores should be documented when cut scores are established as part of the scale score reporting.

GUIDELINE K: Cut scores should be valid; that is, participants above a cut point should demonstrate a qualitatively greater degree and/or different type of skills/knowledge compared to those below the cut point.

GUIDELINE L: The method employed in a judgmental standard-setting process should be documented. The documentation should include the following:

1. The selection and qualifications of judges;
2. The nature of the request for their judgments;
3. The training provided to the judges;
4. The feedback of information to judges;
5. The opportunities for judges to confer with one another concerning their judgments; and
6. The methods used to aggregate the judgments and translate them into cut scores.

GUIDELINE M: The judgmental methods used to establish cut scores should meet the following three criteria:

1. The judgmental method should involve peer review and pre-testing.
2. The judgments to be provided should not be so cognitively complex that the judges are unable to provide meaningful judgments.
3. The process used to set cut scores should be described in sufficient detail so the process can be replicated.

GUIDELINE N: An estimate of the amount of variability in cut scores must be provided regardless of whether the standard-setting procedure is replicated.

GUIDELINE O: Equating/linking functions should be invariant across sub-populations when equating or linking is used to determine equivalent scores. Supporting evidence for the interchangeability of tests/test–forms should be provided.

GUIDELINE P: Detailed technical information (i.e., design of equating studies, standard errors of measurement, statistical methods used, size and relevant characteristics of samples used, and psychometric properties of anchor items) should be provided for the methods by which equating or linking is established.

GUIDELINE Q: Users should be warned that scores are not directly comparable when converted scores from two versions of the test are not strictly equivalent.

STANDARD 8: Reporting—Results of the testing should be provided with sufficient detail and contextual information to understand the inferences that can and cannot be made from them.

GUIDELINE A: The analysis of item responses or test scores should be described in detail, such as procedures for scaling or equating.

GUIDELINE B: Appropriate interpretations of all reported scores should be provided. The interpretations should describe what the test covers, what the scores mean, and the precision of the scores. The generalizability and limitations of reported scores should also be presented.

GUIDELINE C: Validity and reliability should be reported for the level of aggregation for which the scores are reported when matrix sampling is used. Scores should not be reported for individuals unless the validity, comparability, and reliability of such scores indicate that reporting individual scores is meaningful.

STANDARD 9: Manual—All evidence of the standards set forth above for each testing instrument used in NCES surveys must be compiled in a manual.

GUIDELINE: A report should provide technical and psychometric information on a test as well as information on test administration, scoring, and interpretation.

REFERENCES

The Use of Tests When Making High-Stakes Decisions for Students: A Resource Guide for Educators and Policymaker, July 6, 2000. U.S. Department of Education, Office for Civil Rights. Author and Publisher.

ETS Standards for Quality and Fairness. (2000). Educational Testing Service. Princeton, NJ. Author and Publisher.

Standards for Educational and Psychological Testing. (1999). Prepared by the Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME), Washington, DC: AERA.

NCES Statistical Standards. (1992, Reprinted in May 1996). U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington DC: U.S. Government Printing Office. NCES 92–021.

Code of Fair Testing Practices in Education. (1980). Prepared by the Joint Committee on Testing Practices, Washington, DC.

COLLECTION OF DATA

3-1 Coverage for Frames and Samples

3-2 Achieving Acceptable Response Rates

3-3 Monitoring and Documenting Survey Contracts

3-4 Documenting a Survey System

SUBJECT: COVERAGE FOR NCES FRAMES AND SAMPLES

NCES STANDARD: 3-1-02

PURPOSE: To provide guidance on the coverage issues involved in developing, maintaining, and using sampling frames.

DEFINITIONS:

Capture/Recapture: This technique uses two independent frames to estimate the number of units missed on both frames. The first step is to match frames to provide counts of units on one frame, but not the other; as well as a count of units on both frames. With this information, it is possible to estimate the number of units missed on both frames.

Coverage Error: Errors in coverage occur when target population units are missed during frame construction (**undercoverage**) or when they are erroneously included, duplicated, or enumerated in error (**overcoverage**).

Dual-Frame Estimation: A dual-frame design combines two frames in the same survey to offer coverage rates that exceed those of any single frame. Sometimes the best available list is known to have poor coverage and there are no known supplemental frames to provide sufficient coverage.

A **supplemental area frame** can be created. This is often done by first, generating a frame of geographic units where all the geographic units are represented providing full geographic coverage. Next, a probability sample of the geographic units is selected. An intensive search procedure is carried out in each selected area. This generates a supplemental area frame for each selected area. Assuming no error in the search process, the supplemental area frame has complete coverage and the cases can be weighted to represent a national estimate. The data from both the main list frame and the supplemental area frame are then combined so that the weighted sample estimates provide complete coverage.

Frame: A list of all the units that represent the population.

Frame Population: The set of elements that can be identified prior to the selection of a survey sample.

Half-Open Interval: This is a technique used to increase coverage. It is usually applied to a new list that includes cases that were covered in a previous frame, as well as new in-scope units not included in the previous frame. In this technique, new in-scope units between a unit A on the previous frame up to, but not including, unit B (the next unit on the previous frame) are associated with unit A. These new units have the same selection probability as unit A's. This process is repeated for every unit on the previous frame. The new units associated with the actual sample cases are now included in the sample with their respective selection probabilities (**freshening**).

Noncoverage: Units of the target population that are missing from the frame population. Includes the problems of incomplete frames and missing units.

Target Population: The finite set of observable or measurable elements that will be studied, or the conceptual population of analytic units for which data are collected and estimates are made.

Un-Duplication: The process of deleting units that are erroneously in the frame more than once (**overcoverage**).

STANDARD 1: Staff responsible for NCES data collections that are used as sampling frames for other NCES surveys must have coverage evaluated and coverage rates documented at least once every 5 years.

GUIDELINE A: NCES staff should evaluate frames by comparing against alternative frames found inside and outside of the Department of Education, considering total list count comparisons, matching operations, dual-frame estimation, using capture/recapture procedures to estimate noncoverage, and providing an estimation of missing units.

GUIDELINE B: Staff responsible for NCES data collections that are used as sampling frames should maintain two-way communications with survey staff who use their collection as a frame. Procedures such as sharing preliminary data files with survey staff in order to develop frames may be instituted. (For example, staff that use an administrative list of public schools for their frames should be alerted when new data are available and each time there is a major change in the list.)

STANDARD 2: NCES data collections that are used as sampling frames for other NCES surveys must strive for completion rates in excess of 95 percent overall and for each major subdomain.

STANDARD 3: Staff using NCES frames and samples must be cognizant of coverage issues, and must take the steps necessary to provide satisfactory coverage for the survey. If there is not evidence of a coverage rate of at least 85 percent, then frame enhancements such as frame supplementation or dual frame estimation must be incorporated into the study design.

GUIDELINE A: The first time a survey is conducted, background design and coverage work should be done before choosing the frame. Alternative frames, if applicable, should be considered and compared.

GUIDELINE B: Coverage errors such as over- and under-coverage, bad contact information, classification, temporal errors, and other listing errors should be minimized before the use of a frame. Techniques such as list supplements, half-open intervals, and un-duplication can be used to reduce these errors and improve coverage of the frame.

GUIDELINE C: Any possible changes to frame variables identified by sample survey staff should be reported to the staff responsible for the data collection being used as the frame. For example, the relevant variables to maintain and consider include: 1) eligibility (e.g., grade span); 2) contact information (e.g., name, address, and phone number); 3) stratification variables (e.g., state and school level); and 4) measures of size (e.g., grade enrollment).

GUIDELINE D: To reduce coverage error, whenever a frame has important deficiencies with respect to the measurement unit, dual-frame estimation should be considered to correct these deficiencies. Since dual-frame estimation can be expensive, the effect dual-frame estimation has on increasing the variance estimates should also be considered when deciding to use dual-frame estimation.

STANDARD 4: For each sample survey, a description of the frame and its coverage must be included in the survey documentation. This description must include, but is not limited to, the target and frame populations (and exclusions thereof); the name and date of the data collection which provided the original frame; any supplementing done to the original frame; limitations of the frame including the timeliness of the frame; and, if applicable, an estimation of the missing units on the frame.

GUIDELINE A: For each sample survey documentation should include a discussion of coverage issues such as alternative frames that were considered, what the survey did to improve the coverage of the frame, and how data quality and item non-response on the frame may have affected the coverage of the frame.

GUIDELINE B: Documentation should include estimation techniques used to improve the coverage of estimates. This would include post-stratification procedures. (For example, a telephone survey could post-stratify estimates of all individuals to account for the exclusion of those without telephones.)

GUIDELINE C: NCES survey staff should, unless it is in conflict with confidentiality laws, archive their survey's sampling frames as part of the documentation of the survey system found in Standard 3-4-02. This may be particularly important if a preliminary file was used to develop the frame, or if there is a chance that the frame may be used in the future to further develop research questions.

STANDARD 5: NCES survey staff that use NCES data collections as a frame should share any coverage or usage issues with the NCES data collection staff so that the collection can improve their coverage issues for future use. This guideline is related to Guidelines B and C of Standard 2 in this Standard. (For example, after the survey is complete, the survey staff should provide a memo to the NCES data collection staff for the survey used as a frame, reviewing the major limitations of the coverage or the data quality found in using the data collection as a sampling frame.)

SUBJECT: ACHIEVING ACCEPTABLE RESPONSE RATES

NCES STANDARD: 3-2-02

PURPOSE: High response rates help ensure that results are representative of the target population. Data collection programs conducted by or for the NCES must be conducted in a manner that encourages high rates of response across all strata.

DEFINITIONS: Critical data items include the minimum set of items required for a case to be considered a respondent.

STANDARD 1: The data collection, independent of collection methodology (e.g., whether mailed, over the Internet, or administered by an interviewer either in person or by telephone), must be designed and administered in a manner that encourages respondents to participate.

GUIDELINE A: The method of data collection (e.g., mail, telephone, Internet, etc.) should be appropriate for the target population and the objectives of the data collection.

GUIDELINE B: The data should be collected at the most appropriate time of year.

GUIDELINE C: The data collection period should be of adequate and reasonable length to achieve good response rates.

GUIDELINE D: When appropriate, respondent incentives should be considered.

STANDARD 2: An explanation of the need for data, the goals and objectives of the data collection, and examples of uses of the data that benefit the respondent must be provided to the respondent (Privacy Act of 1974, as amended, 5 U.S.C. 552a).

GUIDELINE A: The materials describing the data collection should be sent to the respondent in advance, when possible.

GUIDELINE B: For data collection programs requiring an interviewer, training should emphasize techniques for obtaining respondent cooperation and techniques for building rapport with the respondent, including: manner, follow-up skills, knowledge of the goals and objectives of the data collection, and knowledge of the uses of the data.

GUIDELINE C: Prior to conducting a data collection program, endorsements, support, and the active cooperation of interested groups, such as professional organizations, professional associations, education community leaders, and state

and local school district officials, should be obtained and communicated to the respondent.

STANDARD 3: All NCES data collections must provide information concerning the confidentiality of responses. Privacy and confidentiality assurances citing the appropriate legislation must be provided, as applicable (see Standard 4-2-02).

STANDARD 4: In keeping with the goals of the particular data collection effort, respondent burden must be minimized, as required by the Office of Management and Budget clearance process.

GUIDELINE A: The questionnaire should be pretested and piloted for the difficulty and interpretability of questions.

GUIDELINE B: The questionnaire should be pretested for ease in navigation of self-administered questionnaires.

GUIDELINE C: Questions should be clearly written and skip patterns easily followed.

GUIDELINE D: The questionnaire should be of reasonable length.

STANDARD 5: All data collection programs require some follow-up of nonrespondents to achieve desirable response rates. Therefore, followup strategies designed to achieve acceptable response rates need to be planned and included in the RFP and contract, or in the Interagency Agreement.

GUIDELINE A: Internal reporting systems that provide timely reporting of response rates and the reasons for nonresponse throughout the data collection should be developed. These systems should be flexible enough to identify important subgroups with low response rates for more intensive follow-ups.

GUIDELINE B: For longitudinal surveys, obtain as much locating information about respondents as possible during initial contact (e.g., for a student, school address, home address, name of advisor, phone numbers of parents).

GUIDELINE C: If response rates are low after the initial phases of data collection, and if further data collection on the full sample is deemed too costly, take a random subsample of nonrespondents and use a more intensive data collection method. This subsample will permit a description of nonrespondents' characteristics, provide data needed for nonresponse bias analysis, and allow for possible weight adjustments or for imputation of missing characteristics.

GUIDELINE D: Determine a set of critical data items to obtain when a respondent is unwilling to fully cooperate. These variables may then be targeted in followup to meet the minimum standard for unit response. These variables may also be used in a nonresponse bias analysis that compares characteristics of respondents and nonrespondents using the sample data for those items. These critical items may also be used for item nonresponse imputation systems.

SUBJECT: MONITORING AND DOCUMENTING SURVEY CONTRACTS

NCES STANDARD: 3-3-02

PURPOSE: To assist NCES staff in monitoring and documenting contract activities.

STANDARD 1: The Contracting Officer's Representative (COR) must work to ensure that the contractor meets: (a) contract specifications, (b) contract schedules, (c) NCES standards, (d) performance cost controls, and (e) beneficial effort/method of performance criteria in fulfilling the contract. Education Department Directive C:GPA:2-105 dated 6/15/92 established the *Standards and Guidelines for the Monitoring of Contracts*.

In some instances, the contractor may request technical redirection for unanticipated problems. For simple matters that are clearly within the scope of the contract, such requests may be made verbally. For problems that may require a change in scope, all requests must be in writing and outline the issue(s) and potential options. The COR must use this information in discussions with other NCES senior management in determining the appropriate course of action. All changes in any contract scope of work require action by the Contracting Officer. Whatever course of action is taken, it must be documented and placed in the project files.

GUIDELINE A: The COR should maintain close communication with the contractor. Depending on the nature of the survey, the COR should maintain communication through the use of meetings, phone calls, e-mails, visits, and/or the electronic management information system (MIS) for the purpose of tracking and monitoring the progress of the survey.

GUIDELINE B: The COR should review and verify progress reports, vouchers, technical products and documentation, written correspondence, and other documents for the following purposes:

1. Monitoring adherence to project schedules and requirements;
2. Assuring deliverables meet NCES standards and comply with the conditions of the contract and other quality requirements (e.g., accuracy and completeness); and
3. Identifying potential problems that would substantially affect the successful completion of the survey or alter the terms and conditions of the contract (e.g., cost or time increases, quality decreases).

GUIDELINE C: The status of each unit of observation should be kept current and available to the COR at each stage of the data collection process. Critical status events may include, but are not limited to, dates of questionnaire mailout, returns, deletions (out-of-scopes), scan editing, data entry, machine editing, callback(s), and

addition to the final data files. The COR should request direct and rapid access to the information.

GUIDELINE D: To help decide whether any adjustments or corrective actions are needed, soon after initial startup of field operations, and less frequently thereafter, the COR should evaluate the quality of survey operations by comparing a sample of the original returned questionnaires with the information on the data file for the following purposes:

1. Detect any data processing errors,
2. Learn of any problems with reporting or questionnaire design, and
3. Ensure that editing/update procedures are being correctly implemented.

GUIDELINE E: On an as-needed basis, CORs may request a copy of "completed" records from the current master file (sometimes referred to as a "pull") and analyze the information for conformance to contract requirements. The extent of the statistical analysis of a pulled database should vary with survey objectives. Simple cross-tabulations and frequencies of discrete variables should normally point out internal coding inconsistencies and also provide interim item response rates. Simple descriptive statistics for continuous variables should provide interim item response rates, measures of dispersion, and outliers.

GUIDELINE F: The COR should ensure that software used for weighting, imputations, and variance estimation is accurate. This may be done through a series of "checkpoints" imbedded within the program(s). Another alternative is to have the contractor provide printouts from a series of discrete steps with review by the COR along the way.

GUIDELINE G: The COR should keep the CO and NCES management informed of the result of reviews. As an integral part of this work, the COR should offer recommendations for solving any problems, acceptance of deliverables, performance awards, and approval or disapproval of any proposed changes.

STANDARD 2: The COR must maintain the following documents in the COR contract file: (a) progress reports, (b) vouchers, and (c) deliverables as required by the contract. Together with the RFP, contract proposal, proposal evaluation, and signed contract, these documents are subject to audit. Also document any modifications or changes in (a) key personnel, (b) project schedule, (c) deliverables, and (d) scope of work, and their implications for the project completion date, deliverables, and costs.

GUIDELINE A: It is advisable to include in the contract file all correspondence, such as logs of phone conversations, e-mail and written correspondence, and documentation, describing the approval of or decisions made regarding changes.

GUIDELINE B: The COR should keep accurate and complete records of contractor performance, such as lateness, unacceptable deliverables, and cost overrun. Actions or decisions taken by the COR or CO to remedy the problems should also be clearly documented.

STANDARD 3: CORs should require that all computer programs **(**software**)** be self-documenting.

GUIDELINE: The programmer should insert "comments" within the program(s) to describe each discrete section of code. Relationships between programs and data files should be flowcharted or described in a separate document. This includes record layouts and file structures.

STANDARD 4: Upon completion and/or termination of the contract, the COR must archive those items specified in the Standard for Documenting a Survey System (3-4-02) and Standard for Survey Documentation in Center Reports (7-2-02).

SUBJECT: DOCUMENTING A SURVEY SYSTEM

NCES STANDARD: 3-4-02

PURPOSE: To insure that complete documentation is kept on NCES surveys and survey systems and their associated contract deliverables. Complete documentation is defined as those materials necessary to understand how to properly analyze data from each survey. It also includes information necessary to replicate and evaluate each survey. In addition, survey system documentation includes information necessary to design and estimate resource requirements of future similar surveys.

STANDARD 1. Survey system documentation must include all information necessary to properly analyze the data. This information shall, at a minimum, include the following:

1. The final data set(s);
2. The final instrument(s) or a facsimile thereof;
3. Definitions of all variables;
4. Data file layout;
5. Descriptions of variables on the data file constructed from responses to other variables on the file;
6. Descriptions of variables used to uniquely identify cases in the data file;
7. Descriptions of sample weights and how to apply them;
8. Descriptions of the strata and primary sampling unit (PSU) identifiers to be used for analysis;
9. Descriptions of how to calculate variances appropriate for the survey design;
10. Descriptions of all imputation methods applied to the data and how to remove imputed values from the data; and
11. Descriptions of known data anomalies.

GUIDELINE: If the data are collected through a Web-based collection or through a CATI or CAPI interview, the following information should be included in the documentation of the final instruments:

1. All items in the instrument (e.g., questions, check items, and help screens);
2. Items extracted from other data files to pre fill the instrument (e.g., dependent data from a prior round of interviewing); and
3. Items that are input to the post data collection processing steps (e.g., output of an automated instrument).

STANDARD 2: To insure that a survey can be replicated and properly evaluated, the survey system documentation must, at a minimum, include the following:

1. Justifications for the items on the survey instrument, including how these final items were selected;
2. All instructions to respondents and/or interviewers either about how to properly respond to a survey item or how to properly present a survey item;
3. Description of the data collection methodology;
4. Sampling plan and justifications for why it was implemented, and, if possible, the final sample frame;
5. Selected sample;
6. Description of the magnitude of sampling error associated with the survey, and how it was calculated;
7. Description of the sources of nonsampling error associated with the survey (e.g., coverage, measurement);
8. Unit response rates (weighted and unweighted);
9. Item response rates; and
10. Total response rate.

GUIDELINE A: The survey system documentation should also include the following:

1. Final weighting plan specifications, including calculations for how the final weights were derived, and justifications for why it was implemented;
2. Final imputation plan specifications and justifications for why it was implemented;
3. Data editing plan specifications and justifications for why it was implemented; and
4. Data processing plan specifications and justifications for why it was implemented;

GUIDELINE B: Where appropriate, methods for bounding or estimating the nonsampling error from each source identified in the evaluation plan should be developed and implemented.

GUIDELINE C: Where possible, nonsampling error estimates and bounds should make use of data from other surveys or from administrative records or censuses, taking into account the limitations of the external data.

GUIDELINE D: For recurring surveys, a quality profile report that itemizes all sources of identified error should be produced. Where possible, estimates or bounds on the magnitudes of these errors should be provided; the total error model

for the survey should be discussed; and the survey should be assessed in terms of this model.

STANDARD 3: To insure that NCES has sufficient information to design future surveys and to accurately estimate their resource requirements, survey system documentation must include the following:

1. All information germane to the contractual operation of the survey, including the request for proposals used to solicit the contract(s);
2. The independent government cost estimate;
3. The contract(s) used to develop, conduct, and report on the survey;
4. Any modifications to the contract(s);
5. Final contract deliverables, progress reports, and vouchers; and
6. The Office of Management and Budget (OMB) clearance package and correspondence with OMB about survey clearance.

STANDARD 4: At a minimum, survey documentation must be stored electronically in a format that can be viewed without proprietary software. Final data sets shall be stored in ASCII format. Additional copies in other formats are allowed, but ASCII versions are required. In addition, substantive reports written to release the data shall also be stored, at a minimum, in the format originally used to produce the report, and PDF or ASCII (see Standard 7-1-02).

STANDARD 5: All reports, documentation, and public-use data must be stored on the Web, a CD-ROM, or an NCES dedicated server. Restricted data files and associated documentation shall be transmitted to the Statistical Standards Program for secure storage.

PROCESSING AND EDITING OF DATA

4-1 Imputation of Item Nonresponse

4-2 Maintaining Confidentiality

4-3 Evaluation of Surveys

4-4 Nonresponse Bias Analysis

SUBJECT: IMPUTATION OF ITEM NONRESPONSE

NCES STANDARD: 4-1-02

PURPOSE: To establish guidelines to reduce potential bias, ensure consistent estimates, and simplify analysis, by substituting values for missing or inconsistent data in a data set (i.e., using imputation).

Definitions: Imputation involves substituting values for missing or inconsistent data in a data set.
At a minimum, **key items** include items for which aggregate estimates are commonly published by NCES. They include, but are not restricted to, variables most commonly used in table row stubs. They include key composites and other policy-relevant variables.
A **freshened sample** includes cases from the longitudinal sample of a data set, plus new cases added to produce cross-sectional estimates of the population at the time of a subsequent wave of a longitudinal data collection.

STANDARD 1: Key items in data sets used for cross-sectional estimates must be imputed. This applies to cross-sectional data sets and to data from longitudinal data sets that are used to produce cross-sectional estimates (i.e., base year and subsequent freshened samples). (See appendix B for a discussion of alternative imputation procedures).

> **GUIDELINE A:** In census (universe) data collections, it may not be appropriate to impute data in certain situations (e.g., peer analysis situations or when data for a particular establishment—school, university, or library—are being examined individually).
>
> **GUIDELINE B:** When using non-NCES data sets, it is desirable to impute for missing data in those items being used in NCES publications. This is only appropriate when adequate auxiliary information is available.
>
> **GUIDELINE C:** Imputation procedures should be internally consistent, be based on theoretical and empirical considerations, be appropriate for the analysis, and make use of the most relevant data available. If multivariate analysis is anticipated, care must be taken to use imputations that minimize the attenuation of underlying relationships.

STANDARD 2: In cases where imputation is not used, data tables must include a reference to a methodology table or glossary that shows the actual weighted response rates for each unimputed variable included in the report (see Standard 1-3-02 for the item response rate formula). For individual variables with item response rates less than

85 percent, the variable must be footnoted in the row or column header. The footnote must alert readers to the fact that the response rate is below 85 percent.

STANDARD 3: In the case of longitudinal data sets, two imputation approaches are acceptable: cross-wave imputations or cross-sectional imputations. Cross-wave imputations may be used to complete missing data for longitudinal analysis or cross-sectional imputations may be used. (Guideline 1C of this Standard applies here, as well.)

STANDARD 4: When imputations are used, documentation indicating the weighted proportion of imputed data must be presented for all published estimates based on NCES data.

GUIDELINE A: Information about the amount of imputed data in the analysis can be included in the technical notes and does not have to accompany each table. Report the range of the amount of imputation for the set of items included in each analysis. Also, for items with response rates less than 70 percent report the amount of imputation for each item.

GUIDELINE B: Items with response rates lower than 70 percent should be footnoted in the tables.

STANDARD 5: All imputed values on a data file must be clearly identified as such.

GUIDELINE: Imputed data should be flagged in associated "flag" fields. The imputation method should be identified in the flag. Blanks are not legitimate values for flags.

STANDARD 6: If nonimputed items are used in the estimation of totals or ratios, the risks of not using imputed data must be described.

- Estimated totals using nonimputed data implicitly impute a zero value for all missing data. These zero implicit imputations will mean that the estimates of totals will underestimate the true population totals. Thus, when reporting totals based on a nonimputed item, the response rate for that item must be footnoted in the data table.
- Ratios (averages) using nonimputed data will implicitly impute the cell ratio for all missing data within the cell. This can cause inconsistencies in the estimates between tables.

SUBJECT: MAINTAINING CONFIDENTIALITY

NCES STANDARD: 4-2-02

PURPOSE: To protect the confidentiality of NCES data that contain information about individuals (individually identifiable information). For this reason, staff must be cognizant of the requirements of the law and must monitor the confidentiality of individually identifiable information in their daily activities and in the release of information to the public.

DEFINITIONS: Individually identifiable data refers specifically to data from any record, response form, completed survey, or aggregation about an individual(s) from which information about particular individuals may be revealed.

A **public-use data file** includes a subset of data that have been coded, aggregated, or otherwise altered to mask individually identifiable information, and thus, is available to all external users. Unique identifiers, geographic detail, and other variables that cannot be suitably altered are not included in public-use data files.

Public-use edits are based on an assumption that external users have access to both individual respondent records and secondary data sources that include data which could be used to identify respondents. For this reason, the editing process is relatively extensive. When determining an appropriate masking process, the public-use edit takes into account and guards against matches on common variables from all known files that could be matched to the public-use file. The analysis used to determine which records require masking is called a **disclosure risk analysis.**

Statistical disclosure limitation techniques are used to prepare microdata files for release. One set of these techniques, **Type 1**, directly alters the individual respondent's data for some variables, but preserves the level of detail in all variables included in the microdata file. Blanking and imputing for randomly selected records; blurring (e.g., combining multiple records through some averaging process into a single record); adding random noise; and data swapping or switching (e.g., switching the sex variable from a predetermined pair of individuals) are all examples of Type I disclosure limitation techniques. A second set of these techniques, **Type 2** disclosure limitation techniques, preserve the individual respondent's data by reducing the level of detail used to report some variables. Examples of this technique include: recoding continuous variables into intervals; recoding categorical data into broader intervals; and top or bottom coding the ends of continuous distributions.

Confidentiality edits are defined as edits that are applied to microdata for the purpose of protecting data that will be released in tabular form. Confidentiality edits are implemented using Type 1 disclosure limitation techniques. These techniques are used to alter the responses in the microdata file before tabulations are produced. Thus, all tables are protected in a consistent way. Because the Type 1 techniques that are used are designed to preserve the level of detail in the microdata file, confidentiality edits maximize the information that can be provided in tables, without requiring cell suppression or controlled rounding.

Data swapping is one form of a confidentiality edit. A simplistic example of data swapping would be to assume a data file has two potential individual identifying variables, for example, sex and age. If a sample case needs disclosure protection, it is paired with another sampled case so that each element of the pair has the same age, but different sexes. The data on these two records are then swapped. After the swapping, anyone thinking they have identified either one of the paired cases gets the data of the other case, so they have not made an accurate match and the data have been protected.

A **Data Analysis System (DAS)** is an analysis software system that generates tabular estimates and correlation coefficients in a framework that allows external users to analyze individually identifiable data without allowing the user direct access to individual data records. Users are denied access to individual data records because the data are not in a directly readable format. Additional safeguards come through the use of population subsampling and differential weighting from the sample design, as well as confidentiality edits. The degree of editing required is a direct function of the capabilities of the DAS. As an example, a DAS that provides weighted totals (i.e., a direct measure of population size) within cells would require more confidentiality editing than one that does not provide weighted cell totals, because there is a greater risk of disclosure in groups with small population size.

A **restricted-use data file** includes individually identifiable information that is confidential and protected by law. Restricted-use data files are not required to include variables that have undergone Type II disclosure risk edits.

LEGAL REQUIREMENTS: Three laws cover protection of the confidentiality of individually identifiable information collected by NCES—the Privacy Act of 1974, as amended, the National Education Statistics Act of 1994, as amended, and the US Patriot Act of 2001.

Privacy Act of 1974, as amended—"The purpose of this Act is to provide certain safeguards for an individual against invasion of personal privacy by requiring Federal agencies…to collect, maintain, use or disseminate any record of identifiable personal information in a manner that assures that such action is for necessary and lawful purpose, that the information is current and accurate for its intended use, and that adequate safeguards are provided to prevent misuse of such information." A willful disclosure of individually identifiable data is a misdemeanor, subject to a fine up to $5,000.

National Education Statistics Act of 1994, as amended—This law requires that no person may:

a. Use any individually identifiable information furnished under the provisions of this section for any purpose other than statistical purposes for which it is supplied, except in the case of terrorism (see discussion of the Patriot Act);

b. Make any publication whereby the data furnished by any particular person under this section can be identified; or

c. Permit anyone other than the individuals authorized by the Commissioner to examine the individual reports.

Further, individually identifiable information is immune from legal process, and shall not, without the consent of the individual concerned, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceeding, except in the case of terrorism. Employees, including temporary employees, or other persons who have sworn to observe the limitations imposed by this law, who knowingly publish or communicate any individually identifiable information will be subject to fines of up to $250,000, or up to 5 years in prison, or both (Class E felony).

US Patriot Act of 2001—This law permits the Attorney General to petition a Judge for an ex parte order requiring the Secretary of the Department of Education to provide data relevant to an authorized investigation or prosecution of an offense concerning national or international terrorism. Any data obtained by the Attorney General for these purposes must be subject to a confidentiality agreement negotiated between the Secretary and the Attorney General.

STANDARD 1: All NCES staff, without exception, must pledge not to release any individually identifiable data, for any purpose, to any person not sworn to the preservation of confidentiality. Individually identifiable data are confidential and individually identifiable data are protected from legal process unless the individual provides written consent, except in the case of terrorism.

STANDARD 2: All contractors whose activities might involve contact with individually identifiable information must provide NCES Project Officers with a list of all staff who might have contact with such data; all such staff must have a signed notarized affidavit of nondisclosure on file at NCES. These affidavits and the staff list must be kept current as new staff members are assigned to NCES projects with individually identifiable information.

STANDARD 3: All contractor staff with access to individually identifiable information must only use that information for purposes associated with the data collection and analysis specified in the contract.

STANDARD 4: Respondents must be told in a cover letter or in instructions that all responses that relate to or describe identifiable characteristics of individuals will be kept confidential, and will be protected to the fullest extent allowable under law. (In the case of NAEP, the legislation extends this protection to the identification of individual schools.) Furthermore, the routine statistical purposes for which the data may be used must be explained.

STANDARD 5: Data files, questionnaires, and other reports having individually identifiable data must be kept secure at all times through the use of passwords, physical

separation of individual identity from the rest of the data, and secure data handling and storage. (See the Restricted-Use Data Procedures Manual, 2000.)

STANDARD 6: When confidentiality edits are used they must be applied to all analytical files (e.g., public-use files, DAS files, and restricted–use files).

STANDARD 7: NCES distributes Data Analysis Systems (DAS) that produce tabular estimates from restricted-use files. In this case, the following conditions must be met:

1. NCES may not release the exact sample size for restricted-use data files that are distributed through a DAS.
2. Only restricted-use data files with Disclosure Review Board (DRB) approved confidentiality edits may be used to produce a DAS.
3. A DAS may not publish unweighted counts.

The confidentiality protection required in a DAS is a function of the type of estimate(s) to be produced. For example, a DAS that produces cell counts may require the use of more extensive confidentiality edits.

STANDARD 8: For public-use data files, NCES minimizes the possibility of a user matching outliers or unique cases on the file with external (or auxiliary) data sources. Because public-use files allow direct access to individual records, Type 1 and Type 2 disclosure limitation techniques may both be required. The Type 1 disclosure limitation techniques must include the techniques applied in a confidentiality edit (if one is performed) and may include additional Type 1 disclosure limitation techniques as well. Thus, the public-use files (i.e., the edited restricted-use files) must undergo a disclosure risk analysis in preparation for release to the public. The steps are as follows:

1. At an early stage in designing and conducting this analysis, staff must consult the Disclosure Review Board (DRB) for guidance on disclosure risk analysis and on the use of NCES disclosure risk software. Any modifications that are necessary as a result of the analysis must be made, and the entire process must be documented.
2. The documentation of the disclosure risk analysis must be submitted to the DRB. The documentation must include descriptions of the risk of disclosure and the types of edits used to avoid disclosure. Decisions over the type of confidentiality edits must take into account the procedures needed to avoid disclosure of individually identifiable information, age of the data, accessibility of external files, detail and specificity of the data, and reliability and completeness of any external files. The documentation should also include the results demonstrating the disclosure risk after adjustments to the data.
3. The DRB will review the disclosure risk analysis report and make a recommendation to the Commissioner of NCES about the file release.
4. The Commissioner then rules on the release of the data file.

STANDARD 9: Inasmuch as confidentiality edits are intended to protect individually identifiable data, files that incorporate the results of the DRB approved confidentiality edit plan may be used to produce tables without confidentiality concerns over minimum cell sizes. When this is done:

1. All versions of a data file must reflect the same confidentiality edits. Staff must consult the DRB on the confidentiality plan, data file dissemination plan (restricted, public use, and/or DAS), and disclosure risk analysis plan, concurrently.
2. Documentation of the confidentiality edit must be included along with the documentation of the disclosure risk analysis that is submitted to the DRB.
3. NCES may not release the exact sample size for restricted-use data files to the public.
4. NCES may not publish unweighted counts for restricted-use data files.

STANDARD 10: A survey program may decide not to apply a confidentiality edit to a restricted-use file (and the associated public-use file). In this situation, when tabulations are produced, any table with a cell with 1 or 2 unweighted cases must be recategorized to insure that each cell in the table has at least 3 unweighted cases. This rule excludes table cells with zero cases because there are no data to protect in the cell. As an example, a principal salary table by race and years of experience may only have 2 Asian respondents with more than 20 years of experience. To implement this standard, one possibility would be to either combine the Asian category with another race group or combine the 20+ years of experience category with the next lower experience category. This process would continue until all cells have either at least 3 unweighted cases or no unweighted cases.

STANDARD 11: At the discretion of the Commissioner of NCES, data security staff may release individually identifiable data to persons for statistical uses compatible with the purposes for which the data were collected. Persons receiving individually identifiable data from NCES shall execute a restricted-use data license agreement, sign affidavits of nondisclosure, and meet such other requirements as deemed necessary in accordance with other confidentiality provisions of the law.

STANDARD 12: Before external data users may gain access to public-use data files, they must agree that they will not use the data to attempt to identify any individual whose data is in the file. This may be accomplished by using the following wording:

"WARNING"

> Under law, data collected by the National Center for Education Statistics (NCES) may be used only for statistical purposes.

Any effort to determine the identity of any reported case by public-use data users is prohibited by law. Violations are subject to Class E felony charges of a fine up to $250,000 and/or a prison term up to 5 years.

NCES does all it can to assure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted or modified in the dataset to protect the true characteristics of individuals. Any intentional identification or disclosure of a person violates the assurances of confidentiality given to the providers of the information. Therefore, users shall:

- Use the data in this dataset for statistical purposes only.
- Make no use of the identity of any person discovered inadvertently, and advise NCES of any such discovery.
- Not link this dataset with individually identifiable data from other NCES or non-NCES datasets.
- To proceed you must signify your agreement to comply with the above-stated statutorily based requirements."

REFERENCE

Restricted-Use Data Procedures Manual. 2000. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington DC: U.S. Government Printing Office.

Include a

SUBJECT: EVALUATION OF SURVEYS

NCES STANDARD: 4-3-02

PURPOSE: The results of the statistical evaluation must enable the users of the survey data to understand the quality and limitations of the data and must provide information for planning future surveys or replications of the same survey. The evaluation should also include a systematic assessment of all sources of error for key statistics that will be studied or reported in NCES publications.

STANDARD: All proposed and ongoing surveys conducted by NCES must include an evaluation component in the survey design plan. The survey evaluation must include the following:

1. Identify the range of potential sources of error;
2. Provide for the measurement of the magnitude of sampling error and sources of the various types of nonsampling error expected to be a problem;
3. Include studies designed to determine what factors are associated with differential levels of error and assess procedures for reducing the magnitude of these errors;
4. Assess the quality of the final estimates, including comparisons to external sources, and where possible, comparisons to prior estimates from the same data collection; and
5. technical report or series of technical reports summarizing results of evaluation studies; for example, a quality profile or total survey error model.

GUIDELINE A: Review past surveys similar to the one being planned to determine what statistical evaluation data have been collected in prior surveys and any potential problems that have been identified. Based on this review, prepare a written summary of what is known about the sources and magnitude of error.

GUIDELINE B: Indicate how each issue will be addressed, including the identification of required data internal and external to the study, a discussion of the comparisons that could be made, the experiments that may be built into the survey, and evaluation methods.

GUIDELINE C: Watch for additional problem areas arising during the course of the survey and, where possible, collect and analyze appropriate data to assess the magnitude of the problem.

GUIDELINE D: Analyze data from the survey evaluation prior to or concurrent with the analysis of the survey data so that the results of the evaluation can be taken into account when processing, analyzing, and interpreting the study data.

GUIDELINE E: List 1 may be used to help guide the development of evaluation plans during the survey planning stage and to develop a monitoring system for possible problems that may emerge during data collection and processing. The list identifies five categories of errors, and enumerates potential sources of error, methods to measure or evaluate them, and possible modifications for correcting them.

LIST 1: MEASURING AND EVALUATING ERROR

Sample Selection, Frames and Coverage - Adequacy of Frame

- Sources of error:
  - Limitations of the frame—undercoverage/overcoverage of schools/institutions, duplicates, cases of unknown eligibility;
  - Listing error—failure of initial respondents to include or exclude prospective respondents per instruction; and
  - Selection of sampling units and respondent units within sampling units.

- Evaluation of survey coverage—examples:
  - Comparison of estimated counts to reliable independent sources;
  - Matching studies to earlier versions of the same data source or to other data sources and the use of dual system estimation;
  - Analysis of survey returns for deaths, duplicates, changes in classification, and out-of-scope units; and
  - Field work - such as area listings.

- Correcting for Coverage error – examples:
  - Use a dual frame approach for survey estimation and
  - Employ post-stratification procedures.

MEASUREMENT ERRORS–DATA COLLECTION

- Sources of error:
  - Questionnaire design, content, wording and instructions;
  - Length of reference period;
  - Interview mode(s);
  - Interviewers—Characteristics, training, and supervision;
  - Respondent rules—self versus proxy respondents;
  - Use of records by respondents;
  - Other respondent effects;
  - Consistency and time-in-sample bias for longitudinal studies;
  - Responses to related multiple measures within a questionnaire;
  - Statistics derived for related measures from different questionnaires within a survey system; and
  - Responses to related measures from multiple respondents in a sampled unit (e.g,. parent/student).

- Evaluation of measurement errors—examples:
  - Pretest or field test survey and procedures;
  - Cognitive research methods;
  - Reinterview studies;
  - Response variance;
  - Randomized experiments;
  - Behavior coding;
  - Interviewer variance studies;
  - Interviewer observation studies;
  - Record check studies; and
  - Comparisons of related measures within questionnaires, across respondents; and across questionnaires within a survey system.

- Correcting for measurement errors—examples:
  - Use the results from a pretest or field test to modify questionnaire and/or procedures;
  - Use input from cognitive research to modify questionnaire;
  - Where possible, use results from comparisons of related measures; and
  - Employ interviewer retraining and feedback.

DATA PREPARATION ERROR

- Sources of error:
  - Pre-edit coding;
  - Clerical review;
  - Data entry; and
  - Editing.

- Evaluation of processing errors—examples:
  - Pre-edit coding;
  - Clerical review verification;
  - Data entry verification;
  - Editing verification for manual edits;
  - Edit rates;
  - Coder error variance estimates; and
  - Rating and scoring error variance estimates.

- Correcting for data preparation errors—examples:
  - Resolution of differences identified in verification;
  - Increased training;
  - Feedback during rating and coding; and
  - Edits for lack of internal agreement, where appropriate.

SAMPLING AND ESTIMATION ERRORS

- Sources of error:
  - Weighting procedures;
  - Imputation procedures; and
  - Sample survey estimation and modeling procedures.

- Evaluation of sampling and estimation errors—examples:
  - Variance estimation;
  - Analysis of the choice of variance estimator;
  - Indirect estimates for reporting sampling error—use of generalized variance functions, small area estimates, and regression models;
  - Comparison of final design effects with estimated design effects used in survey planning;
  - Analysis of the frequency of imputation and the initial and final distributions of variables; and
  - Analysis of the effect of changes in data processing procedures on survey estimates.

- Correcting for estimation errors—examples:
  - Re-estimation using alternative techniques (e.g., outlier treatments, imputation procedures, and variance estimation procedures) and
  - Explore fitting survey distributions to known distributions from other sources to reduce variance and bias.

Nonresponse Errors

- Sources of error:
  - Household/school/institution nonresponse;
  - Person nonresponse; and
  - Item nonresponse.

- Evaluation of nonresponse errors—examples (see Standard 4-4-02):
  - Comparisons of respondents to known population characteristics from external sources;
  - Comparisons of respondents and nonrespondents across subgroups on available sample frame characteristics or, in the case of item nonresponse, on available survey data;
  - Comparisons of characteristics of early and late responding cases;
  - Follow-up survey of nonrespondents for a reduced set of key items to compare with data from respondents; and
  - Descriptions of items not completed, patterns of partial nonresponse, and characteristics of sampling units failing to respond to certain groups of characteristics.

- Correcting for nonresponse errors—examples (see Standards 3-2-02, 4-1-02, and 4-4-02):
  - If response rates are low during initial phases of data collection and funds are not available for intensive follow-up of all respondents, take a random subsample of nonrespondents and use a more intensive data collection method;
  - Use nonresponse weight adjustments for unit nonresponse; and
  - Use item imputations for item nonresponse.

- Methods for reducing nonresponse—examples (see Standards 3-2-02, 4-1-02, and 4-4-02):
  - Employ pretest or embedded experiments to determine the efficacy of incentives to improve response rates;
  - Use internal reporting systems to monitor nonresponse during collection;
  - Follow-up strategies for nonrespondents to encourage participation; and
  - Target a set of key data items for collection with unwilling respondents; and
  - For ongoing surveys, consider separate research studies to examine alternative methods of improving response rates.

REFERENCES

Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., and Sudman, S. (eds.). (1991). *Measurement Errors in Surveys*. New York: John Wiley & Sons.

Bradburn, N.M., Sudman, S., and Associates. (1979). *Improving Interviewing Meth*ods and *Questionnaire Design: Response Effects to Threatening Questions in Survey Research.* San Francisco: Jossey-Bass.

Brick, J.M. and Kalton, G. (1996). "Handling Missing Data in Survey Research." *Statistical Methodology in Medical Research,* 5: 215-238.

Brick, J.M., Cahalan, M., Gray, L., and Severynse, J. (1994). *A Study of Selected Nonsampling Errors in the 1991 Survey of Recent College Graduates.* Washington, DC: U.S. Department of Education, National Center for Education Statistics (Technical Report NCES 95-640).

Federal Committee on Statistical Methodology. (1978). *An Error Profile*: *Employment as Measured by the Current Population Survey*. Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 3).

Federal Committee on Statistical Methodology. (1988). *Quality in Establishment Surveys.* Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 15).

Federal Committee on Statistical Methodology. (1990). *Survey Coverage.* Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 17).

Forsman, G. and Schreiner, I. (1991). "The Design and Analysis of Reinterview: An Overview." In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (eds.) *Measurement Errors in Surveys.* New York: John Wiley & Sons, 279-302.

Gonzalez, M.E., Ogus, J.L., Shapiro, G., and Tepping, B.J. (1975). "Standards for Discussion and Presentation of Errors in Survey and Census Data." *Journal of the American Statistical Association,* 70: 351, Part II.

Groves, R.M. and Couper, M.P. (1998). *Nonresponse in Household Interview Surveys.* New York: John Wiley & Sons.

Groves, R.M.(1989). *Survey Errors and Survey Costs*. New York: Wiley.

Groves, R.M., P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg, (eds.). (1988). *Telephone Methods for Surveys*. New York: Wiley.

Jabine, T.B.; King, K.E.; and Petroni, R. J. (1990). *Survey of Income and Program Participation Quality Profile*. Washington, DC: U.S. Department of Commerce, Bureau of Census.

Kalton, G. (1998). *Survey of Income and Program Participation (SIPP) Quality Profile.* 3rd Edition. Working Paper #230. Washington, DC: U.S. Bureau of the Census.

Kulka, R. (1995). "The Use of Incentives to Survey 'Hard-to-Reach' Respondents: A Brief Review of Empirical Research and Current Research Practices." *Seminar on New Directions in Statistical Methodology.* Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 23, 256-299).

Lessler, J.T. and Kalsbeek, W.D. (1992). *Nonsampling Error in Surveys.* New York: John Wiley & Sons.

Lyberg, L. and Dean, P(1992). "Methods for Reducing Nonresponse Rates: A Review." Paper prepared for presentation at the 1992 meeting of the American Association for Public Opinion Research.

Lyberg, L., Biemer, P., Collins, M, de Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (eds.). (1998). *Survey Measurement and Process Quality.* New York: John Wiley & Sons.

Paxson, M. C., Dillman, D. and Tarnai, J. (1995). "Improving Response to Business Mail Surveys." In Cox et al. (eds.) *Business Survey Methods.* New York: John Wiley & Sons.

Salvucci, S., Walter, E., Conley, V., Fink, S., and Saba, M. (1997). *Measurement Error Studies at the National Center for Education Statistics.* Washington, DC: U.S. Department of Education, National Center for Education Statistics, NCES 97-464.

Sudman, S. and Bradburn, N. (1974). *Response Effects in Surveys: A Review and Synthesis.* Chicago, IL: Aldine.

Sudman, S., Bradburn, N., and Schwarz, N. (1996). *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology.* San Francisco: Jossey-Bass.

U.S. Bureau of the Census and U.S. Bureau of Labor Statistics.(2000). *The Current Population Survey: Design and Methodology.* Technical Paper 63. Washington, DC.

United Nations. (1982). *National Household Survey Capability Programme, Non-Sampling Errors in Household Surveys: Sources, Assessment and Control.* New York: United Nations Department of Technical Cooperation for Development and Statistical Office (preliminary version).

SUBJECT: NONRESPONSE BIAS ANALYSIS

NCES STANDARD: 4-4-02

PURPOSE: To identify the existence of potential bias due to unit and item nonresponse.

DEFINITIONS: **The base weight** is the inverse of the probability of selection when sampling is conducted without replacement.
Nonresponse bias is likely to occur as a result of not obtaining 100 percent response from the selected cases.
The **potential magnitude of nonresponse bias** can be estimated by taking the product of the nonresponse rate and the difference in values of a characteristic between respondents and nonrespondents for means, proportions, or totals.

STANDARD 1: Any survey stage of data collection with a unit or item response rate less than 85 percent must be evaluated for the potential magnitude of nonresponse bias before the data or any analysis using the data may be released. (See Standard 1-3-02 for how to calculate overall and total response rates.) Estimates of survey characteristics for nonrespondents and respondents are required to assess the potential nonresponse bias. The level of effort required is guided by the magnitude of the nonresponse.

STANDARD 2: Nonresponse bias analysis to assess the potential magnitude of unit nonresponse bias must be conducted using base weights for the survey stage with nonresponse. The following guidelines should be considered in such analysis.

> **GUIDELINE A**: Comparisons of respondents and nonrespondents on available sample frame characteristics across subgroups provide information about the presence of nonresponse bias. This approach is limited because observed frame characteristics are often unrelated or weakly related to more substantive items in the survey.
>
> **GUIDELINE B**: Formal multivariate modeling can be used to compare characteristics of respondents and nonrespondents to determine if nonresponse bias exists and, if so, to estimate the magnitude of the bias. These multivariate analyses are used to identify the characteristics of cases least likely to respond to an interview (such analyses are often referred to as nonresponse propensity models). Cases are coded as either responding to or not responding to the interviews and multivariate techniques are used to identify which case characteristics significantly relate to unit nonresponse. The predictor variables should have very high response rates. This approach may be limited by the extent to which such predictors exist in the data.

GUIDELINE C: Comparisons of respondents to known population characteristics from external sources can provide information about how the respondents differ from a known population. This approach is limited by information available from existing sources on the population of interest. Known population characteristics are often unrelated or weakly related to more substantive items in the survey.

GUIDELINE D: For collections in which successive levels of effort (e.g., increasing number of contact attempts, increasing incentives to respond) are employed to reduce nonresponse, comparisons of characteristics can be made between the later/more difficult cases and the earlier/easier cases to estimate the characteristics of the remaining nonrespondents. This approach may be less effective if overall or total response rates are relatively low or if a collection period is relatively short in duration.

GUIDELINE E: More intensive methods and/or incentives can be used to conduct a followup survey of nonrespondents on a reduced set of key measures. Comparisons between the nonrespondent followup survey and the original survey can be made to measure the potential magnitude of nonresponse bias in the original survey. This approach may be costly and less useful for modeling nonresponse bias if the nonrespondent followup survey response rates are also below 70 percent.

GUIDELINE F: The estimated bias can be summarized using the following measures. One measure is the ratio of the bias to the standard error, using the base weight. A second measure is the ratio of the bias to the reported survey mean, using the base weight. If weighting adjustments are used to reduce bias, these measures should also be reported using the final weighted estimates.

STANDARD 3: To analyze potential bias from item nonresponse, the guidelines below must be considered.

GUIDELINE A: For an item with a low total response rate, respondents and nonrespondents can be compared on sampling frame and/or questionnaire variables for which data on respondents and nonrespondents are available. Base weights must be used in such analysis. Comparison items should have very high response rates. This approach may be limited to the extent that items available for respondents and nonrespondents may not be related to the low response rate item being analyzed.

GUIDELINE B: Formal multivariate modeling can be used to compare characteristics of respondents and nonrespondents to determine if nonresponse bias exists and, if so, to estimate the magnitude of the bias. These multivariate analyses are used to identify the characteristics of cases least likely to respond to an item (such analyses are often referred to as nonresponse propensity models). Cases are coded as either responding to or not responding to the item and multivariate techniques are used to identify which case characteristics significantly relate to item

nonresponse. Base weights must be used in such analysis. The predictor variables should have very high response rates. This approach may be limited by the extent to which such predictors exist in the data.

GUIDELINE C: If the overall response rate is acceptable, nonresponse bias analysis may be conducted using data from survey respondents only. Unit level respondents who answered the low response rate item can be compared to unit level respondents who did not answer the item. Final weights and unimputed variables should be used in such an analysis. The comparison items should have very high item response rates. This approach may be limited because it does not directly analyze nonresponse bias that may originate because of unit level nonresponse.

ANALYSIS OF DATA/PRODUCTION OF ESTIMATES OR PROJECTIONS

5-1 Statistical Analysis, Inference, and Comparisons

5-2 Variance Estimation

5-3 Rounding

5-4 Tabular and Graphic Presentations

SUBJECT: STATISTICAL ANALYSIS, INFERENCE AND COMPARISON

NCES STANDARD: 5-1-02

PURPOSE: To ensure that statistical analyses, comparisons, and inferences included in NCES products are based on appropriate statistical procedures.

DEFINITIONS: Statistical inference involves using sample data to test whether an estimated value of a parameter, such as the difference between two means, is sufficiently different from a hypothesized value that the null hypothesis, designated H_0 (no significant difference), can be rejected. The result is the acceptance of the alternative hypothesis, H_1 (significant difference is observed). The following table depicts the relationships between the various error rates in a hypothesis test.

		H_0 (NULL HYPOTHESIS)	
		Actually true	Actually false
Decision	Fail to reject H_0	1-α (correct)	β (type II error)
	Reject H_0	α (type I error)	1 - β (power) (correct)

Type I error is made when the tested hypothesis, H_0, is falsely rejected because the test statistic falls into the region of rejection, when in fact the null hypothesis is true. The probability of making a Type I error is denoted by alpha (α). For example, with an alpha level of 0.05, the analyst will conclude that a difference is present when it is actually not present in 5 percent of tests where the null hypothesis is true.

Type II error is made when the tested hypothesis, H_0, is not rejected because the test statistic does not fall into the region of rejection when in fact the null hypothesis is false. The probability of making a Type II error is denoted by beta (β). The power of a test is 1 - β. It is the probability that the null hypothesis is rejected given that it is false. For example, with a beta level of 0.20, the analyst will conclude, in 20 percent of all cases in which the null hypothesis is false, that no difference is present.

The **tail** of the sampling distribution of the test statistic contains the rejection region for the hypothesis tested, H_0.

The **rejection region** is defined by the alternative hypothesis H_1.

Simple comparison is a test, such as a t-test, of the difference between two means or proportions.

Effect size refers to the standardized magnitude of the effect or the degree of departure from the null hypothesis. For example, the effect size may measure the amount of change over time, or the size of the difference between two populations, divided by the appropriate population standard deviation.

The **minimum substantively significant effect (MSSE)** is the smallest difference considered to be important for the analysis of key variables. The minimum substantively significant effect should be determined during the design phase. For example, the planning document should provide the minimum change in key variables

that the survey should be able to detect for a specified population domain, or subdomain of analytic interest. The MSSE should be based on a broad knowledge of the field, related theories, and supporting literature. In the event an analysis is being done on a data file that does not have a predefined MSSE, broad knowledge of the field, related theories, and supporting literature may be used to determine the MSSE
In analysis, the **effective sample size** is the sample size divided by the design effect.
The **power** of a test is defined as the probability of rejecting the null hypothesis when it is actually false (1- β). In analysis, it represents the probability that the statistical test will be significant given that H_0 is actually false. For example, with $\beta = 0.20$, the power is 0.80. The analyst can expect to reject H_0 80 percent of the time if the parameter value has the specific hypothesized value.

STANDARD 1*:* Statistical analyses must be approached from an analysis plan based on such considerations as relevance to policy, prior findings in existing literature, and/or results of previous survey research. The analysis plan must specify the main research questions, and justify the choice of statistical methodology.

STANDARD 2: Analyses of sample survey data based on a stratified sample design with disproportionate sample allocation must use case weights to correct for the unequal probabilities of selection. NCES sample surveys predominantly use sample designs in which probabilities of selection are unequal. Unless the unequal probabilities of selection are taken into account, and corrected through the use of weights, point estimates will be biased.

STANDARD 3: All statistical tests must be performed at the $\alpha = 0.05$ level of significance. When estimates are compared to one another based on exploratory research and presented in descriptive reports, observed deviations in either direction are of interest and the rejection region lies within both tails of the distribution of the test statistic. The conclusions stated in the text are to be supported by two-tailed tests of significance (t-tests).

> **GUIDELINE:** For some research questions, the region of rejection of the null hypothesis H_0, that is the observation of a significant difference, is contained in only one tail of the sampling distribution. If prior research indicates that differences between estimates would be meaningful only in a single direction or an established trend is to be updated with a new year of data, one-sided tests (t-tests) of significance may be used to optimize power to detect significance.

STANDARD 4: Reported analyses must focus on differences that are substantively important (i.e., it is not necessary, or desirable, to discuss every statistically significant difference in a report). Statistical analysis techniques must be used that are appropriate for the specific research question. The efficacy of individual statistical approaches

depends on the assumptions of the techniques having been met; therefore, the assumptions underlying the techniques must be discussed.

GUIDELINE A: When conducting multiple comparisons, appropriate procedures may be used to control the level of Type I error for simultaneous inferences. Multiple comparison procedures include, for example, Bonferroni, Scheffe, and Tukey tests (see, for example Hochberg, Y. and Tamhane, A.C. 1987. *Multiple Comparison Procedures.* New York: John Wiley & Sons.)

GUIDELINE B: Alternative presentation of the results, such as confidence intervals or coefficients of variation, may also be used as appropriate.

GUIDELINE C: When testing for an increasing or decreasing pattern in data, a trend test should be performed (e.g., regression, ANOVA, Wilcoxen sign test).

GUIDELINE D: When it is appropriate, the use of multiple regression and multivariate analysis techniques should be considered to examine relationships between a dependent variable (e.g., test score) and a set of independent variables (e.g., race, sex, and family background). Such techniques can provide an integrated approach to testing many simultaneous relationships.

GUIDELINE E: If the units of measurement are meaningful (e.g., number of years of schooling), then unstandardized regression coefficients or mean differences should be provided in addition to standardized coefficients.

GUIDELINE F: When the results of an analysis are statistically significant, it may be useful to consider the substantive importance of a difference. For this purpose, the observed difference can be converted into an effect size to allow the interpretation of the size of the difference.

For a t test, the effect size (d_{AB}) is:

$$d_{AB} = \left|(\bar{x}_A - \bar{x}_B)\right| / s$$

$\bar{x}_A - \bar{x}_B$ is the observed difference between estimates.

$$s = \sqrt{\frac{\sum_{i=1}^{n_A} w_i (x_i - \bar{x}_A)^2 + \sum_{i=1}^{n_B} w_i (x_i - \bar{x}_B)^2}{\sum_{i=1}^{n_A} w_i + \sum_{i=1}^{n_B} w_i}}$$

s is the weighted standard deviation of each population. The denominator is the sum of the sampling weights, w_i in both populations. The individual population pieces can be computed in SAS by computing a weighted standard deviation within each population, using the VARDEF=weight option.

The s used in d_{AB} assumes the population variances are different for populations A and B and that $n_A \neq n_B$.

In correlation analysis, *r* is the effect size. Consult Cohen (1988) for measures of effect size using additional statistical procedures.

Cohen's (1988) proposal for interpreting effect sizes may be used where for t-tests an effect size of 0.2 is small, 0.5 is medium, and 0.8 is large; for correlations, an *r* of 0.1 is small, 0.3 is medium, and 0.5 is large.

GUIDELINE G: Another approach to considering the substantive importance of a significant difference is to compare the size of the difference to the minimum substantively significant effect (MSSE) size that is determined *a priori.*

STANDARD 5: When comparing totals, means, or proportions to determine whether differences between groups are significant, the appropriate t-test formula must be used.

GUIDELINE A: The following equations are used with direct estimates and their variances, as opposed to differences between the estimates and the variance of those differences. Because of this, these tests will yield conservative results due to the potential for a missing covariance term:

1. When the estimates of totals, means, or proportions from independent groups are being compared, the formula to be used is:

$$t = \frac{E_1 - E_2}{\sqrt{s_1^2 + s_2^2}},$$

where E_1-E_2 is the difference between the two estimates being compared, and s_1 and s_2 are the estimated standard errors of the estimates computed as detailed in the Standards on Variance Estimation (5-2-02).

2. When comparing estimates from groups that are not independent, the estimated standard error of the difference can be calculated directly, using a statistical package that uses appropriate methods of estimating variances. Alternatively, the *t*-test can be adjusted by estimating the covariance between the two estimates. If the comparison is between the mean of a subgroup and the mean of the total group, the following formula can be used:

$$t = \frac{E_{total} - E_{subgroup}}{\sqrt{s_{total}^2 + s_{subgroup}^2 - 2ps_{subgroup}^2}}$$

Where E is an estimate, s is the estimated standard error of the estimate, and p is the proportion of the total group contained in the subgroup.

3. An alternative formula is derived from the fact that the t statistic for comparing a subgroup to the total sample in the case of counts or proportions is equivalent to the t statistic for comparing a subgroup to its complement, or to the remainder.

$$t = \frac{E_{subgroup} - E_{remainder}}{\sqrt{s_{subgroup}^2 + s_{remainder}^2}}$$

where $E_{\text{remainder}}$ is equal to $E_{\text{total}} - E_{\text{subgroup}}$. Since the subgroup and complement are independent, the correlation term between the two estimates is zero and drops out of the equation.

4. When comparing any two subgroups of a percentage distribution that adds to 100 percent, the formula is:

$$t = \frac{E_1 - E_2}{\sqrt{s_1^2 + s_2^2 - 2rs_1 s_2}}$$

where E_1 and E_2 are estimates, s_1 and s_2 are the estimated standard errors of the estimates, and r is the correlation between the two estimates.

GUIDELINE B: Available software, or analyzing complex sample survey data, may be used to estimate the differences and their variance, and to use these data in significance testing.

STANDARD 6: Failure to reject the null hypothesis does not imply acceptance of the null hypothesis. When the null hypothesis is not rejected, the following options are available:

1. Say nothing.
2. Report that differences were not detected.
3. If the significance is between .05 and .10, and the observed differences are believed to be real, based on research or other evidence, but are not significant at the .05 level due to small sample sizes and/or large standard errors, this may be noted.
4. If the estimate is "unreliable," the reader may be informed that the standard error is so high that the observed large differences are not statistically significant.
5. If a statistically significant difference for a total group under study is observed, but similar subgroup differences of the same magnitude are associated with

smaller sample sizes and/or larger standard errors and are not statistically significant, this may be noted.

REFERENCES

Cohen, B.H. (2001). *Explaining Psychological Statistics*. New York: Wiley.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press

Cohen, J, and Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Hillsdale, NJ: L. Erlbaum Associates.

Hays, W. L. (1994). *Statistics*. Fifth Edition. Fort Worth, TX: Harcourt College Publishers.

Hoenig, J.M. and Heisey, D.M. (2001). "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis." *The American Statistician 55*(1) pp. 19–24.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E., and Nizam, A. (1998). *Applied Regression Analysis and Other Multivariate Methods*. Pacific Grove: Duxbury Press.

NCES Statistical Analysis Manual 2002 (forthcoming). Washington, DC: NCES

SUBJECT: VARIANCE ESTIMATION

NCES STANDARD: 5-2-02

PURPOSE: To ensure that there is planning for estimation of variance in sample surveys and that the appropriate procedures are used and properly documented.

DEFINITIONS: Most NCES sample designs have one or more of the following three characteristics: unequal probabilities of selection, stratification, and clustering.
Strata are created by partitioning the frame; and are generally defined to include relatively homogeneous units within strata.
Clustered samples are those in which a naturally occurring group is first selected, such as a school or a residential block, and then units are sampled within the selected groups.
Simple random sampling (SRS) uses equal probability sampling with no strata or clusters. Most statistical analysis software assumes SRS and independently distributed errors.
A **design effect (DEFF)** is the ratio of the true variance of a statistic (taking the complex sample design into account) to the variance of the statistic for a simple random sample with the same number of cases. **DEFT** stands for the square root of a design effect. Design effects differ for different subgroups and different statistics; no single design effect is universally applicable to any given survey or analysis.

STANDARD 1*:* Variance estimates must be derived for all reported point estimates whether reported as a single, descriptive statistic (e.g., 6 percent of 1988 eighth-graders dropped out of school by 1990) or used in an analysis to infer or draw a conclusion (e.g., more 12th graders took advanced-level mathematics courses in 1998 than in 1982).

STANDARD 2: Variance estimates must be calculated by a method appropriate to a survey's sample design, reflecting sample design characteristics including unequal probabilities of selection, stratification, clustering, and the effects of nonresponse, post-stratification, and raking. These estimates must reflect the design effect resulting from the complex design.

Approximate variance estimation methods that adjust for most of the impact of clustering and stratification include Bootstrap, Jackknife, Balanced-Repeated Replication (BRR), and Taylor-Series Linearization. Replication methods (Bootstrap, Jackknife, and BRR) also adjust for nonresponse, post-stratification, and raking. When replication methods are used, the number of replicates should be large enough to enable stable variance estimation (e.g., ≥ 30) and small enough (e.g., ≤ 100) for efficient calculation.

> **GUIDELINE A**: The preferred way to derive appropriate variance estimates for totals, means, proportions and regression coefficients is to use a statistical package that does not assume simple random sampling (SRS). Such packages include

SUDAAN, WesVar, DAS, or Stata, and use such techniques as Taylor-Series Linearization or one of the replication methods mentioned above.

GUIDELINE B: Consideration should be given to incorporating an adjustment for imputations in variance estimation procedures.

GUIDELINE C: In some cases, alternative approximation strategies can be used to produce variance estimates. For example, software for multilevel models can be used to produce estimates that take into account some aspects of complex survey design. Care must be taken to include any clustering of the sample as a level in the model(s). In addition, any design variables and weights, such as those associated with strata or measures of size, should be taken into account.

STANDARD 3*:* Data files are to contain all relevant information necessary for point estimation and variance estimation (e.g., probabilities of selection, weights, stratum and PSU codes), subject to confidentiality constraints (see Standard 7-1-02 on Machine Readable Data Products).

REFERENCES

Kish, L., Frankel, M. R., Verma, V., and Kaciroti, N. (1995). "Design effects for correlated (P_i-P_j)," *Survey Methodology*, 1995, 21: 117–124 (for an example on design effects for estimates of differences between proportions).

Pfeffermann, D. (1996). "The use of sampling weights for survey data analysis," *Statistical Methods in Medical Research*, 1996, *(5)* pp. 239–261.

Skinner, C. J., Holt, D., and Smith, T. M. F. (Eds.). (1989). *Analysis of Complex Surveys*, New York: Wiley.

Lehtonen R. and Pahkinen, E. J. (1995). *Practical Methods for Design and Analysis of Complex Surveys.* New York, NY: Wiley.

Pothoff, R. F., Woodbury, M. A., and Manton, K. G. (1992). "Equivalent sample size and equivalent degrees of freedom: refinements for inference using survey weights under superpopulation models." *Journal of the American Statistical Association, 87*, pp. 383–396.

Goldstein, H. and Rasbash, J. (1998) Weighting for Unequal Selection Probabilities in Multilevel Models, *Journal of the Royal Statistical Society, Series B, (60)*, pp. 23–40.

Jones, K. (1992). "Using Multilevel Models for Survey Analysis." In Westlake, A. (Ed.), *Survey and Statistical Computing*. New York: North Holland. pp. 231–242.

Goldstein, H. (1991). "Multilevel Modeling of Survey Data." *The Statistician, 40*, pp. 235–244.

SUBJECT: ROUNDING NUMBERS AND PERCENTAGES FOR REPORTING IN TEXT AND DISPLAYING IN SUMMARY TABLES AND FIGURES

NCES STANDARD: 5-3-02

PURPOSE: To ensure consistent practices for rounding and displaying numbers and percentages in text and summary tables/figures.

STANDARD 1: Calculations performed to produce summary data, and computations performed to estimate standard errors must be done on unrounded numbers. The final rounded value must be obtained from the original figure available, not from a series of roundings (e.g., 7.1748 can be 7.175 or 7.17 or 7.2 or 7 but not 7.18). This situation typically arises when researchers round percentages from tables in tenths of a percent to full percents to be used in text.

STANDARD 2: Sums of column or row counts in a table must be derived using unrounded numbers, with appropriate rounding of the total after its derivation. All tables that should logically sum to either 100 percent, or to a numeric total, must include a notes that states: NOTE: Detail may not sum to totals because of rounding."

STANDARD 3: Specific rules for rounding:

If the first digit to be dropped is less than 5, the last retained digit is not changed.

6.1273 is rounded to 6.127

If the first digit to be dropped is greater than or equal to 5, the last digit retained is increased by 1.

6.6888 is rounded to 6.69

5.451 is rounded to 5.5

STANDARD 4: In adding, multiplying, or dividing numbers using data from secondary sources, the result can only be stated in terms of the component number with the fewest significant digits. (For example, if 4.5 and 5.75 are rounded numbers, the product can be stated only as 26, with 4.5 having two significant digits and 5.75 having three.)

STANDARD 5: Before rounding numbers for publication, a decision must be made about the appropriate number of decimal places to be reported. [See *the OERI Publication Guide* (http://www.ed.gov/offices/OERI/MIS/guide]

1. Percentages must be rounded from 4 decimal places.

2. Percentages appearing in text must be rounded to whole numbers unless fine differences require finer breakdowns. Summary tables must be rounded to no more than one decimal place.
3. Percentages appearing in reference and methodological tables must be rounded to no more than two decimal places except in certain methodological tables where finer breakdowns may be necessary.
4. Standard errors must be rounded to one decimal place more than the estimates for which they are computed.
5. Universe data may be reported unrounded. Sample survey data must be rounded.
6. When dealing with small values, a measured zero (i.e., none of something) must always appear in a table or a figure as 0. However, when it is logically impossible to have a response in a cell (i.e., not applicable), that must be denoted by the symbol †. Numbers that round to zero must be represented in tables and figures by the symbol **#.**

GUIDELINE: Numbers appearing in text and summary tables should adhere to the following conventions:

1. Round four- and five-digit numbers to hundreds (e.g., 1,255 is rounded to 1,300; 56,789 is rounded to 56,800);
2. Round six-digit numbers to thousands (e.g., 156,789 is rounded to 157,000); and
3. Round millions and larger numbers to no more than two decimal places (e.g., 1,234,567 is rounded to 1.2 or 1.23 million; 1,912,345,678 is rounded to 1.9 or 1.91 billion).

REFERENCE

OERI Publication Guide. (1999). U.S. Department of Education, Office of Educational Research and Improvement. Washington, DC: U.S. Government Printing Office.

SUBJECT: TABULAR AND GRAPHIC PRESENTATIONS

NCES STANDARD: 5-4-02

PURPOSE: To ensure that tables and graphics displayed in NCES products communicate information accurately, clearly, and efficiently. This will allow the reader to easily and correctly interpret the presentation as a stand-alone display.

DEFINITION: The **reference year** is the year the data were collected about, whereas the **survey year** is the year in which the data were collected.

STANDARD 1: All tables must be produced in accordance with the "NCES Guidelines for Tabular Presentations" (appendix C).

STANDARD 2: Graphics must highlight key points.

STANDARD 3: All figures (graphs, maps, or charts) must be understandable without reference to the text.

- Each figure must have a title concisely stating the subject of the figure and the reference period for the survey.
- Each figure must include all notes necessary to convey information not immediately evident from the main graphic, such as notes that define acronyms, explain special terms, or define the population included in the analysis.

  GUIDELINE: Bar and pie charts should include point estimates for each category displayed.

STANDARD 4: All figures must be consistent with best practices for graphical display. All figures must adhere to the following:

1. Omit distracting detail. For example, avoid the use of three-dimensional effects when only two dimensions are displayed.
2. Be easy to read. For example, all elements (font, lines, labels, symbols, segments, etc.) should be large enough to read with ease in the printed form, easily differentiated, and legible when photocopied or printed in black and white.
3. Be consistent with and prepared in the same style as other figures in the same publication or product. For example, lettering should be of similar size and font, lines of the same weight, symbols, or legends should be used for the same categories.

4. Use consistent scales with consistent spacing when presenting similar units of measurement.
5. With exception of time-series, continuous scales should start with zero or the minimum value of the scale. If used, scale breaks should be clearly visible;
6. When using time-series data, varying time intervals should be plotted on a scale with equal time units and actual data points should be labeled.
7. Include labels for all variables and categories.
8. Clearly label all axes and include tick marks on axes.
9. Prepare figures with patterns, screens, or colors selected to print clearly across different media. In addition, all tables and figures must be in compliance with Section 508 standards that require that information on Web pages be made "accessible" to people with a wide range of disabilities, including vision and hearing impairments, dexterity problems, color blindness and even rare conditions such as photosensitive epilepsy triggered by rapidly flashing lights. For the full text of the law, see:

   http://www.cio.gov/Documents/section%5F508%5Faugust%5F1998%2Ehtml

STANDARD 5: All figures must incorporate a complete source note. A complete source note identifies all the sources relevant to the data presented in the figure.

GUIDELINE A: For figures based on data from one or more reports the Source should cite the report, relevant survey(s) or sub-survey(s), data reference year, file version number, department name, and agency name. In the case of unpublished data, use the month and year of the tabulation or data file. If the data are drawn from multiple years: for one to three years, report each year; for more than three continuous years, use the year span; and for more than three noncontinuous years use "selected years" and the year span. (See appendix D for list of survey titles.)

Following are some typical examples:

Data from one or more reports:
Revenues and Expenditures for National Public Elementary and Secondary Education: School Year 1997-98, Common Core of Data (CCD), "National Public Education Financial Survey" (NPEFS), 1997-98, Version 1, U.S. Department of Education, National Center for
Education Statistics.

Data from unpublished tabulations and a published NCES report:
SOURCE: U.S. Department of Commerce, Bureau of the Census, Current Population Survey, previously unpublished tabulations (April 1998); and U.S. Department of Education, National Center for Education Statistics, *Dropout Rates in the United States.* Selected years 1972–97.

GUIDELINE B: For figures based on data from a compendium report, the source note should cite the compendium report and the original survey or survey report (e.g., *1998 Digest of Education Statistics*, Integrated Postsecondary Education Data System, Fall Enrollment 1997).

GUIDELINE C: For figures based on unpublished tabulations from surveys that are not the main focus of the report, the source note should indicate the data source followed by "previously unpublished tabulation."

GUIDELINE D: For figures based on online data tools, the source note should cite the data source and the data tool.

STANDARD 6: Supporting data for figures must be included in the publication or product. In the case of extracts that only summarize existing publications, supporting data are not required, but summary products must refer to the full report. In the case of short publications (i.e., 15 pages or less), if supporting data are not available in a published report, they must be available on the Web and the publication must refer to the URL. (See Web standards for URL format.)

STANDARD 7: An explanatory note must accompany all figures that include data that may not sum to the expected total (i.e., "Detail may not sum to totals because of rounding.")

STANDARD 8: Figures in the executive summary must be assigned alpha characters consecutively and figures in reports must be assigned numbers. Figures in appendices must be assigned the letter of the appendix and a number suffix (e.g., figures in Appendix A must be labeled A-1, A-2, etc,)

STANDARD 9: Data for the outlying areas must be excluded from U.S. summary totals, unless separate totals are shown.

STANDARD 10: When presenting multiple related figures on one page, a summary title must appear at top of the page and each figure must have its own title. When using multiple related figures from one source on the same page, the source note must be provided at the bottom of the page. When using multiple related figures from different sources on the same page, source notes must be provided for each figure. These source notes must follow the guidelines in Standard 4.

REFERENCES:

Data Documentation Initiative, http://www.icpsr.umich.edu/DDI.

Harris, R. L. (1999). *Information Graphic A Comprehensive Illustrated Reference: Visual Tools for Analyzing, Managing and Communicating.* New York: Oxford University Press.

NCES Guidelines for Tabular Presentation, 2002. Available on the web at http://nces.ed.gov/statprog

OERI Publication Guide. (1999). U.S. Department of Education, Office of Educational Research and Improvement. Washington, DC: U.S. Government Printing Office.

Schmid, C. F. and Schmid, S.E. (1979). *Handbook of Graphic Presentation*. New York: Wiley.

Tufte, E.R. (1983). *The Visual Display of Quantitative Information.* Cheshire, Conn.: Graphics Press.

Tufte, E.R. (1997). *Visual Explanations: Images and Quantities, Evidence and Narrative.* Cheshire, Conn.: Graphics Press.

ESTABLISHMENT OF REVIEW PROCEDURES

6-1 Review of Reports and Data Products

SUBJECT: REVIEW OF REPORTS AND DATA PRODUCTS

NCES STANDARD: 6-1-02

PURPOSE: To ensure that NCES produces and releases high quality products suitable for a variety of audiences.

DEFINITIONS:
Level 1. Review and Adjudication: Requires PD/STA/AC/ODC/OC review and signoff, and outside reviewers are included on the review committee.
Level 1a. Rolling Review: Requires PD/STA/AC/ODC review and approval as parts of the whole are completed. Final product requires full Level 1 review.
Level 2. Statistical Review: Requires PD/AC/ODC review and approval, but no outside review or adjudication. The inclusion of an STA review is at the discretion of the AC.
Level 3. AC/ODC/OC: Requires PD/AC/ODC/OC review and approval, but no outside review or adjudication. The inclusion of an STA review is at the discretion of the AC.
Level 4. AC: Requires PD/AC review and approval, but no ODC/OC or outside review or adjudication.
Level 5. NCES/RIMG/OMB: Requires PD/STA/AC approval within NCES, plus review/approval by RIMG and OMB, and copy to Chief Statistician.
Level 6. Author/Web publisher: Requires full review/adjudication as appropriate for the original NCES numbered product.

STANDARD 1: Prior to the release of a new micro data file, a report presenting the key data contained on the file must be adjudicated and made available to the public. Key data include the major variables that were identified in the analysis plan, and those items that will be maintained over time as part of an NCES data series.

STANDARD 2: All NCES products must be reviewed for technical details and overall quality. The level of review required for each type of product is identified in Table A.

STANDARD 3: Reports requiring Level 1 Review and Adjudication must go through the review procedures outlined in list A and chart A.

STANDARD 4: All NCES Web products/applications require review as outlined in table B.

STANDARD 5: The NCES publication process and related timelines must be documented on the publication sign off sheet (Form A).

Table A. --NCES Products: Required Reviews

Product:	Type of Review Required					
	Level 1. Review and Adjudication	Level 1a. Rolling Review	Level 2. Statistical Review	Level 3. AC/ODC/OC	Level 4. AC	Level 5. NCES/RIMG/ OMB
Compendium	X					
Directory	X					
NCES Handbook	X					
Updated indicators		X				
Pre-release data					X	
Statistical Analysis Report	X					
R&D Report	X					
Technical/Methodological Report	X					
Statistics in Brief	X					
EDTab	X					
Issue Brief/NAEP Facts	X					
Quarterly			X			
Re-packaged Excerpts only			X			
Guide (e.g., Programs & Plans)			X			
Working Papers					X	
Data File (including CD ROM/DAS/WEB)			X			
Data File Documentation /User's manuals (must accompany data file)			X			
Video/Data			X			
Conference Report					X	
Non-data Videotape (e.g., conference, Commissioner's statements)					X	
Brochure/Pamphlet				X		
Newsletters				X		
Co-op Products (e.g., FORUM, NPEC)					X	
Questionnaires						X
Glossaries				X		

Level 1. Review and Adjudication	Requires PD/STA/AC/ODC/OC review and signoff, and outside reviewers are included in the review committee
Level 1a.Rolling Review	Requires PD/STA/AC/ODC review and approval as parts of the whole are completed. Final product requires full Level 1 review.
Level 2. Statistical Review	Requires PD/AC*/ODC review and approval, but no outside review or adjudication.
Level 3. AC/ODC/OC	Requires PD/ AC*/ODC/OC review and approval, but no outside review or adjudication.
Level 4. AC	Requires PD/AC* review and approval, but no outside review or adjudication. No official NCES distribution but made available via web or special request.
Level 5. NCES/RIMG/OMB	Requires PD/STA/AC approval within NCES plus review/approval by RIMG & OMB, and copy to Chief Statistician.
Note:	AC* review may or may not require STA review at the discretion of the AC.

Table B. --NCES WEB Products: Required Reviews

Product:	Type of Review Required				
	Level 1. Review and Adjudication	Level 2. Statistical Review	Level 3. AC/ODC/OC	Level 4. AC	Level 6. Author/Web Publisher
Web Applications:					
NCES Products: (with #)					
pdf file	X				X
Html	X				X
ASCII/ Excel/ data base file*		X			X
Conference Reports/Co-op Products				X	X
Tools:					
Locator			X		X
Peer Tool: Public Access			X		X
Peer Tool: Limited Access*				X	X
Data Tool			X		
Questionnaire Tool			X		
Glossary Search - based on approved product with NCES #)			X		X
Table/ Figure Search			X		
DAS		X			
WEB sites; pages; information sources:					
Survey /Program site			X		X
Web Publications	X				
Quick Facts					X
Video					
Informational Videos			X		X
Data Videos			X		X
PowerPoint Presentations			X		X
Quick tables/figures (quarterly)					X
Unadjudicated Co-op Products				X	X
Working Papers				X	

*	Excludes pre-release data
X	All tools with micro data will be subjected to data snooping tests as well as appropriate review. A full adjudication review is required only for new products. Updates to current products only require review of the update information as appropriate.
Level 1. Review and Adjudication	Requires PD/STA/AC/ODC/OC review and signoff, and outside reviewers are included in the review committee
Level 1a.Rolling Review	Requires PD/STA/AC/ODC review and approval as parts of the whole are completed. Final product requires full Level 1 review.
Level 2. Statistical Review	Requires PD/AC*/ODC review and approval, but no outside review or adjudication.
Level 3. AC/ODC/OC	Requires PD/ AC*/ODC/OC review and approval, but no outside review or adjudication.
Level 4. AC	Requires PD/AC* review and approval, but no outside review or adjudication. No official NCES distribution, but made available via web or special request.
Level 5. NCES/RIMG/OMB	Requires PD/STA/AC approval within NCES plus review/approval by RIMG & OMB, and copy to Chief Statistician.
Level 6. Author/Web Publisher	Assumes full adjudication as appropriate for the original NCES numbered product.
Note:	AC* review may or may not require STA review at the discretion of the AC.

LIST A: KEY STEPS IN THE REVIEW AND ADJUDICATION PROCESS

NCES reports that include data or the analysis of data undergo both internal and external peer review.

PROGRAM REVIEW PROCESS

Decision: NCES Author submits draft report to Program Director for review.
Sign-off: Program Director

DIVISION REVIEW PROCESS

Decision: NCES Author submits draft report to Senior Technical Advisor for review. The Senior Technical Advisor sends signed-off draft to the Associate Commissioner for clearance and to the Chief Statistician for a pre-review.
Sign-off: Senior Technical Advisor, Associate Commissioner, and Chief Statistician

APPROVAL OF PROPOSED REVIEWERS

Decision: NCES Author submits reviewer memo through the Associate Commissioner to the Office of the Commissioner (OC) 3 weeks before the report due to OC date. The reviewers must include two relevant specialists from other NCES programs, and one or more external reviewers for additional subject matter or technical expertise.
Sign-off: Associate Commissioner and Commissioner

SUBMIT REPORT TO THE OFFICE OF THE COMMISSIONER

Decision: NCES Author submits approved peer review list and the publication to the Office of the Commissioner for clearance for distribution for review.
Sign-off: Commissioner

REVIEW BY ASSISTANT SECRETARY

Decision: Five (5) working days for review by the Office for Educational Research and Improvement (OERI).
Sign-off: Assistant Secretary

INTERIM REVISION PERIOD

Decision: Ten (10) working days for author to make revisions requested from OERI.
Sign-off: Associate Commissioner, in consultation with the Chief Statistician and the Commissioner

SCHEDULE ADJUDICATION MEETING

Decision: NCES Author requests a Statistical Standards Program (SSP) chair for an adjudication meeting. After a chair is selected, an adjudication meeting is scheduled.
Sign-off: Chief Statistician

LIST A: KEY STEPS IN THE REVIEW AND ADJUDICATION PROCESS (continued)

DISTRIBUTION FOR REVIEW

Process: NCES Author sends peer review draft to internal and external reviewers. This draft should include supporting documentation for statistical testing. At the same time, the Office of the Commissioner sends the peer review draft for Principal Operating Component (POC) review allowing 2 days for distribution.

Review period: Eighteen (18) or more working days for all reports. NCES Authors are to allow 15 days for peer review, with a request for written comments from reviewers no later than 3 days prior to the scheduled adjudication meeting.

PREPARATION OF REVIEWERS COMMENTS

***Process*:** NCES Author delivers one copy of all POC and peer reviewer comments to the Chief Statistician and one copy to the adjudicator two working days before the scheduled adjudication meeting. To concentrate the adjudication meeting on areas needing resolution; when possible, a pre-adjudication memo should be provided at the adjudication with author agreement and suggested responses to comments.

ADJUDICATION MEETING DECISION

Decision: If, and only if, comments from **all** reviewers are received and are minimal, the author may recommend not holding the adjudication meeting.

Sign-off: Chief Statistician

ADJUDICATION MEETING

Process: The Adjudicator chairs a meeting of the author and reviewers. The Author presents major points from the written comments of reviewers; these are discussed and resolved by the participants. The Adjudicator makes decisions if no consensus is reached. Prior to the end of the meeting, the author is responsible for summarizing the description of all revisions agreed upon during the meeting. The Author obtains assurance from the Adjudicator that the publication with proposed changes will meet NCES standards. Any appeals to decisions may be made to the Chief Statistician. In cases where the revisions result in new analysis and/or extensive rewriting, a second adjudication meeting may be held.

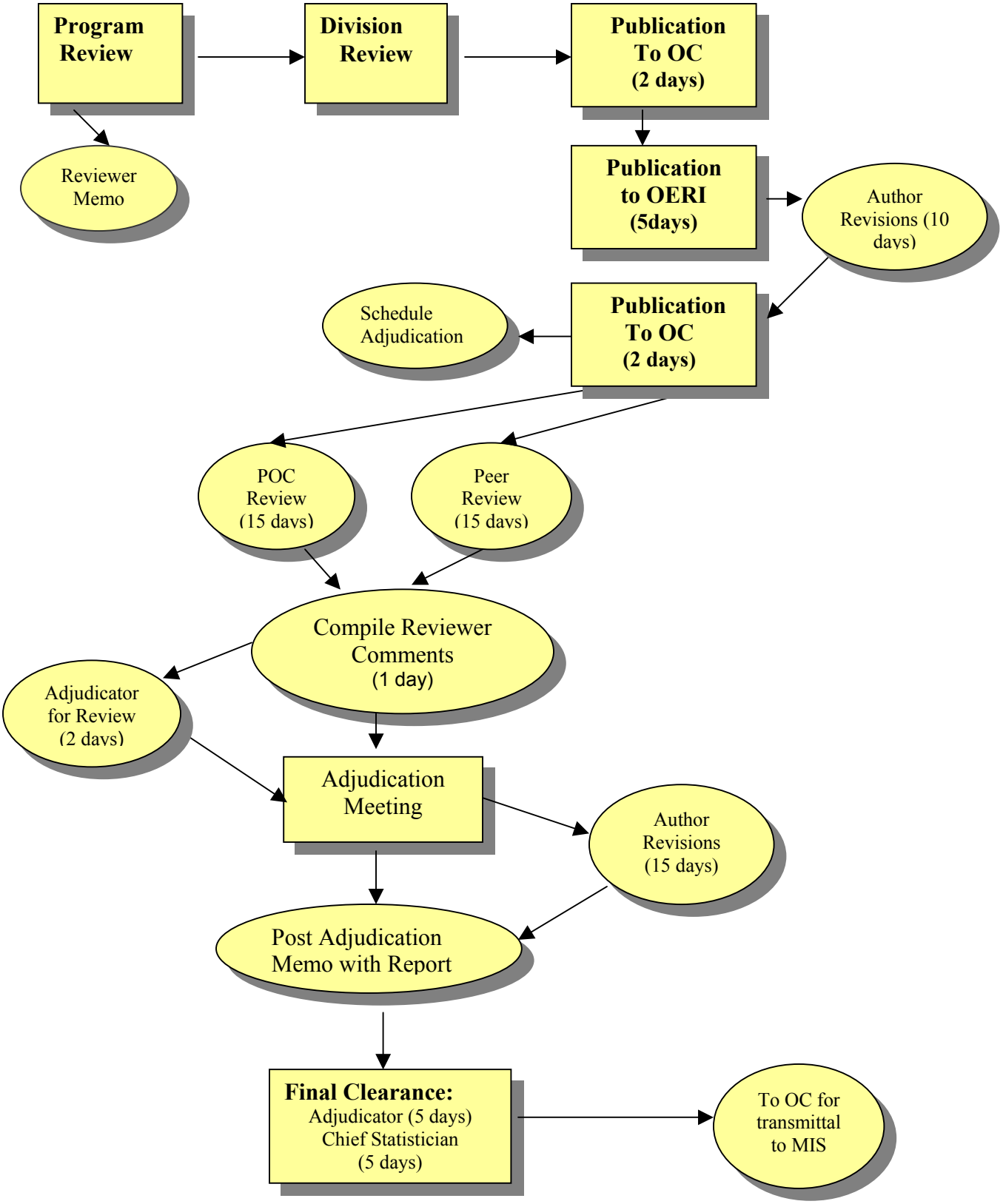
POST-ADJUDICATION REVISIONS AND CLEARANCE

Decision: Within fifteen (15) working days, the NCES author submits the revised publication, along with a post-adjudication memo that describes all changes, to the adjudicator for review.

Sign-off: Chief Statistician, based on recommendation of the Adjudicator.

NOTE: The Commissioner of NCES is the final judge of the content of NCES publications. If the Commissioner delegates this authority, decisions may be appealed to the Commissioner.

CHART A: NCES Publication Review Process



National Center for Education Statistics
Publication Review Form

Pub #: _____

Title: _____

NCES Author: _____ Division: _____ Program: _____ Room: _____

Phone: _____ No. Pages: _____

Adjudicated Publications

User's Manual/Data Files	[ ]
Issue Brief/NAEP Fact	[ ]
Statistics in Brief	[ ]
Statistical Analysis	[ ]
Technical	[ ]
R&D Report	[ ]
Compendium	[ ]
Guide	[ ]
Handbook/Directory	[ ]
ED Tab	[ ]

Other Publications/Products

Non-data video	[ ]
Brochure	[ ]
Pamphlet	[ ]
Newsletters	[ ]
Glossaries	[ ]
Cooperative product	[ ]
Conference Report	[ ]
Working Papers	[ ]
Questionnaires	[ ]
Compilations	[ ]

REVIEWERS	DATE In	Out	DATE In	Out	DATE In	Out	INITIALS
NCES Staff Submits Pub							
Program Director Review							
Review Memo to Program Dir.							
Review Memo to Assoc. Com.							
Review Memo to Commissioner							
Senior Technical Advisor Review							
Associate Commissioner Review							
Chief Statistician (CS) Pre-review *							
Pub to Pub. Coordinator							
Assistant Secretary Review							
Author Submits Revised Pub to AC/Chief Stat./Commissioner							
Schedule Adjudication with CS							
Copies to Pub Coordinator							
Copies to Peer Reviewers							
Reviewer Comments/Memo to Adj.							
Adjudication Meeting							
Post-Adjudication Pub, Memo, Web form, Abstract to Adjudicator							
Post-Adjudication Clearance CS							
Camera Copy to Pub Coordinator							
Announcement to Pub Coordinator							
MIS Review/GPO							
Members Web Form							
PDF to Webmaster							
Pub to GPO							

*To occur concurrent with the Associate Commissioner's review.

DISSEMINATION OF DATA

7-1 Machine Readable Products

7-2 Survey Documentation in Reports

7-3 Release and Dissemination of Reports and Data Products

SUBJECT: MACHINE READABLE PRODUCTS

NCES STANDARD: 7-1-02

PURPOSE: To ensure that data files created by NCES staff and contractors are fully usable by researchers within and outside NCES.

DEFINITION: Metadata contain information about the microdata.
The **reference year** is the year the data were collected about, whereas the **survey year** is the year in which the data were collected.

STANDARD 1: Machine-readable products, including relational databases and spreadsheets, must be released in ASCII format. Each record must contain a unique case identifier such as ID. When there are multiple records per case, each record type must also contain a unique record identifier (e.g., record number, year of data). Data files must be in one of two acceptable formats:

- delimited, text quoted file format that is importable, or
- positional files where the location (i.e., file, record within file, and position within record) of ALL variables are identified.

GUIDELINE A: Data producers are invited to provide additional data sets in alternate formats that may be helpful to users. For guidance on Web-based formats, see the NCES public Web publishing standards; request a copy by sending an e-mail to *NCESwebmaster@ed.gov*.

GUIDELINE B: To facilitate the sharing and use of data elements, national and international standards organizations have produced drafts of several standards for the creation of metadata on data elements. Examples are the International Organization for Standards "Specification and Standardization of Data Elements" standard (ISO/IEC 11179) and the more detailed American National Standards Institute "Metadata for the Management of Shareable Data" Standard (ANSI X3.285). These standards continue to be refined. Data producers should determine what metadata standards are current at the time data files are produced and produce associated metadata for their files that are in compliance with applicable standards.

STANDARD 2: A file description and record layout must be provided for each file. The file information/metadata header must include the following:

1. The title of the survey (survey name, part, and year as applicable);
2. The name(s) of each file;
3. The reference year for the data;
4. The version number and date of release;

5. The logical record length (in positional files) or number of variables on the file (delimited files);
6. The number of records per case or observation; and
7. The number of cases in the data file. For delimited files also include the delimiters (e.g., comma, space).

STANDARD 3: For each variable on the file, the file description must include the following:

1. Variable name;
2. Data type (alpha or numeric);
3. The record number (if multiple records per case); and
4. The position within the record (beginning—end, or variable number if delimited) within the record, field length, and variable label.

STANDARD 4: Data set naming conventions must be standardized and must conform to Information Systems Security Organization (ISSO) 8.3 (or more recent) standards for pressing a CD.

STANDARD 5: Jewel box covers and Web links or URL links must identify the survey system (e.g., HS&B, CCD), component, year of survey, and version number.

STANDARD 6: All variables must be clearly identified and described.

1. The description of variables must include the universe for the variable and all survey items used to construct the variables.
2. In the case of composite variables, the description must include the algorithm used to construct the variables.
3. Upper and lower case labels that clearly describe the variables must be used.
4. For all categorical variables, each value must be associated with a frequency, a percentage of total cases and a label for each category. In public-use and restricted-use file documentation, unweighted frequencies must be included.
5. For all continuous variables, the distribution of values (e.g., minimum, maximum, mean, and standard deviation) must be provided.

GUIDELINE A: FIPS Standards are used where applicable. NCES standard definitions and codes are used where applicable (see Standard 1-4-02).

GUIDELINE B: Variables names should be consistent across surveys within a survey system, within and across years.

GUIDELINE C: In a printable record layout file, line length should be specified so that it prints correctly without wrapping and without special modification (e.g., 72 characters, 12 point type).

STANDARD 7: Data file documentation must be completed for all data files. This includes:

- An abstract or summary that cites the methodology report or technical notes associated with the survey;
- A description of survey methodology that is consistent with the NCES standard for survey system documentation (see Standard3-4-02). In general, survey methodology documentation must include the following:
    1. A description of data collection methods;
    2. Weighting procedures;
    3. Description of editing, error resolution, and imputation flags;
    4. Guidelines for processing the data;
    5. Dates of coverage;
    6. Unweighted frequency counts, and response rates;
    7. Information on how to use replicate weights or PSUs and stratum for variance estimation; and
    8. Procedures for using weights to produce estimates.

STANDARD 8: The following data element conventions must be used:

1. Numeric-fields must contain only numbers or blanks. Reserve codes for numeric fields should be extreme negative values (e.g., lower than the lowest real value).
2. "0" must represent zeros. Blanks or "—" may not be used to represent 0s.
3. Unique values must be used to distinguish between legitimate skips and nonresponse.
4. Suppression symbols must be removed from numeric fields and stored in associated "flag" fields.
5. Separate record locations must be used for all data items.
6. Imputed data must be flagged in associated "flag" fields. Imputation method must be identified in the flag. Blanks are not legitimate values for flags.

GUIDELINE: When practical, numeric data fields containing continuous variables should be identical in length.

SUBJECT: SURVEY DOCUMENTATION IN CENTER REPORTS

NCES STANDARD: 7-2-02

PURPOSE: To provide the appropriate amount of documentation on the survey data, methodology, and other important aspects of a survey in each NCES report. Survey documentation in the report should enable the reader—even the non-statistical user—to understand its contents, and the use and limitations of data readily and clearly.

STANDARD 1: All NCES reports must include documentation that allows the reader to understand the nature and limitations of the results presented. The level of detail included will vary depending on the type of report. The general areas to be covered include: executive summary, status of data, methodology, data collection, and data presentation. The attached list outlines the types of documentation to be included in the various types of NCES reports. “C” for "Complete" indicates the full item is to be included. “B” for "Brief" indicates that a brief description should be included; and “-” indicates not applicable.

STANDARD 2: Standard errors must be available for all estimates included in reports. Standard errors (se’s) or confidence intervals (CI’s) for statistics in tables and graphs can be included in reports in their entirety. In which case, se’s or CI’s for each table for graph are reported either in a separate table in an appendix, or in columns accompanying the statistics being presented. Alternatively, especially for publications that are targeted to general audiences, a separate table of exemplar standard errors on key statistics may be presented in the technical appendix with the detailed standard error tables for all tables and graphs included in a report available on the Web.

Checklist for documentation to be included in NCES reports	Issue Brief, NAEPfact	Compendia	EDTab	Statistics In Brief	R&D Report, Statistical Analysis Report	Survey Technical Report
EXECUTIVE SUMMARY[1]						
History and purpose of the survey	-	B	B	-	B	B
Target population	-	B	B	-	B	B
Time and geographic coverage of the survey	-	B	B	-	B	B
Main findings	-	B	B	-	B	-
STATUS OF DATA						
Identification of data as preliminary, revised or final	-	C	C	C	C	-
Schedule of revisions	-	-	C	C	C	-
Relationship of survey to previous surveys in same series	-	B	C	C	C	C
SAMPLE DESIGN						
Target population	B	B	B	B	C	C
Size of target population	B	-	B	B	C	C
Survey frame, including source of frame, reference date, and number of units	-	-	B	B	B	C
Units selected for sample at each stage	-	-	B	-	B	C
Number of sampling units at each stage	-	-	B	-	B	C
Sample allocation procedure at each stage	-	-	B	-	B	C
Sample selection process at each stage	-	-	B	-	B	C
Total sample sizes[2]	B	-	B	B	C	C
Response rates and their derivations	-	-	B	B	B	C
Measures of size defined for sampling with probability proportional to size	-	-	-	B	B	C
Summary of sources of bias	B	-	B	B	B	C
DATA COLLECTION						
Nature of instruments used, e.g., the contents or kinds of data sought in major sections of the instrument(s) and number of questions in each major section	-	-	B	B	B	C
Method(s) of administering the instrument(s)	-	B	B	B	B	C
Copies of interview scripts/forms/ questionnaire, or copies upon request	-	-	B	B	B	C
Quality control procedures used in data process and results of their implementation	-	-	-	-	-	C
Results of pretest and independent evaluations	-	-	-	-	-	C
Problems, if encountered	-	B	B	-	B	C
DATA PRESENTATION						
Definitions of critical terms/concepts and constructed variables	B	B	B	C	C	C
Supporting numbers for graphs	C[3]	C	C	C	C	C
Selected exemplar standard errors for tables and graphs	-	B	B	B	B	-
Full standard errors for tables and graphs available on the Web	C	C	C	C	C	C[4]

NOTE: The above list outlines the types of documentation to be included in the various types of NCES reports. "C" for "Complete" indicates the item is to be included. "B" for "Brief" indicates that a brief description should be included. "-" means not applicable.

[1] Required if report is longer than 15 pages.
[2] Can be rounded to nearest 100 for restricted data files.
[3] Numbers not included in graphics in the report must be cited to an existing report.
[4] Standard error tables must be included in technical reports.

SUBJECT: RELEASE AND DISSEMINATION OF NCES REPORTS AND DATA PRODUCTS

NCES STANDARD: 7-3-02

PURPOSE: To ensure that all NCES products are disseminated in ways which help to promote the widespread use of NCES data, and to increase the awareness of NCES data among potential users.

STANDARD 1: All NCES products must be disseminated according to a plan that identifies intended and potential users.

GUIDELINE A: To ensure that the contents of a product reflect the needs of intended users, authors should consider user needs early in the publication development process.

GUIDELINE B: In designing a publication or product, the author should consider the Web presentation of the final product.

GUIDELINE C: Once a product has been approved for release by the Chief Statistician, an author should arrange a meeting with OC to review proposed dissemination strategies including press releases, targeted mailings, the number of copies to be printed, Web release, the use of print on demand, and the use of both print and electronic announcements.

GUIDELINE D: Innovative ways to disseminate NCES data should be explored. Presentations at annual meetings, seminars on specific publications, training on the use of data bases, outreach to external groups, and special research efforts using NCES data should be encouraged.

GUIDELINE E: NCES should have strategies in place to collect user feedback on the utility of its products and solicit recommendations for making NCES data more useful.

STANDARD 2: NCES products should utilize a variety of dissemination techniques, as outlined in Table A. All publications must be produced in PDF format, and all mandatory publications must also be produced in HTML format.

GUIDELINE: Efforts should be made to produce other publications in HTML format as well.

Table A. --NCES Products: Required Product Formats

Product:	Type of Product: Print			WEB Product			WEB Tool		
	MIS/GPO	Article in Quarterly	Print on Demand	pdf file	html	ASCII/ Excel/ data base file	Locator	Peer Tool: Public Access	Peer Tool: Limited Access
Standard Products:									
Compendium	XX	XX		XX	1				
Directory	XX	XX	X	XX			X	X	
NCES Handbook	XX	XX	X	XX	X				
Updated indicators				XX	1				
Pre-release data						X			X
Statistical Analysis Report	X	XX	X	XX	X/1				
R&D Report	X	XX	X	XX	X				
Technical/Methodological Report	X	XX	X	XX	X				
NAEP Facts	X	XX	X	XX	X				
EDTab	X	XX	X	XX	X				
Issue Brief	X	XX	X	XX	X				
Quarterly	XX		X	XX	XX				
Re-packaged Excerpts only	X		X	XX					
Guide (e.g., Programs & Plans)	X	XX	XX	XX	1				
Working Papers			XX	XX					
Data File (including CD ROM/DAS/WEB)	XX - Restricted X - Public					XX	2	2	2
Data file Documentation/User's Manuals (must accompany data file)	X			XX	X				
Video/Data									
Conference Report	X		X	XX	X				
Non-data Videotape (e.g., conference, Commissioner's statements)									
Brochure/Pamphlet	XX			XX					
Newsletters				X					
Co-op Products (e.g., FORUM, NPEC)	X		X	X					
Questionnaires				XX	X				
Glossaries				XX	X				

XX	Must be produced for this format
X	Consider producing in this format
1	Required for all Priority 1 publications, optional others
2	Suggested for Universe Files-any format

APPENDIX A

All combinations of 5 races and 1 ethnicity (64 combinations)

[Single Race]	Hispanic or Latino (cell number)	Not Hispanic or Latino (includes not reported)
White	1	33
Black or African American	2	34
Asian	3	35
American Indian or Alaska Native	4	36
Native Hawaiian or Other Pacific Islander	5	37
[Combination of Two Races]		
White and Black or African American	6	38
White and Asian	7	39
White and American Indian or Alaska Native	8	40
White and Native Hawaiian or Other Pacific Islander	9	41
Black or African American and Asian	10	42
Black or African American and American Indian or Alaska Native	11	43
Black or African American and Native Hawaiian or Other Pacific Islander	12	44
Asian and American Indian or Alaska Native	13	45
Asian and Native Hawaiian or Other Pacific Islander	14	46
American Indian or Alaska Native and Native Hawaiian or Other Pacific Is	15	47
[Combination of Three Races]		
White and Black or African American and Asian	16	48
White and Black or African American American Indian or Alaska Native	17	49
White and Black or African American and Native Hawaiian or Other Pacific Islander	18	50
White and Asian and American Indian or Alaska Native	19	51
White and Asian and Native Hawaiian or Other Pacific Islander	20	52
White and American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander	21	53
Black or African American and Asian and Native Hawaiian or Other Pacific Islander	22	54
Black or African American and Asian and American Indian or Alaska Native	23	55
Black or African American and Native Hawaiian or Other Pacific Islander and American Indian or Alaska Native	24	56
Asian and Native Hawaiian or Other Pacific Islander and American Indian or Alaska Native	25	57
[Combination of Four Races]		
White and Black or African American and Asian and American Indian or Alaska Native	26	58
White and Black or African American and American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander	27	59
White and Asian and American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander	28	60
White and Black or African American and American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander	29	61
Black or African American and Asian and American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander	30	62
[Combination of Five Races]		
White and Black or African American and Asian and American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander	31	63
[No race specified or refused]		

APPENDIX B

EVALUATING THE IMPACT OF IMPUTATIONS FOR ITEM NONRESPONSE

Marilyn Seastrom, Steve Kaufman, Ralph Lee

An incomplete data record for a survey respondent results in item nonresponse that cannot be ignored. Survey nonresponse can result in an increase in the mean square errors of survey estimates and a distortion of the univariate and multivariate distributions of survey variables, and thus may result in biased estimates of means, variances, and covariances (OMB, 2001).

Measuring Bias

The degree of nonresponse error or bias is a function of two factors: the nonresponse rate and how much the respondents and nonrespondents differ on survey variables of interest. For example, in the case of item nonresponse on family income, a comparison of the characteristics of the respondents and nonrespondents on other items that were completed by the item nonrespondent can be used to assess whether there are any systematic differences. In the case of our example, parent's education, parent's occupation, and race-ethnicity (or a longer list) might be good candidates to examine for an indication of the amount of bias associated with the missing income data.

The mathematical formulation to estimate bias for a sample mean is:

$$B(\bar{y}_r) = \bar{y}_r - \bar{y}_t = \left(\frac{n_m}{n_t}\right)(\bar{y}_r - \bar{y}_m)$$

where:

$\bar{y}_t$ = the mean based on all sample cases

$\bar{y}_r$ = the mean based only on respondent cases

$\bar{y}_m$ = the mean based only on nonrespondent cases

n_t = the number of cases in the sample (i.e., $n_t = n_r + n_m$)

n_m = the number of nonrespondent cases

n_r = the number of respondent cases

$\bar{y}_r$ is approximately unbiased if either the proportion of nonrespondents (n_m/n) is small or the nonrespondent mean, $\bar{y}_m$, is close to the respondent mean, $\bar{y}_r$.

The relative bias provides a measure of the magnitude of the bias:

$$\text{Rel } B(\bar{y}_r) = \frac{B(\bar{y}_r)}{\bar{y}_r}$$

Where:

Rel $B(\bar{y}_r)$ = the relative bias with respect to the estimate, $\bar{y}_r$.

The bias ratio provides an indication of how confidence intervals are affected by bias:

$$\text{Bias Ratio} = \frac{B(\bar{y}_r)}{\sigma_{\bar{r}}}$$

Where:

$c_{\bar{r}}$ = the standard error.

Next, since the estimate total for variable *y* is the sum of the estimates for the respondents and the nonrespondents:

$$y_t = y_r + y_m$$

which is also equal to the product of the number of respondents times the mean value for the respondents added to the number of nonrespondents times the mean value for nonrespondents:

$$y_t = y_r + y_m = n_r \overline{y_r} + n_m \overline{y_m}$$

The bias for the estimate of a total, y_r, is:

$$B(y_r) = y_r - y_t = -y_m = -n_m \bar{y}_m$$

Thus, the bias is small if the number of nonrespondents is small or if the mean for nonrespondents is low.

The bias for an estimate of variance is:

$$B(s_r^2) = \left[\left(\frac{n_m}{n_t}\right)\left(s_r^2 - s_m^2\right)\right] - \left[\left(\frac{n_r}{n_t}\right)\left(\frac{n_m}{n_t}\right)\left(\bar{y}_r - \bar{y}_m\right)^2\right]$$

Note first, that the first term is similar to the equation for the bias of the mean, in that it is the product of the nonresponse rate and a difference—in this case the difference is that between the variance of the respondents and the nonrespondents. The second term is the product of the response rates for respondents and nonrespondents and the squared difference between the means for the respondents less the nonrespondents.

Suppose the variances for respondents and nonrespondents are similar (a more reasonable assumption than assuming this for the means), then the nonresponse rate times zero or a small difference is negligible. When this is the case, the bias in the variance is a function of the product

of the response and nonresponse rates and the contribution from the squared difference in the mean values for respondents less nonrespondents. In other words, the bias in the variance is a function of the amount of nonresponse and the difference in the means for respondents and nonrespondents and it will will always result in an underestimate of the variance.

Consider the example in which the variance is the same for respondents and nonrespondents and the response rate is 70 percent. The bias formula reduces to the second term:

$$\mathrm{B}(s_r^2) = -\left[\left(\frac{n_r}{n_t}\right)\left(\frac{n_m}{n_t}\right)\left(\bar{y}_r - \bar{y}_m\right)^2\right]$$

The product of the response rates is .21 and the squared difference of the means, some value *z,* will be positive regardless of which mean is larger. The bias is then equal to:

$$\mathrm{B}(s_r^2) = \text{-}.21(z).$$

If the variances of the respondents and nonrespondents are the same, the variance will always be underestimated.

However, in some cases the variances associated with respondents and nonrespondents may not be equal. For example, consider the case of income reporting where nonrespondents are likely to be concentrated at the upper and lower ends of the distribution, leaving the respondents more clustered in the middle. It will result in a larger variance associated with the nonrespondents than the variance for the respondents. Thus the difference between the two variances will be negative. Continuing with the earlier example, the bias for an estimate of the variance becomes:

$$\mathrm{B}(s_r^2) = .30(\text{-j}) - [.21(\mathrm{z})] = .30(\text{-j}) - .21(\mathrm{z})$$

Where j is the difference between the two variance estimates. Again, the variance is underestimated. In fact, j is likely to always be smaller than z, since variances decrease as the sample size increases. While the differences in the means are not affected by sample size and as a result are likely to be larger in large scale surveys. Thus, more of the bias is due to the differences in the means and the variance will always be underestimated.

For the bias for an estimate of covariance, consider the case where respondents are defined as those who answered both y and a second variable x. Here r' is the number of cases with answers to both x and y, with the prime used to indicate the joint response. The bias for an estimate of covariance is:

$$B\left(s'_{rxy}\right) = \left[\left(\frac{n_m}{n_t}\right)\left(s'_{rxy} - s'_{mxy}\right)\right] - \left[\left(\frac{n_r}{n_t}\right)\left(\frac{n_m}{n_t}\right)\left(\bar{x}'_r - \bar{x}'_m\right)\left(\bar{y}'_r - \bar{y}'_m\right)\right]$$

If $s'_{wy} = s'_{mwy}$ the covariance is not necessarily underestimated. When the estimates of covariance are equal for respondents and nonrespondents, the bias will be negative (i.e. an underestimate of the

covariance) if the signs on $\left(\bar{x}_r' - \bar{x}_m'\right)$ and $\left(\bar{y}_r' - \bar{y}_m'\right)$ are both positive or both negative the bias will be negative and the covariance will be underestimated. On the other hand, if these two terms have opposite signs the bias will be positive and the covariance will be overestimated.

The Problem with Ignoring Item Nonresponse

The reason item nonresponse cannot be ignored is because once it exists, any analysis of the data item requires either an implicit or explicit imputation. To ignore the missing data and restrict analyses to those records with reported values for the variables in the analysis, implicitly invokes the assumption that the missing cases are a random subsample of the full sample, that is, they are missing completely at random (MCAR). This means that missingness is not related to the variables under study. This requires that all respondents are equally likely/unlikely to respond to the item and that the estimate is approximately unbiased. These are strong assumptions. As noted by Brick and Kalton, 1996, "The use of imputation can improve on this strategy."

Little and Rubin included a discussion of "Quick Methods for Multivariate Data with Missing Data" in their 1987 book *Statistical Analysis with Missing Data.* In introducing these methods they state "Although the methods appear in statistical computing software and are widely used, we do not generally recommend any of them except in special cases where the amount of missing data is limited." Included in this discussion are complete-case analyses where only the cases with all variables specified in the analysis included (i.e. the number of cases is fixed for all variables in an analysis) and available-case methods that include all cases where the variable of interest is present (i.e. the sample base changes from variable to variable). They conclude this discussion by stating "Neither method, however, is generally satisfactory."

Lessler and Kalsbeek also explored a variety of imputation methods in their 1992 book, *Nonsampling Errors in Surveys.* While they caution that there is no substitute for complete response, "…it is better when attempting to reduce nonresponse bias to use a well-chosen method than to do nothing at all, unless the rate of nonresponse is low."

Examples

A few numerical studies can help illustrate this point. Lessler and Kalsbeek, 1992 reported on a 1978 analysis that they conducted on data from the National Assessment of Educational Progress (NAEP). Their goal was to measure the effect of nonresponse on 17-year-old students, since they have lower response rates than the 13- or 9-year-old students. Their comparison of data from a subsample of nonresponding 17-year-olds with data from the original group of sample respondents showed that the size of the nonresponse bias relative to the variance component of most estimates in this survey was high. They noted that since bias does not depend on sample size, but variance diminishes as the sample size increases; nonresponse bias tends to be significant for large surveys. They also observed a direct relationship between the extent of nonresponse bias and a lowering of the actual confidence levels.

A second example may be drawn from "A study of selected nonsampling error in the 1991 Recent College Graduates Study," (U.S. Department of Education, 1995). The estimate of interest is the

percent of graduates with a bachelor's degree who are education majors. Although technically the institution is the first stage of sample selection and the graduate is the second stage, for the purposes of this example the institution will be taken as the respondent and the item nonresponse is determined by whether the graduate responded or not. The institution response rate of 95 percent is posited to allow for a relatively accurate estimate of the item nonresponse bias.

The nonresponse rate for graduates was 16.4 percent. The institutions reported data showing that 7.79 percent of the nonrespondents majored in education, compared to 10.54 percent of the respondents. The bias can be estimated as:

$$[.164*(.1054 - .0779)] = .00451 = 0.5\%$$

In other words, if the estimate were based only on the respondents, it would overestimate the percentage who are education majors by one-half a percent.

The relative bias with respect to the estimate, is:

$$(.00451/.1054) = .0428 = 4.3\%$$

Thus, the bias is relatively small in this case. However, when the bias ratio is considered, a different picture emerges. In general, a bias ratio of 10 percent or less has little effect on confidence intervals or test of significance. That is to say, with a bias ratio of 10 percent, the probability of an error of more than 1.96 standard deviations from the mean is only 5.11 percent, compared with the usual 5 percent (table 1). In the graduate example, when the estimate of bias is compared to the standard error, the bias ratio is:

$$(.00451/.0003047) = 14.8 = 148\%$$

The bias ratio of 148 percent means that that there is a 32 percent chance of a Type I error, (i.e., rejecting a true hypothesis) in computing the confidence interval or conducting a significance test in this example.

This bias ratio is so large because the estimated standard error is small, as is typically the case with large sample sizes. Thus, although the actual bias and the relative bias are relatively small, the bias ratio illustrates the fact that the impact on statistical inferences can still be quite large. This has important implications for Federal statistical agencies that conduct large sample surveys.

If we assume that the variance associated with the estimate of education majors is the same for respondents and nonrespondents. Then, the bias of the variance estimate in this example is:

$$B(s_r^2) = - [(.164)(.836)](.1054 - .0779)^2 = - .000104$$

The variance in this example is underestimated by .01 percent.

Table 1.— Bias ratio by size of probability of a Type I error

Bias Ratio (Percent)	Probability of Type I error
2	.0500
4	.0502
6	.0504
8	.0508
10	.0511
20	.0546
40	.0685
60	.0921
80	.1259
100	.1700
150	.3231

Cochran, 1977

Explicit Methods of Imputing for Item Nonresponse

The alternative to ignoring missing item responses is to adopt a strategy to "fill-in," or in other words, impute the missing responses. A number of different methods have been proposed and used in survey research. Before discussing the specific methods and the relative advantages and disadvantages of each one. It is worthwhile to consider the pros and cons of explicit imputations in general.

Most authors in this area caution that imputations carry both potentially positive and negative outcomes. For example, Kalton and Kasprzyk, 1982, identified three positive aspects of explicit imputations. They are intended to reduce biases from item nonresponse in sample survey data. By filling in the holes, they allow analyses to proceed as though the data set were complete, thus making analysis easier to conduct and results easier to report. They result in consistent results across analyses, because all analysts should be working with the same set of "complete" cases. They also identified potential drawbacks. They cautioned that imputation methods do not necessarily lead to a reduction in bias, relative to the incomplete data set. And, they warned against the danger of analysts treating the "complete" cases as actual responses, thus overstating the precision of the survey estimates. Brick and Kalton, 1996, concur with these statements and add that imputation methods may also distort the association between variables. They note that although methods can be selected to maintain the associations of the variable subject to imputation with certain key variables, associations with other variables may be attenuated.

Imputations can be categorized along two dimensions. First, by whether they are deterministic or stochastic. In the case of deterministic imputations, the residual term is set to zero. This yields the best prediction of the missing value, however it results in an attenuation of the variance of the imputed estimate relative to that of the unobserved estimate and it distorts the distribution of the values of the item in question. Thus deterministic imputations give more precise estimates of means (e.g. an average score), but produce biased estimates of distributions (e.g. the percent of students scoring above a certain point). In stochastic imputations, the residual or error term is randomly assigned. This addition of random noise improves the shape parameters by yielding more

realistic distributions. Brick and Kalton, 1996, concluded that given "the importance of shape parameters in many analyses, stochastic imputations are generally preferred."

The second dimension has to do with whether or not auxiliary variables are used in the imputation method. Within the set of imputation methods that use auxiliary variables, they may be either categorical, categorizing sample members into imputation classes, or they may be continuous, as in the case of regression imputation methods.

As mentioned earlier, a number of different types of imputation methods have been developed and used in survey research. A partial, although probably not complete, listing includes historical imputation, deductive imputations, mean imputations, random imputation, overall mean imputations within classes, random imputation within classes, hot-deck imputation, cold-deck imputation, flexible matching imputation, ratio imputation, predicted regression imputation, random or stochastic regression imputation, EM algorithm imputation, distance function matching, composite methods, Bayesian Bootstrap imputation, and multiple imputation methods. There are a number of sources that review the methods and properties of these varied imputation techniques (Little and Rubin, 1987; Kalton, 1983; Kalton and Kasprzyk, 1982, 1986; Lessler and Kalsbeek, 1992; HU, Salvucci, and Cohen, 199).

The rest of this discussion will focus on those methods that are either currently used at NCES or the most promising alternatives for future work.

Table 2, taken from a forthcoming NCES report by Salvucci, et.al, shows the imputation methods used in recent NCES data collections. In the case of the universe data collections (CCD, PSS, IPEDS) the imputation methods most used include ratio imputation, mean imputation, and cold-deck imputation. In a few cases deductive or logical imputations are employed, and hot-deck imputation methods are also used in a few cases. Historical imputations should be added to this list, inasmuch as they are used in the Digest of Education Statistics and perhaps in the Condition of Education.

Table 2.—Imputation methods employed in NCES data collections

Survey	Imputation methods
CCD	Ratio imputation and adjustment
PSS	Ratio adjustment, deductive and sequential hot-deck imputation
IPEDS-IC	Mean and ratio imputation
IPEDS-EF	Mean and ratio imputation
IPEDS-C	Mean, ratio, and cold-deck imputation
IPEDS-SA	Within class mean and ratio imputations
IPEDS-F	Ratio adjusted cold-deck and sequential hot-deck imputation

IPEDS-S	Ratio adjusted cold-deck and hot-deck imputation
IPEDS-L	Logical imputation, ratio adjustment
IPEDS-ALS	Ratio and cold-deck imputation
NSOPF	Sequential hot-deck and within-class random imputation
SASS	Deductive and sequential hot-deck imputation
SASS-TFS	Deductive and sequential hot-deck imputation
RCG	Deductive, hot-deck, and within-class random imputation
NHES	Manual and hot-deck imputation
NPSAS	Deductive, hot-deck, and regression imputation
FRSS	Mean, median, and sequential hot-deck imputation
PEQIS	Ratio adjustment and sequential hot-deck imputation
NAEP	Multiple imputation based on Bayesian models for scores
TIMSS	Multiple imputation based on Bayesian models for scores

The sample survey data collections primarily use sequential hot-deck imputation along with deductive imputations. There has also been limited use of within-class random imputation, regression imputation, multiple imputation, and a few of the methods listed above under universe data collections.

Deductive or logical imputations

Sometimes the value of a missing item can be logically deduced with certainty from responses to other items. It is unclear whether this should be considered a form of imputation or a form of data editing. If strict rules of logic are followed, then the value is clear and has no impact on any of the resulting statistics. While deductive imputation is the ideal form of imputation, it is frequently not possible. Some argue that these data corrections are best treated as edits.

Historical imputations

Historical imputations are used for variables that tend to be stable over time (e.g. the number of teachers in a state). This method uses previously reported data from the same unit to impute for missing data in a current data collection. This method attenuates both the size of trends and the incidence of change. A variation on this method helps correct for these problems, by using some measure of trend, frequently derived from other cases.

This method works best when the relationship over time is stronger than the relationship between variables at one point in time.

Cold-deck imputation

Cold-deck imputation uses a constant value from a source external to the current data collection to "fill-in" the missing item. Frequently a previous iteration of the same survey serves as the external source. Little and Rubin, 1987, acknowledge that current practice is to ignore these imputations, treating these data as a complete sample. They go on to state that there is no satisfactory theory for the analysis of data obtained by cold deck imputation. Lessler and Kalsbeek, 1992, describe cold-deck imputation as being of historical interest, but rarely used in practice. This method seems to be very close to historical imputations.

Mean value imputation

Mean value imputation uses the mean of the reported values to "fill-in" the missing value. In the case of overall mean value imputation, the mean is taken from the entire distribution; while in within-class mean value imputation the mean is taken from the specific imputation class. (Median value imputation is very similar, using the median of the reported value.)

This method can only provide unbiased estimates for means and totals if the missing values meet the strong assumption of missing completely at random. Because this procedure creates a spike at the mean value, it does not preserve the distribution or the multivariate relationships in the data. Furthermore, because the sample size is effectively reduced by nonresponse, standard variance formulas will underestimate the true variance. Overall mean value imputation is not recommended. Kovar and Whitridge in Cox et. al., 1995 caution that if all else fails, within-class mean value imputations can be used with carefully chosen classes for means and totals, but that it does not work for other statistics. Salvucci et. al., 2001 point out that if the missing values depend on any variables not included in the auxiliary variables used to form the imputation class, the means and totals will be biased, the distribution will be distorted, and the variances will be substantially underestimated. Little and Rubin, 1987, make the point that the distortion of the distribution is particularly problematic when the tails of the distribution or the standard errors of the estimates are the focus of study.

Ratio Imputations

Ratio imputations, like within-class mean value imputations, use auxiliary variables that are closely related to the variable to be imputed and that have data available for all or nearly all of the sampled units. The imputed value for case *i* is obtained by multiplying the ratio of the mean for the responding cases for the variable to be imputed to the mean of all cases for the auxiliary variable times the case *i* value for the auxiliary variable. The requirement for a highly correlated auxiliary variable can yield accurate imputations, but it is more often the case that the variable to be imputed is correlated to several auxiliary variables. Thus a ratio imputation that is, by definition, tied to one auxiliary variable is not fully efficient. In addition, if the auxiliary item is identical across several units used in the imputation, the related imputed items will mirror that pattern, thus distorting the distribution of the imputed variable.

It is important to note here that the ratio imputations used by at least some NCES data collections do not follow this description exactly. Instead, what is done for example with state level fiscal data in CCD, is to partition the responding cases, remove the value of the variable in question from the total for each state, compute the ratio of the value for each responding state to their reduced total,

compute the average of these ratios across all responding states, and then multiply the total for each state with missing data by the average ratio.

Regression Imputation

Predicted regression imputation is very closely related to the ratio imputation approach, the primary difference being that a set of highly correlated auxiliary variables are used to predict missing values in the imputed variable. In this case the imputed values are only as good as the model used to predict them. Random regression imputation follows the same procedures used in predictive regression imputation, with the addition of a stochastic component through the residual terms. There are several alternative assumptions that can be used to define the way these residual terms are generated in an imputation procedure—normally distributed, chosen at random from the respondent's residuals, or chosen at random from respondents who are similar on the auxiliary variable. One drawback that is unique to regression imputations is their ability to yield improbable results.

In this case, as in other forms of imputation, the component of variance that is attributable to survey nonresponse is not accounted for in standard variance estimation software; resulting in an underestimation of the true variance.

Hot-Deck Imputation

Hot deck originally got its name from the decks of computer cards that were used in processing data files, with the term hot referring to the same data file. There is actually a class of imputation procedures that share this label. The common thread is that missing values are replaced one at a time with an available value from a similar respondent in the same study. This is general approach is probably the most widely used imputation method. One of the reasons there is variability among types of hot-deck methods, is that its popularity has caused it to evolve. In general, the procedures starts with a set of imputation classes and the cases within each class are processed and compared. This procedure preserves the distribution of the estimates, and increases the variance relative to the mean imputation method. Thus, the underestimation of the variance of the estimate is decreased.

In the case of the sequential hot-deck imputation each class starts with a single value for the item subject to imputation; each record is compared to that item, if the record has a value for that item, it replaces the starter value, on the other hand, if the record is missing that item the starter value or the value that has replaced it is "filled-in" on the case with the missing value. One problem occurs with this approach when several records with missing values occur together on the file. This results in the current donor value being assigned to multiple records, thus leading to a lack of precision in the survey estimates (Kalton and Kasprzyk, 1986). A variation on this approach is known as random imputation within classes; the difference here being that the donor respondent is chosen at random within the imputation class for assignment to the nonrespondent. Lessler and Kalsbeek, 1992, pointed out that if this is done with replacement, the multiple use of a donor problem persists; however, they also noted that this can be avoided by sampling without replacement. While this procedure is more cumbersome, it has the advantage of providing a basis to correctly formulate the mean square error of estimators using a hot-deck imputation.

Another way to avoid the problems associated with sequential hot-deck imputation is the hierarchical hot-deck imputation. This method sorts respondents and nonrespondents into a large

number of imputation classes based on a detailed categorization of a large set of auxiliary variables. Nonrespondents are then matched with respondents in the smallest class first, if no match is found that class is collapsed with the next one, and so on until a donor is found—hence the label hierarchical.

As problems have been identified, alternative schemas have been devised to solve those problems. Regardless of the specifics, all hot-deck procedures take imputed values from a respondent in the same data file, thus yielding imputations that are valid, although not necessarily internally consistent for the respondent values. In order to evaluate the hot-deck imputation used for any specific data collection, detailed information is required.

Data Analysis with Imputed Data

This brief review has highlighted the fact that imputed data sets can provide good estimates of means and totals, and that with some care and attention in the selection of the imputation method, the distributions can be reasonably well preserved. However, as Kovar and Whitridge, 1995 point out "The situation is not as favorable when it comes to estimates of variances and correlations." They note that numerous studies have shown that imputations can have a deleterious effect on the statistics of the estimates. In particular, correlations between imputed variables are attenuated to varying degrees, but good auxiliary variables can help this problem (Santos 1981; Kalton and Kasprzyk, 1982, 1986; and Little, 1986).

When standard formulas are used for the computation of statistics for estimates based on imputed data, the variances of estimated means and totals are underestimated (Rubin, 1978). This underestimation occurs because standard computing software treats imputed values for missing data as observed data and thus, ignores the component of variance that is due to imputation. Kovar and Whitridge, 1995, report that standard variance formulas underestimate the variance with imputations present by about 2 to 10 percent with a response rate of 5 percent and by as much as 10 to 50 percent with 30 percent nonresponse. The size of the underestimate varies with different types of imputation.

Brick and Kalton, 1996, discuss two methods for reducing imputation variance. The first method involves the use of sampling strategies. Selecting donors without replacement within each imputation class minimizes the multiple use of donors resulting in a lower imputation variance compared to sampling with replacement. When there is more than one respondent in a class, stratified sampling with a class or systematic sampling from an ordered list can also help reduce imputation variance. The second method relies on fractional imputation. With this approach individual respondent records are divided into parts, with weights distributed accordingly, and separate donors are chosen for each part of the respondent's record.

The underestimation of the variance results in short confidence intervals and a tendency to declare significance when none exists. Sarndall, 1990 demonstrated that these statistical problems become more severe as the amount of missing data increases. Lessler and Kalsbeek, 1992 point out that the size of the nonresponse bias associated with totals, means, variances, and covariances is linked to differences between respondents and nonrespondents.

There are several recently developed techniques designed to estimate the variance due to imputation. Rubin pioneered the use of multiple imputations in this arena, estimating the variance by replicating the process a number of times and then estimating the between replicate variances. Sarndall, 1990, proposed a method using model-assisted estimators of variance. Rao and Shao, 1992, use a method that corrects the usual jackknife variance estimator. Brick, Kalton, Kim and Fuller are currently under contract to NCES, conducting an evaluation of these new methodologies. The Statistical Standards Program at NCES is also supporting work by Aitken on an alternative approach using the EM algorithm.

Despite these limitations and cautions associated with various imputation methods, Little and Rubin, 1987, note that "It is important to emphasize that in many applications the issue of nonresponse bias is often more crucial than that of bias. In fact, it has been argued that providing a valid estimate of sampling variance is worse than providing no estimate if the estimator has a large bias…"

Comparisons of Alternative Imputation Methods

There are a number of extant studies comparing alternative imputation methods. Two of them were conducted using NCES data, and a third involving a set of simulations was supported by NCES.

IEA Reading Literacy Study

One example using NCES data from the U.S. component of the IEA Reading Literacy Study, compared complete case (CC) analysis, available case (AC) analysis, hot-deck (HD) imputation, and the EM algorithm (EM). The first three methods were described above. The EM algorithm uses an iterative maximum likelihood procedure to provide estimates of the mean and variance-covariance matrix based on all available data for each respondent. The algorithm assumes the data are from a multivariate normal distribution, and that, conditional on the reported data, the missing data are missing at random. To conduct this comparison, regression equations were estimated using the four methods of imputation.

A linear regression model was used to predict a student's performance on a reading literacy test. The three reading scores used as the dependent variables were the narrative, expository, and document performance scores. These scores were derived using Item Response Theory models scaled for international comparison (Elley, 1992). The predictor variables used in all models were gender, age, race, father's and mother's education, family structure, family composition, family wealth/possessions, and use of a language other than English at home. The amount of missing data ranged from 0 to 18 percent with 31 percent missing data for one or more variables.

Unweighted ordinary least squares regressions were run using each of the four imputation methods for the three independent variables. For each independent variable, the regression coefficients estimated using the HD, EM, and AC methods were very similar. The estimates using the CC analysis method were dissimilar. This analysis also used adjusted mean scores to examine the performance of subgroups of students after controlling for other characteristics. The adjusted scores for a number of subgroups (e.g. gender, minority status, and parent's education) showed mean

scores using CC that were approximately 10 points higher than the mean scores using HD, EM, and AC. These differences are presumably explained by the fact that the CC analysis excludes the 31 percent of the students who had missing data on one or more items.

This analysis was repeated for a comparison of CC, AC, and HD using weighted data. Although the use of the weights reduced the size of the gap somewhat, the differences persisted, with the CC analysis method yielding higher estimates than the AC and HD methods (which yielded similar results). The authors of this report (Winglee, et. al., 1994) concluded that the CC analysis method was clearly inefficient. Rather than the missing cases being randomly distributed, they found evidence that the students with missing data differed from those with complete data in reading performance, race/ethnicity, type of community, region of the country, and control of the school. They further concluded that given the similarity of results between the remaining three methods (AC, HD, and EM) since the HD method is the easiest to implement it is the best to use for the IEA study.

NELS:88

The second example from the analysis of NCES data uses data from the National Education Longitudinal Study of 1998 (NELS:88) to compare two imputations methods that were used for test scores—within-class random hot-deck imputation and model based random imputation. The goal of this study was to select an imputation method to use to impute missing reading and math scores in the base year to second follow-up cohort. Sixty-five percent of the cohort took all four cognitive assessments in the three waves of the survey. The nonresponse rates by key demographic subgroups ranged from 20.5 to 27.5 percent, with the highest rates among minority students and low SES students, causing some concern over potential bias in the NELS estimates of academic performance.

The authors of this analysis first identified a set of auxiliary variables, and then using the subset of cases with complete cases they simulated different levels and patterns of missingness assuming about 20 percent missing data. Following the simulation, the incomplete data were compared with the imputed data using the average imputing error, the bias of the variance, and the mean bias. The average imputation error was found to be consistently lower in the model-based approach compared to the hot-deck approach.

Looking first at math, although a comparison of the bias of the mean across the two imputation methods and the incomplete data showed no consistent pattern; the means computed with the incomplete data were outperformed by one or both of the other two imputation methods in all but one comparison (i.e., the bias was smaller for on of the other two methods). The relative bias of the variance was consistently smaller in the model-based approach than it was in the other two approaches. The same results were observed in reading.

The authors concluded that the model-based approach was the “preferred method” and proceeded to use PROC IMPUTE to implement the imputations for the NELS data set.

Simulation Study

In the NCES sponsored simulation study, Hu, Salvucci, and Cohen, 2000, used 6 evaluation criteria to compare 11 imputation methods for 4 types of distributions, 5 types of missing mechanisms, and 4 types of missing rates. The imputation methods evaluated include: mean imputation, ratio

imputation, sequential nearest neighbor hot deck imputation, overall random imputation, mean imputation with disturbance, ratio imputation with disturbance, approximate Bayesian bootstrap, Bayesian bootstrap, modeling non-ignorable missing mechanism (PROC IMPUTE), data augmentation (Schaefer's software), and adjusted data augmentation method.

The evaluation criteria used include: bias of parameter estimates, bias of variance estimates, coverage probability, confidence interval width, and average imputation error.
They found that the results varied across different types of missing data; the five types considered are: missing completely at random (MCAR), tails more likely missing, large values more likely missing, center values more likely missing, tail values more likely missing with confounded (missingness in y depends on y itself).

In the case where large values are missing ratio imputation (with or without disturbances), and data augmentation (Schafer) correct the bias in the mean; and within class random imputation and the sequential nearest neighbor hot-deck improved the biases substantially. However, the authors cautioned that the findings for ratio imputation may well be an artifact of their manipulation of the data. In summary, they note that although the improvement is much less when there is a right skewed distribution, in most cases these methods provide improvement when considerable biases exist in the means with the incomplete data.

In summarizing the results for variance estimation, the authors concluded that all imputation methods studies, except the mean imputation method, yield acceptable variance estimates when the data are missing completely at random. For the three unconfounded types of missing data—tails missing, large values missing, and center missing—data augmentation (Schafer) worked best, but ratio imputation, within class random imputation, and the sequential nearest neighbor hot-deck method all can improve the biases of variance estimates dramatically. (However, there is a caution that the ratio imputation method tends to overestimate the variance.) For the confounded missing data pattern, where the missingness is related to the variable itself, only the ratio imputation methods (with and without disturbances) results in a substantial improvement in the bias of the variance.

When coverage rates and confidence interval widths are considered together, data augmentation (Schafer) and adjusted data augmentation are the least likely to provide bad estimates. Finally, when average imputation error is considered, ratio imputation, data augmentation (Schafer), and within class random imputation perform best, followed by hot-deck, ratio with disturbance, and mean imputation methods.

Looking across the entire set of results, data augmentation (Schafer) is the one imputation method that scores high on all accounts. Two other methods that are more commonly used at NCES—within class random imputation (PROC IMPUTE) and the sequential nearest neighbor hot-deck method— also performed well in estimating means and variances and perform reasonably well on coverage rates and average imputation error (although within class random imputation (PROC IMPUTE) usually edges out the hot-deck method).

APPENDIX C

NCES Guidelines

For Tabular Presentations

(2002 Edition)

SAMPLE TABLE

Table 6.—Number of public high school completers, by state: School year 1999–2000 **TITLE**

BOXHEAD

State	Total	High School Completers		
		Diploma Recipients	Other High School Completers	High school equivalency recipients[1]
U.S.	—	**2,546,102**	[2]**41,638**	—
Alabama	43,459	37,819	2,535	3,105
Alaska	7,968	6,615	53	1,300
Arizona	—	38,304	375	—
.				
.				
.				
Wyoming	—	6,462	27	—

SPANNER HEAD

COLUMN HEADER

TABLE BODY

— *Data missing.*
† Not applicable.

SPECIAL NOTES

[1] Total other high school completers does not include New Hampshire, New Jersey, Washington, and Wisconsin.
[2] Includes recipients age 19 or younger, except in Minnesota, where they are age 20 or younger.

REFERENCE NOTES

NOTE: High school completer categories may include students not included in 12th-grade membership.

GENERAL NOTE

SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data, "State Nonfiscal Survey of Public Elementary/Secondary Education," 2000–01.

SOURCE NOTE

CONTENTS

	Page
INTRODUCTION	127
MAJOR TYPES OF TABLES	128
Summary/Text Tables	128
Reference Tables	128
Methodological Tables	128
TABULAR FORMAT	129
Printed Position	129
Single-Page Tables	129
Multi-Page Tables	129
Double-Page-Spread Tables	129
Hybrid Tables	129
TABLE TITLES	131
Table Identifiers	131
Wording of Table Titles	132
Placement of Titles	133
Titles for Multi-Page Tables	133
Headnotes	133
BOXHEADS	134
Parts of the Boxhead	134
Wording and Punctuation	134
Sequence of Columns	134
Spacing in the Column Head	134
Units of Measurement in the Column Head	134
Column Numbers or Letters	135
Breaking and Hyphenating Words	135
THE TABLE STUB	136
Organization of the Stub	136
Indention in the Stub	137
Vertical Spacing in the Stub	137
Boldface Type	138
Wording and Punctuation in the Stub	138
Leaders	138

Line Numbers 138
Continuations 139

THE BODY 142

Units of Measurement in the Body 142
Spanners 142
Decimals, Zeros, and Dollar and Percent Signs 143
Placing Figures in the Columns 143
Arranging Figures for Ease of Comparison 144

TABULAR NOTES 145

Headnotes 145
Special Notes 145
Footnotes 146
General Notes 146
Reference Notes 146
Source Notes 147

SIZING A TABLE 148

Pruning 148
Internal Revision 149
Spacing Reduction 149
Font Reduction 149

Appendix 150

INTRODUCTION

Tabular presentation is a way to bring together and present related material in columns or rows. The object is to show in a concise and orderly manner information that could not be shown so clearly in any other way. To many users and potential users, however, columns and rows of figures are not easy to understand. Important facts and figures may be buried in the masses of data shown. To enable the inexperienced user to accurately interpret the data, and the experienced statistician to do so more readily, table design should be kept as simple and direct as the subject matter and available space allow. In general, good design is as simple as possible, focuses attention on the data, and makes their meaning and significance clear. Poor design obscures the meaning and distracts attention.

A consistent "style" of presentation can help avoid distracting the user's attention. Subtle differences in terminology may cause the perceptive reader to ponder if a difference in meaning is involved. So, one of the general standards of good presentation is to use the same terminology in title, stub, headings, footnotes, etc.

To that end, these guidelines stress the importance of table design to satisfy the needs of the user, not of the producer. A consistent style builds a "normal expectation" through uniform treatment of many details. Unaccountable variation may distract the user and weaken the user's understanding of the content of the table. And by avoiding meaningless "differences," the table producer can capitalize on meaningful differences, and strengthen understanding, when deliberate small changes are made in words, phrases, or table structure.

The guidelines developed here attempt to adapt some widely accepted principles of tabular presentation to the subject matter, production methods, and operating procedures dealt with in NCES. Further, as with any set of guidelines, some arbitrary choice among acceptable alternatives is involved here. The guidelines are intended to help the development of clear and concise tabular presentations tailored to NCES needs.

Much of the material in the 1972 *NCES Guidelines for Tabular Presentation* was adapted from the *Census Manual*. The Government Printing Office *Style Manual* and the *Manual of Statistical Presentation* (January 1970), prepared by the Division of Research Grants, National Institutes of Health were also consulted for appropriate details. This 2001 edition draws heavily on the 1972 edition. Some of the revisions reflect technological changes. The *Publication Manual of the American Psychological Association* was also consulted for current practices. Beyond that, the modifications that have been made represent the experiences of a number of NCES analysts and NCES contractors.

MAJOR TYPES OF TABLES

There are three types of tables that are used in NCES publications. Taken in order of complexity, they are—summary tables, reference tables, and methodological tables.

Summary/Text Tables
Summary, or text, tables focus on selected data to show important comparisons and relationships. In reports containing analytical text, these tables are often placed at or near the first textual reference to them because they are closely related to the discussion. If numerous, they may be grouped at the ends of chapters or at the end of the report, preceding the reference tables, if there are any.

Reference Tables
Reference tables are the most detailed tables. Large quantities of information and comprehensive collections of data appear in reference tables. They normally form a separate section usually placed following the text at the end of the report. Sometimes, fairly short reference tables appear at the ends of chapters if summary tables are interspersed in the text.

Methodological Tables
Methodological tables contain standard errors or confidence intervals for data in a report. Place these tables in an appendix.

TABULAR FORMAT

Printed Position
Tables may be printed on the page in either portrait or landscape position in a variety of structural forms. In portrait tables, the words and data extend across the printed page (normal "width"), as these sentences do. Most tables present statistical data in this format. Landscape tables are rotated a quarter turn to the left, with the words and data extending up the page -- the top of the table at the left, the bottom at the right. Landscape tables should be avoided if possible (particularly when interspersed in a report with the text and other tables in an upright position) because smooth transition is interrupted from text to table and from table to table.

Single-page Tables
Occupying one page or less, these tables are easy to examine and highly desirable, especially as summary tables. If well designed, they convey easily grasped amounts of information as complete units. Frequently, careful pruning will allow a table that is either a little too long, a little too wide, or both to fit on a single page.

Multi-page Tables
Although single page tables are preferred, there are times when a table is too long to fit on one page; if these tables cannot logically be split into smaller tables, they must be continued on one or more additional pages. The title (with "—Continued") and the boxhead are repeated on successive pages of multi-page tables. The end of each page preceding the last page of a multi-page table should carry a note advising the reader to "See footnotes at end of table." The notes for a multi-page table appear on the last page of a multi-page table.

Double-page-spread Tables
The double page spread is a special kind of portrait multi-page table that extends across facing pages, instead of one page, with about half of the column headings on each page. It may continue on successive facing pages. The entire stub should be repeated at the right side of the right-hand page; but if there is not enough room, line numbers may be used instead. (See Line Numbers, page 13.) The title is repeated on the second and subsequent pairs of pages (with "—Continued"). Otherwise, the double page spread is treated much like a one-page-size portrait table with the advantage of accommodating about twice as many columns.

Hybrid Tables
Two types of portrait tables that combine some of the aspects of both page-wide and double-page spread tables are the "divide" and the "double-up" tables.

Divide tables are portrait multi-page tables in which the title is repeated (with "—Continued"), the stub is repeated on the left of each page, and the column heads continue across a second page or more. If only two pages wide, it may be set up on facing pages, like a double page spread, and the stub may continue for any number of pages. The

divide table is useful if the stub is only one page long but the table must be three or more pages wide. It obviously cannot be both too long for one page and too wide for two.

First page:

Table 1.—Title * * * * *

Stub-head	Column heads				
	A	B	C	D	E

Second page:

Table 1.—Title * * Continued

Stub-head	Column heads				
	F	G	H	I	J

Third page:

Table 1.—Title * * Continued

Stub-head	Column heads				
	K	L	M	N	O

Double-up tables are set up somewhat like a double-page-spread table confined to one-page width. It is especially useful for a long table with few columns. It may continue as a multi-page table. The title occupies the width of the page, but the stub-head and column heads are repeated under it in the two halves, as shown.

Table 1.—Title *

Name of Institution	Enrollment		Name of Institution	Enrollment	
	Men	Women		Men	Women

Or alternatively,

Table 1.—Title *

Name of Institution	Enrollment		Name of Institution	Enrollment	
	Men	Women		Men	Women

TABLE TITLES

The formal tables (summary, reference, and methodological) have headings consisting of identification symbols (numerical or alphabetical); descriptive titles; and, sometimes, headnotes.

Table Identifiers

Tables in Executive Summaries should be lettered alphabetically, and tables in the body of the report should be numbered consecutively. For many reports simple identifiers such as Arabic numerals in sequence–1, 2, and so on–are the best solution. For example, most NCES reports have a short introductory text, with no chapter numbers needed; one series of tables, requiring identifiers; and one appendix, requiring none. However, distinguishing identifiers are needed for more than one series of tables (such as a few summary tables and reference tables) or more than one appendix. An orderly system that takes account of the table identifiers in relation to the other parts of a report is needed. Without this, much confusion would result in a publication with, for example, as many as three or more separate series of tables (summary, reference, and those in one or more appendices) to distinguish them from or relate them with a series of charts, the appendices themselves, and several chapters.

Readily available for identifiers are Arabic numerals and the English alphabet in uppercase and lowercase. Arabic numerals are easiest to comprehend and can extend easily through any number of table titles. In addition, the tables within a particular series may have sub-series that need to be related. For example, a main or "master" table may show particular data for all postsecondary institutions in the United States, followed by a subseries of tables showing identical kinds of data separately for universities, other 4-year institutions, and 2-year institutions. They should be numbered with a basic identifier and an appropriate suffix that is selected to avoid disrupting the standard numbering system and to bring out the table relationships, as shown in the following example:

Table 5.—All institutions
Table 5-A.—Universities
Table 5-B.—4-year institutions
Table 5-C.—2-year institutions

The same scheme might be used for a frequency table showing basic figures followed by a table of percents or medians derived from the basic figures (tables 5 and 5-A).

A slight variation may be used when component parts are shown separately in a series of tables *without* a master table. These tables are basically a single whole table that is split apart into a series of consecutive tables for convenience. They might be numbered 5-A, 5-B, etc.

Table 5-A.—Publicly controlled institutions
Table 5-B.—Privately controlled institutions

Appendices should be lettered; and tables in appendices should be assigned the letter of the appendix and a number suffix. For example, tables in appendix A should be labeled A-1, A-2, etc.; in appendix B, B-1, B-2, etc. If there is a methodology table for each summary/text table, it is helpful to use the appendix letter followed by a number suffix, where the number corresponds to the text table number.

Wording of Table Titles

Titles are catalogs of content and guides for ready reference. They should tell **what, how classified, where,** and **when**. For example:

What:	Basic content and general limits of the group or subgroup that are shown in the table (e.g., enrollments in postsecondary institutions).
How classified:	How the universe data are classified and cross-classified (e.g., by control of institution, age and sex of student, geographic region, and state).
Where:	Area or space segment, such as political division, geographic area, or other coverage designation if necessary for clarity (e.g., by country, by states, or, perhaps, geographic regions)
When:	Time reference (e.g., 2000; September 1999; academic year 1998-99; various years, 1950-90, etc.)

Thus, we might have:

Table 1.—Full-time equivalent fall enrollment in postsecondary institutions, by control, and age: By state, 1998

Note the punctuation, a period and an em-dash, with no extra space between number and first word are used to separate the title from the table identifier. A comma is used before the "by" classification, with commas separating series of three or more components, including a comma before "and." Finally, a colon is used before "where" or "when" reference (use a comma between where and when if both are present). Note also that, besides proper nouns, the first word of the title and the first word after the colon begin with capital letters.

For the "how classified" segment, a definite order should be used. Start with the data-column heads crossing left to right and top to bottom, then the stub. For example, the title above would fit a table set up like this:

	Total, in all institutions		In publicly controlled		In privately controlled	
State	Under 30	Over 30	Under 30	Over 30	Under 30	Over 30
U.S.						
Alabama						
Alaska						
Arizona						

If the purpose is to emphasize one of these elements, rewording of the title might better reflect the content. For example, the element that sets this table apart from others in a series might be the control classification. Then the title could read: "Enrollments in publicly and privately controlled institutions of higher education, by age, sex, and state: 1998" leaving out "control" in the classification segment.

The title must never promise more than the table contains, but the table may contain more. To avoid excessive wordiness, generalizations may be used, but table titles should be detailed and explicit enough to differentiate any one table from all others in a report. For example, if the number of items in the classification segment is lengthy and a subset of items are repeated across a series of tables, the table titles might read:

Table 1.—Fall enrollment in elementary and secondary schools, by free lunch eligibility and selected characteristics: 1999
Table 2.—Fall enrollment in elementary and secondary schools, by minority enrollment and selected characteristics: 1999

The wording should be in topical form, not in sentence form. This means that verbs are omitted from titles, as are articles and other parts of speech that do not convey the basic "numbers of," "percent of," and "distributions of" if the meaning and differentiation from other tables are clear without them. Carefully chosen headnotes and footnotes also may help shorten titles. Abbreviations are used sparingly, and then only those that are commonly accepted or otherwise identified, as in footnotes or text.

Placement of Titles
Start the first line of the title at the left margin and begin each subsequent line under the first word of the title.

Table 1.—The first line of the title extends the first line the width of the table; the second and subsequent lines begin under the first word of the title;

Titles for Multi-page Tables
For each page after the first page of a multi-page table, repeat the table number and the full table title, with the word "—Continued" added, as follows:

Table 1.—Total expenditures for public elementary and secondary education, by function and state: 1995-96—Continued

In the case of a double-page-spread table the word "—Continued" is added after the first pair of facing pages.

Headnotes
The headnote—a general qualifying statement in brackets, centered under the title—should be used only when it applies to all or almost all of the table or clarifies the contents of the table by expanding or qualifying the title. The headnote ends without a period, even if the last statement is a complete sentence; but internal periods are used if required by sentence structure. (See section on Tabular Notes, page 18.)

BOXHEADS

The boxhead consists of the stubhead, column heads, and spanner heads that classify, describe, or qualify the column or columns to which they refer. The heads are placed approximately in the center of areas defined by real or imaginary lines (boxes) directly above the vertical columns of information to which they apply.

Parts of the Boxhead

The column head is the basic unit of the boxhead, and each column should have one. It may or may not be qualified, supplemented, or described by one or more spanner heads above it.

Spanner heads, or multicolumn heads, are placed above two or more subordinate column heads to clarify, describe, or shorten the subordinate heads (See also Spanners, page 15). A single spanner head may also span two or more subordinate spanner heads, as in this example:

State or other area	All students	First-time students only				
		Number			Percent of total number	
		Total	Men	Women	Men	Women

In double-page-spread tables, spanners continue from the left-hand to the right-hand page of the pair, with "—Continued" added following the repeated spanners on the right-hand page.

A banner head, which is a special type of spanner head that is rarely needed, extends over all columns except the stub. The best use of a banner head is as a "read-in" line that clarifies data in the columns in relation to the column heads. In the following example, the banner is appropriate to all data columns and identifies the data shown as different from what the single column heads indicate.

Type of licenses	Licensees and stations in:					
	Aggregate United States	North Atlantic	Great Lakes and Plains	South -east	West & South- west	Out- lying areas

Wording and Punctuation

Column heads should read *horizontally*—almost never vertically. Wording is brief, as in other parts of the table, and requires careful phrasing. Horizontal space almost always can be saved by using multicolumn heads, by putting wide heads on more lines, by hyphenating words at the ends of lines, and by using standard, easily understood abbreviations where necessary. (See section on Breaking and Hyphenating Words, page 10.) To avoid an overly formal appearance, capitalize only the first letter of the first word in each head and the first letters of any proper nouns.

Sequence of Columns
Total and subtotal columns are placed at the left of the columns that they aggregate, except in financial tables prepared specifically for accounting purposes, which require totals at the right. In NCES publications intended for broad readership, tables showing dollar amounts have totals on the left. Derived figures—such as averages, ratios, and percentages—usually are placed in columns to the *right* of the base figures.

Spacing in the Column Head
The illustrations following show minimum, normal, and maximum recommended spacing in the boxhead.

Minimum vertical spacing in the boxhead

Normal vertical spacing in the boxhead

Maximum vertical spacing in the boxhead

In these three examples, the column is approximately centered vertically in the area assigned. For minimum spacing, no blank space is left above or below this head; this spacing should only be used in cases where space is at a premium.

Each column heading in the body of the table should be placed flush right over the column. Within each set of column headings, each column heading should end on the same line. (See also section on Placing Figures in the Column, page 16.)

In a ruled table, all of the column-heading boxes on the same level should be the same height, as determined by the column heading with the most typed lines.

Wrong:

	Research grants	Formula and project grants		Training grants	
1968	1969	1968	1969	1968	1969

Right:

Research grants		Formula and project grants		Training grants	
1968	1969	1968	1969	1968	1969

Units of Measurement in the Column Head
Units of measurement (e.g., pounds, percent, dollars) often appear in the column head. When they do, they should be placed after or below the column-head captions that they modify. Sometimes a unit of measurement comprises an entire column head.

1. If it modifies a caption, enclose it in parentheses and use all lowercase letters; for example: “Expenditures (millions of dollars)” and “Dollars awarded (in

thousands)." Abbreviations, if used, should be clear; for example: "Floor area (1,000 sq. ft.)" or "Floor area (thous. sq. ft.)."

2. If it comprises an entire column head, omit the parentheses and treat it like any other column heading. Capitalize, for example, "Billions of dollars" or "Percent of total" if it is the entire head.

Column Numbers or Letters

Occasionally, tables with many column headings need numbered or lettered columns for ease of reference. The numbers or letters appear just below the boxhead and run in sequence from left to right beginning with the stub. Column numbers or letters may be enclosed in parentheses or separated from the rest of the table by a horizontal ruling.

Stub Caption	Column head	Column head
(1)	(2)	(3)
Total	986	461
Item	0	73
Item	986	388

OR

Stub caption	Column head	Column head
1	2	3
Total	986	461
Item	0	76
Item	986	388

Breaking and Hyphenating Words

Most often in headings (but also in stubs), breaking and hyphenating words is necessary. The guide for breaking words and use of hyphens is the GPO Style Manual and its Word Division Supplement. Comments on some common pitfalls follow:

Break words only between syllables: usually divide doubled consonants (e.g., syl-la-bles
but en-roll-ments).
Never break one-syllable words.
Avoid breaking:
Words that would leave a one-letter syllable on a line (**not** a-mendment).
Words of four or five letters.
Always hyphenate as follows:
Full-time equivalents and full-time-equivalent number.
Nondegree-credit students (<u>but</u> noncredit courses and activities).
Nonscience-related curricula.
Nonengineering-related technologies.
First-professional degrees.

THE TABLE STUB

The stub consists of a heading and the line captions that are listed at the left side of a table and describe each row of figures in the field. Capitalize only the first letter of the first word and the first letters of any proper nouns in both the stub heading and the line captions. Always provide a stub heading that describes, defines, or amplifies the stub captions. Use a word such as "Item" or "Characteristic" for a collection of stub entries that defy brief classification:

Stub caption	First column heading
Total	$979,630
Subtotal	425,111

When the stub is too long for one page and must be continued on another page, the continuation should also be placed at the left side of the second page.

For a double-page spread, the stub should be repeated on the right side of the right-hand page. Line numbers may be substituted for the right-hand stub if space is tight. (See Double-Page-Spread Tables, page 3.)

Left-hand page		**Right-hand page with stub**	
Stub caption	First column heading	Last column heading	Stub caption
Total	$979,630	$1,460,325	Total
Subtotal	425,111	678,013	Subtotal

Organization of the Stub

Place grand totals at the top of the column stub. Then the items in a stub should be displayed in a logical sequence. Some typical categories are alphabetical, geographical, chronological, numerical, quantitative (by size), customary (commonly accepted order), progressive (order of growth or development), and importance. Sometimes the arrangement of items in a stub of a single table may fall into two or more categories. For example, the main order might be geographic (which could be customary also) with the states listed alphabetically, sometimes listed under each geographic region.

Convention requires year entries showing trends to run sequentially from earliest to latest. Stub entries consisting entirely of years are centered in the area allotted to the stub.

Indention in the Stub

When there are multiple levels of subordination to be displayed within a table, indention of the stub can provide a road map to help readers follow the flow of a table. Indention can best be accomplished by setting tabs at a space equivalent to a specified number of the letter "n."

- **Grand totals**—If there is only one other level, indent three spaces (i.e., start in the fourth space). Indent five spaces if there are two or more levels of subordination.
- **Major group or subtotal captions**—Start the caption line at the left edge of the table. Indent any continuation lines three spaces.
- **Subordinate captions**—Tab two additional spaces for each subsequent level of subordination (e.g. two spaces for the third level group and four spaces for the fourth level). Indent any continuation lines three spaces.

For example:

Stub head (centered vertically and flush left)	Spanner head
	Col. head.
Total	1,625
Major group	860
Minor group	514
Item	101
Item	98
Item	193
Item	32
Item	47
Item	43

Vertical Spacing in the Stub

Normal vertical spacing in the stub leaves a blank line between the total and the first group caption, between group captions, and between a subordinate series and following superior group caption. (See also Spacing in the Column Head, page 9 and Sizing a Table, page 21.)

When available vertical space is tight, reducing the height of blank lines can minimize normal spacing. The absolute in minimum spacing allows removal of all blank lines between stub captions, and then bolding total and all major group (subtotal) captions.

Subordinate items under a group caption are usually single-spaced except when there is a long list of such items. Then it is best, if space permits, to group them by three, four, five, or more items with a blank "reader" line between groups.

Boldface type

When vertical spacing is tight, boldface type, instead of line spacing, may be used to set off group captions in the stub. The separation is indicated by bolding the group captions. Boldface type may also be used to make totals and subtotals stand out. But first, the table should be examined carefully to determine whether appropriate spacing and indention of the stub captions without using boldface type could achieve the same result.

Wording and Punctuation in the Stub

Stub captions should be as brief as possible without losing precision and clarity. If space is limited, abbreviations are used only when they can be understood instantly. Minimum punctuation is used to make the meaning clear. Periods are omitted at the ends of stub captions and may also be omitted after abbreviations to save an additional space in very tight stubs, if the meaning is clear.

Leaders

Leaders are rows of periods connecting the last word of a stub caption (last line of an overrun) with the first data column. If used in tables with no vertical rulings two or three spaces should separate the leaders from the longest number in the first data column. Use leaders only when a wide space divides the stub caption and the first column of data in the body of the table.

Leaders are **always** omitted after stub captions without entries opposite them in the field. They are almost always omitted in the duplicate stub at the far side of the right-hand page of a double-page spread.

Line Numbers

The main use for line numbers is for convenience of reference or to alleviate a tight stub situation in the right-hand side of a double-page-spread table. When they are used, all stub captions that identify entries in the field should be numbered consecutively. The line numbers are lined up as a column, two spaces to the left of the stub entry positioned farthest left and are placed opposite the last line of an overrun caption.

In a double-page-spread table the line numbers should be repeated on the right-hand page, as the last column two spaces to the right of the longest line of the duplicate stub. If space is very limited, use line numbers only (omitting the duplicate stub), matching them with the line numbers of the stub on the left-hand page. The illustrations below show how to use line numbers in a normal double-page-spread table and in one in which space is limited.

Left-hand page			Right-hand page with stub			Right-hand page without stub	
Stub Caption		First column heading	Last column heading	Stub caption		Last column heading	
1	Total	$979,630	$1,460,325	Total	1	$1,460,325	1
2	Subtotal	425,111	678,013	Subtotal	2	678,013	2
3	Item	98,409	229,481	Item	3	229,481	3
4	Item	134,862	65,319	Item	4	65,319	4
5	Item with overrun	7,516	187,005	Item with overrun	5	187,005	5

Continuations

When a category with subcategory listings breaks over to another page all superior categories should be repeated, with the word "—Continued."

For example:

Foreign languages . . . . .
- French . . . . . .
- Spanish. . . . . .
- Other foreign languages .

Social sciences . . . . . .
- U.S. history. . . . . .
- World history. . . . .
- Other history:
  - Ancient history. . .
  - Oriental history. . .

Social sciences—Continued
- Other history—Continued
  - Bible history . . . . . .
  - Local history . . . . . .
- Geography . . . . . .
- Government studies . . . . . .
- Current events . . . . . .

Other . . . . . .

THE BODY

Body, or field, is the part of a table that contains the numerical data—below the column heads and to the right of the stub. It consists of cells, rows, and columns. A cell is the space occupied by one entry in the field. A row is a horizontal array of cells opposite a stub caption. A column is a vertical array of cells under a column heading.

Units of Measurement in the Body

Units of measurement usually do not appear in the body. The preferred places for units of measurement are in a headnote, if they apply to all or nearly all of the table (see Headnotes, page 18) or in the boxhead, if they vary by column (see Units of Measurement in the Column Head, page 9).

Spanners

Spanners are multicolumn headings that cross the table within the field instead of in the boxhead. The column heads at the top of the table apply to all levels in the field. The field spanner is most useful when emphasis on a change of category is needed and the label applies directly to the data in the field. They should not be used when they apply to the stub-entry classification.

Table 1.—Number and percentage distribution of families, by family status and presence of own children under 18: 1980 to 1998

Family status	1970	1980	1998	Change, 1970 to 1980	Change, 1980 to 1998
	In thousands			Percent change	
All families	51,456	59,550	70,880	15.7	19.0
Married-couple family	44,728	49,112	54,317	9.8	10.6
No own children under 18	19,196	24,151	29,048	25.8	20.3
With own children under 18	25,532	24,961	25,269	-2.2	1.2
Other family, male householder	1,228	1,733	3,911	41.1	125.7
No own children under 18					
With own children under 18					
Other family, female householder					
.					
	Percent of all families			Change in percentage points	
All families	100.0	100.0	100.0	100.0	100.0
Married-couple family	86.9	82.5	76.6	-4.5	-5.8
No own children under 18	37.3	40.6	41.0	3.3	0.4
With own children under 18	49.6	41.9	35.7	-7.7	-6.3
.					

Field spanners sometimes are used to reduce the length and increase the width of very narrow and long tables. They also may be used for placing long major group captions in the field when there
is not enough room for them in the stub. These advantages are offset to some extent by their unfavorable location in the field where they break across the columns and separate the figures from the descriptive column headings.

Decimals, Zeros, and Dollar and Percent Signs

In a column of figures containing decimal fractions, figures of less than 1 have a zero (0) to the left of the decimal point. However, do not use a zero before a decimal fraction when the number cannot be greater than 1 (e.g. levels of statistical significance, proportions, or correlations). If there are whole numbers (numbers without decimal fractions) in the column, they are recorded with a decimal and zero to the right of the decimal point. All figures in a table that are reported in the same unit of measurement should report data to the *same* decimal value. If the column consists entirely of whole numbers, do not use decimal points and zeroes. The recorded number of decimal places should offer no greater degree of precision than is warranted by the data (see standard 5-3-02, NCES Statistical Standards, 2002).

As shown below, the only exception to these rules is that absolute zero (0) is always expressed as a single zero without a decimal point; in a column of decimal fractions, it is positioned as shown.

TABLE A

Item A	0
Item B	0.7
Item C	4.0
Item D	18.6
Item E..	#

Rounds to zero.

TABLE B

Item A	0
Item B	0.72
Item C	4.00
Item D	18.64
Item E	#

Rounds to zero.

TABLE C

Item A	0
Item B	1
Item C	4
Item D	19
Item E	#

Rounds to zero.

When all of the figures in a column pertain to money, the first figure in the column should be preceded by a dollar sign ($), even though the column heading or a headnote indicates the unit of measurement (e.g., millions of dollars).

A percent sign (%) should not follow figures in the field. If all are percentages, the fact may be indicated in a headnote: if some columns or lines are percents, indicate in a spanner, individual column heads, stub entry, or title, as appropriate (e.g., “in percent”). The word “percent” instead of “percentage” is preferred in this context; the symbol (%) should be used only if there is no room to spell it out.

Placing Figures in the Columns

Allow a minimum of one space on each side of an entry. Entries should be aligned at the right-hand side—including absolute zero in number columns. For two-line stub captions,

entries are placed opposite the second line. Leave no cell empty; if a number is not available, insert the appropriate explanatory special symbol in the cell. (See list in Special notes, page19).

Arranging Figures for Ease of Comparison

The closer numbers are to each other, the easier it is to compare them. Vertical comparisons usually can be made more rapidly than horizontal comparisons. In the following example, arrangements A and B both are satisfactory, but the vertical listing in B is more effective because it is much easier to locate the largest and smallest numbers and to determine differences in the general sizes of the numbers.

A	B				
102,007,666	102,007,666	1,998,464,732	99,428,531	941,325	23,918
1,998,464,732					
99,428,531	(NOTE: The vertical arrangement brings the figures closer				
941,325	together and requires less movement of the eyes.)				
23,918					

The following tabulations show identical data, but the vertical comparison in B is more emphatic than the horizontal comparison.

	A				B				
Item	FY 1964	FY 1965	FY 1966	FY 1867	Fiscal Year	Item A	Item B	Item C	Item D
A	1,192	6,195	8,628	7,107	1964	1,192	647	92	5,430
B	647	502	111	835	1965	6,195	502	86	1,999
C	92	86	75	42	1966	8,629	111	75	3,671
D	5,430	1,999	3,671	4,442	1967	7,107	835	42	4,442
E	775	215	303	629	1968	2,888	229	34	1,041

In any table, the comparisons that are the most important should be placed as close together as possible for maximum emphasis.

TABULAR NOTES

Tabular notes contain supplementary information necessary for a correct understanding of the table or a part of it. They fit into two categories: (1) headnotes at the top of the table are used only occasionally, and (2) footnotes at the bottom of the table are used often. Footnotes include general notes, reference notes, and source notes.

Tabular notes should be kept as brief as possible without sacrificing clarity. Topical style is used, with subject-noun, verb, articles, and other parts of speech omitted if not essential to understanding.

Headnotes

A headnote is a special explanation that should be seen before the rest of the table is read. The headnote should be used only when it applies to all or almost all of the cells in the body of the table or if it clarifies the contents of the table by expanding or qualifying the title. Sometimes, careful wording of title and column heads can eliminate the need for headnotes. Consider, instead of the headnote, a general note (NOTE: Data are . . . . .), or a reference footnote with the symbol attached to column heads or stub. Reference notes attached to the title should be avoided, if possible.

A headnote should be centered above the boxhead; if two lines are needed, the second should be centered under the first. It should be enclosed in brackets and typed in lowercase letters, except for the first letter of the first word and the first letters of proper nouns and adjectives. No period is placed after the last word; if more than one sentence, a period ends all but the last sentence. The following are typical examples of headnotes.

[Based on a 10-percent sample of applications]

[Includes both public and private]

[Millions of dollars]

Sometimes a headnote may indicate a unit of measurement that applies to some, but not all, of the columns of figures:

[Dollar amounts in thousands]

Normally, one blank line separates the headnotes from the table title; but more room may be left, if necessary, to make the table fit the available space. Two blank lines usually separate the headnote from the top line of the boxhead.

Special notes

Special notes are notes that are standard for cells in the body of tables and usually refer to a statistical property of the specific cell (e.g. not applicable, missing, an unstable estimate, statistically significant). Special notes fill cells in the body of tables, and do not require parentheses. When special notes are used, they should always be listed in the

following order. The following list summarizes a number of statistical special notes and related set of symbols that should be used consistently across all NCES reports.

Symbol	Label	Meaning
—	Not available	Data were not collected or not reported
†	Not applicable	Category does not exist
#	Rounds to zero	The estimate rounds to zero
!	Interpret data with caution	Estimates are unstable
‡	Reporting standards not met	State/Country did not meet reporting standards
*	p<0.05	Significance level

Footnotes

General, reference, and source notes fall at the bottom or "foot" of the table. General notes refer to all or much of the table; reference notes, to specifically designated portions; and source notes identify sources of the data. All end with a period.

General notes

General notes, like the headnotes, qualify, describe, or explain whole tables or easily identifiable parts of them. The choice between a general note and a headnote is guided by the degree of emphasis required, and the length and detail included in the note.

The general note is introduced with the word "NOTE" followed by a colon. For example:

NOTE: Detail may not add to totals because of rounding.

Reference notes

Reference notes refer to specifically designated portions of the table. By "keying" the note to the material to be qualified reference notes can be kept brief.

States	Classroom teachers		
	Full-time	Part-time	FTE[1]

[1]Full-time equivalent of full-time and part-time.

The positioning of symbols for reference notes in tables follows definite principles. The symbols are placed at the right of the word the note applies to, in both headings and stubs. They are placed at the left of data in the field of a table (if software permits); and if a footnote stands alone in a cell, it is enclosed in parentheses: (1). Footnotes are numbered sequentially throughout a single table, but a recurrent reference repeats the symbol. Footnotes follow a logical order, generally line for line from left to right and down.

The placement of footnote symbols within a table and the arrangement of notes at the end of the table are illustrated in the table below. Footnotes are placed at the end of the table. Special notes are listed first, followed by reference footnotes, general notes, and then the source.

Table 1.— Families, by family status and presence of own children under 18: 1980 to 1998

Family status	1970	1980	1998
All families	51,456	59,550	70,880

#Special symbols are listed first.
[1]Numerical footnotes follow.
NOTE: The general note comes next.
SOURCE:

Source notes
The *source* note indicates the specific source of the statistic. In general, the source note refers the user to the original (or primary) source and gives credit to the originating report, or in the case of new tabulations, the data file.

The source note should cite the report, relevant survey(s) or sub-survey(s), data reference year, file version number, department name, and agency name. In the case of unpublished data, use the month and year of the tabulation or data file. If the data are drawn from multiple years: for one to three years, report each year; for more than three continuous years, use the year span; and for more than three noncontinuous years use "selected years" and the year span.

Following are some typical examples:

Data from one or more reports:

Revenues and Expenditures for National Public Elementary and Secondary Education: School Year 1997-98, Common Core of Data (CCD), "National Public Education Financial Survey" (NPEFS), 1997-98, Version 1, U.S. Department of Education, National Center for Education Statistics.

Data from unpublished tabulations and a published NCES report:

SOURCE: U.S. Department of Commerce, Bureau of the Census, Current Population Survey, Previously unpublished tabulation (April 1998); and U.S. Department of Education, National Center for Education Statistics, *Dropout Rates in the United States*. Selected years 1972–97.

SIZING A TABLE

Most NCES publications are printed on paper that is 8 1/2 x 11 3/8 inches. The "image" size (area occupied by printed matter) is expected to be about 6 1/2 x 9 1/2 inches, *including space for the page number.*

It is well to note that, although this section focuses on ways to reduce dimensions, do so within reason. The problems of table layout usually are those of too much rather than too little, and too much vacant space within a table is no less a fault than others.

Some ways to improve the appearance and reduce one or both dimensions include pruning, internal revision, and font reduction, now discussed in that order. (See also Spacing in the Column Head, page 9 and Vertical Spacing in the Stub, page 12.)

Pruning

Trimming a table to alter its shape aims to prune its outline to the desired proportion. Of course, internal symmetry also is desirable within reason—such as relatively even spacing among the structural elements of the column heads, data columns, and stub captions. Here are some suggestions.

To reduce the width of a table, try—

1. Typing wide column headings or stub captions on several lines, dividing words if necessary.

2. Using spanner (multicolumn) headings over related column headings to avoid the repetition of duplicating words.

3. Paring unnecessary words in or abbreviating the stub captions and column headings. (See Wording and Punctuation in the Column Head, page 8, and Wording and Punctuation in the Stub, page 13.)

4. Rounding columns of figures.

To reduce the length of a table, try—

1. Typing column headings or stub captions on fewer lines by abbreviating or by placing more words on each line.

2. Removing the blank lines in the column headings or stub captions. (See Spacing in the Column Head, page 8, and Vertical Spacing in the Stub, page 12.)

4. Omitting a blank line above or below headnotes (See Headnotes, page 7.)

5. Omitting the blank line below the first footnote or placing two or more footnotes on one line.

6. Examining the stub to eliminate unnecessary nondata captions.

7. Paring unnecessary words in column heads and stub captions by using spanner headings.

Internal Revision

Sometimes an odd-shaped table can be tailored to fit a single page by revising its internal structure. For example, if the table is very wide and short, the table may be "turned," by reversing the functions and positions of the stub and the boxhead. (See also Arranging Figures for Emphasis, page 17.) Or, the boxhead may be divided into two levels, repeating the stub as below:

Fiscal Year	Number	Dollars (millions)	Number	Dollar (millions)
	Research grants		Training grants	
1970				
1980				
2000				
	Formula grants		Project grants	
1970				
1980				
2000				

Conversely, if the table is narrow and much too long for the page, using a double-up table format may shorten it. (Continuing the stub in two parallel columns on the page and repeating the boxhead across the top see page 4.) In less drastic situations, some data columns or data lines may be eliminated by incorporating low-yield categories, or more of them, in an "other" (residual) category or by eliminating categories entirely if they yield no data.

Spacing Reduction

Blank lines can be variably sized. By reducing the vertical spacing from a full line to three-quarters or one-half a line, the size of a table may be reduced and the number of printed pages may be reduced.

Font Reduction

Smaller fonts can be used to reduce tables that are too long and/or too wide. It is often desirable to reduce statistical tables for other reasons also. With exceptionally long tables, the number of printed pages may be long. By using a smaller font, the number of printed pages may be substantially cut, thus making the publication easier to use as well as lowering the printing, storage, and mailing costs. And many tables are easier to read a slightly smaller size. Note, however, that a minimum practical font size is 9.

APPENDIX
HOW TO PRODUCE TABLES AT NCES

(This is a quick guide; for more details, see the 2000 edition of NCES Guidelines for Tabular Presentations)

GENERAL

1) An NCES report may contain as many as three different types of tables.

   - Summary/text tables—range in size from a few lines to one or two pages and require titles and table numbers. You can place each summary table at or near the first text reference to the table or group them at the ends of chapters or at the end of the report in the order mentioned in the text. Summary tables precede reference tables if there are any.

   - Reference tables—detailed tables containing large quantities of data. They usually form a separate section at the end of the text or in an appendix (in the order mentioned in the text). When these tables include standard errors along with the data they should be placed in an appendix.

   - Methodological tables—contain relevant statistics for the data in a report; for example, sample sizes, coefficients of variation, or standard errors. Place these tables in an appendix and follow the order the tables are presented in the report.

2) If you disperse tables throughout the text, refer to each of them in the narrative, and refer to them sequentially (i.e., the tables should appear in the order mentioned in the text).

TABLE TITLES

1) Start out with the topic of the table, followed by a comma and then the "by" list.

2) In the "by" list, items in the columns are listed first, followed by the items in the rows.

3) End the title with a colon followed by the data year(s).

4) Capitalize only the first word, proper nouns, and the word following the colon.

5) Avoid footnoting a title, use a general note (i.e. NOTE: instead).

6) Year spans—use 1988–97 or 1988 through 1997 for a span of calendar years; 1988 and 1987 for two distinct years. Use Fiscal years 1989–98 or Fiscal years 1989 through 1998 for a span of fiscal years. And, use Academic year 1988–89 for one school year or Academic years 1988–89 through 1991–92 for a span of school years. Use en dashes instead of hyphens between years.

ORGANIZATION OF SIDE STUBS

1) Place the stub header flush left.

2) Report grand totals in the first row of the table.

   - If the table has only two levels, that is, the grand total and one disaggregation, tab over three spaces the size of the letter "n" to start the table grand total.

   - If the table has three or more levels, that is the grand total and at least two disaggregations, tab over 5 "n" spaces to start the grand total.

3) Start the label for the first level of disaggregation (that is, the major group or subtotal) at the left margin of the table.

4) Tab over two "n" spaces to start the second level of disaggregation.

5) Tab over four "n" spaces for the label for a third level of disaggregation (continue this pattern for additional levels of disaggregation).

6) If a row label needs a footnote place it to the right of the label.

7) If the rows are school years, use "School year ending" as the stub and then use the single years across from such a stub.

8) Use full state names in table stubs.

9) Use an en dash to designate "through" when referring to age.

HEADERS

1) Place the side stub head flush left.

2) Column spanners should be centered over the set of columns they describe.

3) Place each column head flush right.

4) If a column header needs a footnote place it to the right of the header.

5) If the columns are school years, use "School year ending" as a column head and then use the single years under such a head.

6) Use an en-dash to designate "through" when referring to age span.

BODY OF THE TABLE

1) Do not mix different measurements of data in the same column (e.g. percents and counts).

2) In tables displaying dollar amounts over time, indicate whether the amounts are current or constant dollars and include the base year (e.g. in constant/current 1997 dollars). Place dollar signs only in the first row.

3) If the rows and columns in a table do not add to the totals presented, add a general note, "Detail may not add to totals because of rounding."

4) LEAVE NO TABLE CELLS BLANK

- Use a † if data for the cell are not applicable. (Do not use NA). Use — for not available (i.e. not reported) and # for estimate too small to report. These symbols fill the cell and do not require parentheses. See Attachment for a list of symbols to use with blank cells.
- If a cell is blank for a reason not covered by a special symbol, footnote the cell with the footnote number or symbol in parentheses flush right in the column.

5) If a number in a cell needs a footnote, place the footnote to the left of the number (if software permits); and if a footnote stands alone in a cell, it is enclosed in parentheses: (1).

6) In order to place a zero in a cell, the measure must actually be zero. (It is preferable to report it 0, not 0.0.)

7) Use a line of periods (leaders) only when a wide space divides the stub and the first column.

8) In text and summary tables, round percentages to no more than one decimal place, round four and five digit numbers to hundreds, and round six digit numbers and over to thousands.

9) In reference and methodology tables, round percentages to no more than two decimal places, except in certain methodological tables, where a finer breakdown may be necessary. Standard errors should be reported to one decimal place more than the related estimate.

BOTTOM OF TABLE

1) Footnotes—bring all lines of a footnote flush left.

2) Symbol footnotes precede numbered ones at the bottom of the table.

3) Place numbered footnotes next.

4) The general note comes next; bring all lines for notes flush left. There may be more than one note, but they are all reported in one NOTE: section.

5) The last entry at the bottom of the table is the SOURCE: Department name, agency name, major survey or publication title, subsurvey title (in quotes), and year of survey or publication.

6) For unpublished data, use the month and year of the tabulation or tape file.

7) For up to three years of data, state each year. For more than three continuous years, give the year span. For more than three noncontinuous years, use "selected years" and the year span.

8) Use a semicolon to separate sources from the same agency. Use a period to separate agencies.

Attachment

The following list summarizes a number of statistical special notes and related set of symbols that should be used consistently across all NCES reports.

—	Not available	Data were not collected or not reported
†	Not applicable	Category does not exist
#	Rounds to zero	The estimate rounds to zero
!	Interpret data with caution	Estimates are unstable
‡	Reporting standards not met	State/Country did not meet reporting standards
*	$p<0.05$	Significance level

APPENDIX D

Under construction