GLOSSARY

-A-

Test **accommodations** are changes that are made in the content, format, and/or administration procedure of a test in order to accommodate test takers who are unable to take the original test under standard test conditions. 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 **African American or Black** person has 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."

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

An **Asian** person has 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.

An **assessment** is any systematic method of obtaining information from tests and other sources, used to draw inferences about characteristics of people, objects, or programs.

-B-

Balanced repeated replication (BRR) is a form of balaced sampling scheme used for variance estimation in sample surveys. Specifically, primary sampling units (PSUs) are grouped into strata, random samples of two PSUs are selected from each stratum and independent estimates of stratum characteristics are obtained from each PSU

The **base weight** is the initial sample weight that is the inverse of the probability of selection.

Bias is the deviation of the average survey value from the true population value. Bias refers to systematic errors that affect any sample taken under a specific design with the same constant error.

A **Black or African American** person has 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."

Bonferroni adjustment is a procedure for guarding against an increase in the probability of a Type I error when performing multiple significance tests. To maintain the probability of a Type I error at some selected value alpha, each of the m tests to be performed is judged against a significance level, alpha/m.

Bootstrap is a computer-intensive resampling method that is a generalization of jackknife in which subsamples are drawn with replacement. For each new subsample drawn, the same estimate is computed (e.g., mean or proportion) and the sampling variance among these replicated estimates is considered to be an approximation of the variance of the original full sample variance.

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.

-C-

The **capture-recapture** method is used to estimate the size of a population, N, that involves two independent draws of a sample from the same population, in which the cases in the first sample, n1, are identified. The second sample, n2, is drawn and the cases in the second sample that were identified in the first sample are counted, m. The size of

the population is then estimated as the ratio of the product of the two sample sizes to the number of cases in the second sample that were identified in the first sample (N=(n1n2)/m).

A **Census designated place *is a*** statistical geographic entity that is equivalent to an incorporated place, defined for the decennial census, consisting of a locally recognized, unincorporated concentration of population that is identified by name.

City, Large: Territory inside an urbanized area and inside a principal city with population of 250,000 or more.

City, Midsize: Territory inside an urbanized area and inside a principal city with population less than 250,000 and greater than or equal to 100,000.

City, Small: Territory inside an urbanized area and inside a principal city with population less than 100,000.

Classical test theory involves a psychometric theory based on the view that an individual's observed score on a test is the sum of a true scorecomponent for the test taker, plus an independent measurement error component.

Cluster sampling is an economical method for sampling a geographically scattered population. The population is divided into a large number of geographically compact regions (clusters) and a random sample of clusters is selected. Clusters may also be naturally occuring groupings in society such as households, schools, or places of business, where cases are sampled within the clusters of interest.

Coarsening involves disclosure control methods that reduce the amount of information included in a public use data file. These methods preserve the individual respondent's data by reducing the level of detail used to report some variables. Examples of coarsening methods include: top and/or bottom coding extreme values with less extreme values, the categorization of data into sufficiently broad categories to minimize the risk of disclosure, and the removal of potentially disclosive items.

Coding involves converting information into numbers or other symbols that can be more easily counted and tabulated.

Cognitive interviews are use to develop questionnaires. In a cognitive interview, respondents are required to report aloud everything they are thinking as they attempt to answer a survey question.

A **cohort** is a specific subpopulation whose members are linked together through a shared characteristic (e.g., a group of individuals in the same grade in school) or a common time factor.

A **Combined Statistical Area** is a geographic entity consisting of two or more adjacent Core Based Statistical Areas (CBSAs) with employment interchange measures of at least 15. Pairs of CBSAs with employment interchange measures of at least 25 combine automatically. Pairs of CBSAs with employment interchange measures of at least 15, but less than 25, may combine if local opinion in both areas favors combination. Note, the employment interchange measure of ties between two adjacent entities is the sum of the percentage of employed residents of the smaller entity who work in the larger entity and the percentage of employment in the smaller entity that is accounted for by workers who reside in the larger entity.

Confidentiality involves the protection of individually identifiable data from unauthorized disclosures.

Confidentiality edits include the use of one or more data perturbation techniques to add disclosure protection while minimizing the impact on the data. These techniques are used to alter responses in the microdata file before tabulations are produced.

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.
Construct validity is a term used to indicate that the test scores are to be interpreted as indicating the test taker's standing on the psychological construct measured by the test. A construct is a theoretical variable inferred from multiple types of evidence, including for example interrelations of the test scores with other variables, internal test structure, observations of response processes, and the content of the test.
A **Core Based Statistical Area (CBSA)** is a statistical geographic entity consisting of the county or counties associated with at least one core (urbanized area or urban cluster) of at least 10,000 population, plus adjacent counties having a high degree of social and economic integration with the core as measured through commuting ties with the counties containing the core. Metropolitan and Micropolitan Statistical Areas are the two categories of Core Based Statistical Areas. The core is a densely settled concentration of population, comprising either an urbanized area (of 50,000 or more population) or an urban cluster (of 10,000 to 49,999 population) defined by the Census Bureau, around which a Core Based Statistical Area is defined.
Covariance is a characteristic that indicates the strength of relationship between two variables. It is the expected value of the product of the deviations of two random variables, x and y from their respective means.
Coverage refers to the extent to which all elements on a frame list are members of the population, and to which every element in a population appears on the frame list once and only once.
Coverage error refers to the discrepancy between statistics calculated on the frame population and the same statistics calculated on the target population. ***Undercoverage*** or noncoverage errors occur when target population units are missed during frame construction, and ***overcoverage*** errors occur when units are duplicated or enumerated in error.
A **crosswalk study** delineates how categories from one classification system are related to categories in a second classification system.
A **cross-sectional** sample survey is based on a representative sample of respondents drawn from a population at one point in time.
Cross-sectional imputations are based on data from a single time period.
Cross-wave imputations are imputations based on data from multiple time periods. For example, a cross-sectional imputation for a time 2 salary could simply be a donor's time 2 salary. Alternatively, a cross-wave imputation could be the change in a donor's salary from time 1 to time 2 multiplied by the time 1 nonrespondent's salary.
A **cut-off sample** is a nonprobability sample that consists of the units in the population that have the largest values of a key variable (frequently the variable of interest from a previous time period). For example, a 90% cut-off sample consists of the largest units accounting for at least 90% of the population total of the key variable. Sample selection is usually done by sorting the population in decreasing order by size, and including units in the sample until the percent coverage exceeds the established cut-off.
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.

-D-
Data protection involves techniques that are used to insure that confidential individually identifiable data are not disclosed.

Data series are repeated collections of sequential cross-sectional or longitudinal data characteristics of the target population over time.

Data swapping involves the swapping or switching of a set of related items across similar, but distinct, records (i.e., between cases that are matched on other items). A simplistic example of data swapping would be to assume a data file has two potential variables that could potentially identify a respondent in the sample, for example, sex and age. If a sample case needs disclosure protection, it is paired with another sampled case on one of the two variables (e.g., age) so that both respondents in the pair have 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 **derived score** is a score to which raw scores are converted by numerical transformation (e.g., conversion of raw scores to percentile ranks or standard score).

The **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. Design effects differ for different subgroups and different statistics; no single design effect is universally applicable to any given survey or analysis.

DEFT is the square root of a design effect.

Differential item functioning (DIF) is a statistical property of a test item in which different groups of test takers who have the same total test score have different average item scores or, in some cases, different rates of choosing various item options. For example, DIF may exist when examinees of equal ability differ on an item solely because of their membership in a particular group.

Direct survey-based estimates are intended to achieve efficient and robust estimates of the true values of the target populations, based on the sample design and resulting survey data.

A **disability** is a physical or mental impairment which results in a need for special education and related services. Physical or mental impairment includes mental retardation, hearing impairments (including deafness), speech or language impairments, visual impairments (including blindness), serious emotional disturbance, orthopedic impairments, autism, traumatic brain injury, other health impairments, or specific learning disabilities, and in the case of children ages 3 through 9, developmental delays

Disclosure means the public release of individually identifiable data that were obtained under a pledge of confidentiality.

Disclosure risk analysis is conducted to identify records that could result in an exact disclosure, records that result in an unacceptably narrow estimation of a respondent's confidential information, and/or records that could be matched with external data sources to reveal the identity of an individual respondent. Such records require masking to produce a public-use data file from a restricted-use data file.

Domain refers to a defined universe of knowledge, skills, abilities, attitudes, interests, or other human characteristics.

With **dual frame sampling** units within the sample of interest are selected with independent probability samples taken from each of two frames. The resulting estimates from the two frames are combined to form a single composite dual frame estimate of the population parameters of interest. Dual frame estimates offer coverage rates that may exceed those of either of the single frames. Sometimes the best available list is known to

have poor coverage and there are no known supplemental frames to provide sufficient coverage. In this instance an area frame could be used as the second frame.

-E-

Editing includes procedures for detecting and correcting errors in the data. The data **edits** use available information and some assumptions to derive substitute values for inconsistent values in a data file.

Effect size refers to the standardized magnitude of the effect or the departure from the null hypothesis. For example, the effect size may be the amount of change over time, or the difference between two population means, divided by the appropriate population standard deviation. Multiple measures of effect size can be used (e.g., standardized differences between means, correlations, and proportions).

The **effective sample size**, as used in the design phase, 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 (i.e., the effective sample size is the original unweighted sample size divided by the design effect).

An **eligible sample unit** is a unit selected for a sample, which is confirmed to be a member of the target population.

Equating involves putting two or more essentially parallel tests on a common scale.

Estimates result from the process of providing a numerical value for a population parameter on the basis of information collected from a survey and/or other sources.

Estimation is the process of using data from a survey and/or other sources to provide a value for an unknown population parameter (such as a mean, proportion, correlation, or effect size), or to provide a range of values in the form of a confidence interval.

Estimation error is the difference between a survey estimate and the true value of the target population.

-F-

Facsimile is an exact copy or reproduction, in this case of the data collection items.

Fairness is the testing principle that every test taker should be assessed in an equitable way. Construct irrelevant personal characteristics such as race, ethnicity, sex, or disability have no appreciable effect on test results or their interpretation.

In a **field test,** all or some of the survey procedures are tested on a small scale that mirrors the planned full-scale implementation.

A **focus group** involves a semi-structured group discussion of a topic.

Forecasts involve the specific projection that an investigator believes is most likely to provide an accurate prediction of a future value of some process.

A **frame** is a mapping of the universe elements (i.e., sampling units) onto a finite list (e.g., the population of schools on the day of the survey).

The **frame population** is the set of elements that can be enumerated prior to the selection of a survey sample.

Freshening a sample involves augmenting a sample in a follow-up wave of a longitudinal study with respondents who are representative of new entrants to the population. This addition supports cross-sectional nationally representative estimates of a specific population at the time of the follow-up. For example, in a longitudinal study that starts with sophomores, the cohort is followed-up in the senior year to examine changes

over time within the initial cohort; but in the senior year the sample may be augumented with new entrants to the population (immigration and students who were skipped or retained a grade) to produce a nationally representative sample of seniors.

-H-

Half-open interval is a linking procedure that is used to increase coverage (i.e., reduce undercoverage). In this technique, new in-scope units between a unit A on the existing frame up to, but not including, unit B (the next unit on the existing 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 existing frame. The new units associated with the actual sample cases are now included in the sample with their respective selection probabilities.
A **Hispanic or Latino** person is 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."
Hypothesis testing draws a conclusion about the tenability of a stated value for a parameter. For example, sample data may be used to test whether an estimated value of a parameter (such as the difference between two population means) is sufficiently different from zero that the null hypothesis, designated H_0 (no difference in the population means), can be rejected in favor of the alternative hypothesis, H_1 (a difference between the two population means).

-I-

Imputation is a procedure that uses available information and a set of assumptions to derive substitute values for missing values in a data file.
An **Individualized Education Plan (IEP)** describes a child's present level of educational performance, presents the short term objectives or benchmarks that represent annual goals for the child, includes a list of the special education and related services provided to the child, explains the extent to which the child will not participate with nondisabled children in regular classes and school activities, describes modifications to state and district-wide assessments required for the child to participate or explain why the test is not appropriate and how the child will be tested, and presents the schedule for services provided for the child. (Title 42 U.S.C. Section 144 (d))
Individually identifiable data refers specifically to data from any list, record, response form, completed survey, or aggregation about an individual or individuals from which information about particular individuals or their schools/education institutions may be revealed by either direct or indirect means.
Instrument refers to an evaluative device that includes tests, scales, and inventories to measure a domain using standardized procedures. It is commonly used in surveys to refer to the device used to collect data, such as a questionnaire or data entry software.
Item nonresponse occurs when a respondent fails to respond to one or more relevant item(s) on a survey.
Item response theory (IRT) involves mathematical models that postulate that the probability of correct responses to a set of test questions is a function of the test taker's level on the psychological trait being assessed (e.g., mathematical ability) and of one or more parameters specific to each test question (e.g., item difficulty).

-J-

Jackknife is a method used for estimating standard errors of estimates obtained from complex sample surveys. The jackknife method uses overlapping samples made up of the whole sample, dropping each of the subsamples in turn in the computation of the estimate of interest and then the variance of these estimates is computed over the set of part-samples.

-K-

Key variables include survey-specific items for which aggregate estimates are commonly published from a study. They include, but are not restricted to, variables most commonly used in table row stubs. Key variables also include important analytic composites and other policy-relevant variables that are essential elements of the data collection. They are first defined in the initial planning stage of a survey, but may be added to as the survey and resulting analyses develop. For example, a study of student achievement might use gender, race-ethnicity, urbanicity, region, and school type (public/private) as key reporting variables.

-L-

A **Latino or Hispanic** person is 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."

Linkage is the result of placing two or more tests on the same scale, so that scores can be used interchangeably.

Longitudinal analysis involves the analysis of data from a study in which subjects are measured repeatedly over time.

A **longitudinal** sample survey follows the experiences and outcomes over time of a representative sample of respondents (i.e., a cohort) who are defined based on a shared experience (e.g., shared birth year or grade in school).

-M-

Response to a **mandatory survey** is required by law.

Measurement error is the difference between observed values of a variable recorded under similar conditions and some fixed true value (e.g., errors in reporting, reading, calculating, or recording a numerical value). Potential sources of measurement error include the respondent, the instrument, and the interviewer.

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource (i.e., data about data). Metadata is essential to ensuring thst resources will survive and continue to be accessible into the future.

A **Metropolitan Division** includes a county or group of counties within a Core Based Statistical Area that contains a core with a population of at least 2.5 million. A Metropolitan Division consists of one or more main/secondary counties that represent an employment center or centers, plus adjacent counties associated with the main county or counties through commuting ties. A main county acts as an employment center within a Core Based Statistical Area that has a core with a population of at least 2.5 million. A secondary county acts as an employment center in combination with a main county or another secondary county within a Core Based Statistical Area that has a core with a population of at least 2.5 million.

A **Metropolitan Statistical Area** is a Core Based Statistical Area associated with at least one urbanized area that has a population of at least 50,000. The Metropolitan Statistical Area comprises the central county or counties containing the core, plus adjacent outlying counties having a high degree of social and economic integration with the central county as measured through commuting. The central county or counties of a Core Based Statistical Area contain a substantial portion of an urbanized area or urban cluster or both, commuting is measured from the central county to determine qualification of outlying counties.

A **micro data** file includes the detailed responses for individual respondents.

A **Micropolitan Statistical Area** is a Core Based Statistical Area associated with at least one urban cluster that has a population of at least 10,000, but less than 50,000. The Micropolitan Statistical Area comprises the central county or counties containing the core, plus adjacent outlying counties having a high degree of social and economic integration with the central county as measured through commuting.

The **minimum substantively significant effect (MSSE)** is the smallest effect, that is, the smallest departure from the null hypothesis, 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 or perhaps, the minimum correlation, r, between two 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.

Missing at random for a given survey variable, refers to a situation in which the probability that a unit is missing that variable is independent of its value, but may not be independent of another variable(s).

Missing completely at random for a given survey variable, refers to a situation in which the probability that a unit is missing that variable is independent of its value, and is also independent of the values or other variables.

Model validation involves testing a model's predictive capabilities by comparing the model results to "known" sources of empirical data.

Multiple comparisons involve a detailed examination of the differences between a set of means.

Multiplicity estimation Multiplicity occurs within a sampling frame when a member of the population is linked to more than one entry on the frame, so that the member has multiple chances of being selected. Multiplicity estimation is a technique used to adjust selection probabilities when the unit of interest has multiple chances of being selected. For example, a survey might ask the household respondent to report how many children in their household are home schooled and how many children on their block are homeschooled (i.e., every homeschooled child on the block is linked to every sample household on the block). If multiple households on the same block are interviewed, multiplicity estimation is used to account for multiple reports of the same set of homeschooled children.

Multistage sampling occurs when the sampling is performed in two or three stages, with the sampling units at each stage being sampled from the units chosen at the previous stage. For example, a sample survey of teachers might start with a sample of school districts, with schools sampled within districts, and with teachers sampled within schools.

Multivariate analysis is a generic term for many methods of analysis that are used to investigate multivariate data. The principal techniques of multivariate

Multivariate data include data for which each record consists of values for more than one random variable.
Multivariate modeling provides a formalized mathematical expression of the process assumed to have generated the observed multivariate data.

-N-
A **Native Hawaiian or Other Pacific Islander** person has origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
The **NCES Locale Codes** include:
City, Large: Territory inside an urbanized area and inside a principal city with population of 250,000 or more.
City, Midsize: Territory inside an urbanized area and inside a principal city with population less than 250,000 and greater than or equal to 100,000.
City, Small: Territory inside an urbanized area and inside a principal city with population less than 100,000.
Suburb, Large: Territory outside a principal city and inside an urbanized area with population of 250,000 or more.
Suburb, Midsize: Territory outside a principal city and inside an urbanized area with population less than 250,000 and greater than or equal to 100,000.
Suburb, Small: Territory outside a principal city and inside an urbanized area with population less than 100,000.
Town, Fringe: Territory inside an urban cluster that is less than or equal to 10 miles from an urbanized area.
Town, Distant: Territory inside an urban cluster that is more than 10 miles and less than or equal to 35 miles from an urbanized area.
Town, Remote: Territory inside an urban cluster that is more than 35 miles from an urbanized area.
Rural, Fringe: Census-defined rural territory that is less than or equal to 5 miles from an urbanized area, as well as rural territory that is less than or equal to 2.5 miles from an urban cluster.
Rural, Distant: Census-defined rural territory that is more than 5 miles but less than or equal to 25 miles from an urbanized area, as well as rural territory that is more than 2.5 miles but less than or equal to 10 miles from an urban cluster.
Rural, Remote: Census-defined rural territory that is more than 25 miles from an urbanized area and is also more than 10 miles from an urban cluster.
A **New England City and Town Area (NECTA)** is a statistical geographic entity that is defined using cities and towns as building blocks and that is conceptually similar to the Core Based Statistical Areas in New England (which are defined using counties as building blocks).
A **New England City and Town Area (NECTA) Division** is a city or town or group of cities and towns within a NECTA that contains a core with a population of at least 2.5 million. A NECTA Division consists of a main city or town that represents an employment center, plus adjacent cities and towns associated with the main city or town, or with other cities and towns that are in turn associated with the main city or town, through commuting ties. A main city or town acts as an employment center within a New England City and Town Area that has a core with a population of at least 2.5 million.
Nonprobabilistic methods—see “probabilistic methods.”

Noncoverage involves eligible units of the target population that are missing from the frame population; this includes the problems of incomplete frames and missing units.
Nonresponse in sample surveys is the failure to obtain information from an eligible sample unit/respondent for any reason. Nonresponse can occur at the unit/respondent level due to a failure to contact or a refusal to respond, or at the item level due to a respondent's failure or refusal to respond to some, but not all, of the items on a survey.
Nonresponse bias occurs when the 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.
Nonresponse error is the overall error observed in estimates caused by differences between respondents and nonrespondents.
Nonsampling error includes measurement errors due to interviewers, respondents, instruments, and mode; nonresponse error; coverage error; and processing error.

-O-
An **Other Pacific Islander or Native Hawaiian** person has origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
An **On-line Analysis Tool (OATS)** 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 OAT. As an example, a OAT 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.
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.
Overcoverage errors occur when units are duplicated or enumerated in error.

-P-
The ***p*** **value** is the probability of the observed data, or data showing a more extreme departure from the null hypothesis, occurring when the null hypothesis is true.
Perturbation techniques are disclosure control methods that change the data before dissemination to decrease the disclosure risk while retaining as much information as possible. This involves purposely introducing error for confidentiality purposes using techniques. 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 are all examples of perturbation techniques.
In a **pilot test,** a laboratory or a very small-scale test of a questionnaire or procedure is conducted.
A **planning document** includes the information needed to reach a decision to proceed with survey design and development, including a study justification, review of related

studies, identification of potential privacy and confidentiality issues, justification for reocurring data items, and preliminary survey design, analysis plan, time schedule, dissemination plan, evaluation plan, and cost estimate.

Point estimate is the realized value of a statistic used to estimate a parameter for a particular sample of the data.

Population—see "target population."

Post-stratification adjustments are applied to survey data, in which sample units are stratified after data collection using information collected in the survey and auxiliary information to adjust weights to population control totals.

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.

The **power** (1 – b) of a test is defined as the probability of rejecting the null hypothesis when a specific alternative hypothesis is assumed. For example, with b = 0.20 for a particular alternative hypothesis, the power is 0.80, which means that 80 percent of the time the test statistic will fall in the rejection region if the parameter has the value specified by the alternative hypothesis.

Precision of survey results refers to how closely the results from a sample can reproduce the results that would be obtained from a complete count (i.e., census) conducted using the same techniques. The difference between a sample result and the result from a complete census taken under the same conditions is known as the precision of the sample result.

A survey **pretest** involves experimenting with different components of the questionnaire or survey design or operationalization prior to full-scale implementation. This may involve **pilot testing**, that is a laboratory or a very small-scale test of a questionnaire or procedure, or a ***field*** **test** in which all or some of the survey procedures are tested on a small scale that mirrors the planned full-scale implementation.

A **Principal City** is the largest city of a Core Based Statistical Area, plus additional cities that meet specified statistical criteria based on incorporated places and Census designated places.

Privacy is an individual's right to decide whether or not to share personal information.

Probabilistic methods for survey sampling are any of a variety of methods for sampling that give a known, non-zero, probability of selection to each member of the target population. The advantage of probabilistic sampling methods is that sampling error can be calculated. Such methods include: random sampling, systematic sampling, and stratified sampling. They do not include: convenience sampling, judgment sampling, quota sampling, and snowball sampling.

Probability of selection in a survey is the probability that a given sampling unit will be selected, based on the probabilistic methods used in sampling. In a simple random selection, this probability is the number of units drawn in the sample divided by the number of units on the sampling frame.

Probability proportional to size sampling (PPS) is a class of unequal probability sampling in which the probability of a unit being sampled is proportional to the level in that unit of some known characteristic of a cluster that is associated with the variable of interest in the sample survey (e.g., the number of students enrolled in a school for a study of schools and teachers).

A **projection** is an estimate of a future value of a characteristic based on current and past trends.

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 include two forms of disclosure control or limitation techniques: 1) confidentiality edits applied to the restricted-use data that underlie a public-use data file and 2) coarsening disclosure methods. Public-use edits are based on an assumption that external users have access to both individual respondent records, or information on the 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.

-Q-
Quality assurance processing includes any procedure or method that is aimed at maintaining or improving the reliability or validity of the data.

-R-
Raking is a multiplicative weighting technique that uses iterative proportional fitting. That is, weights are obtained as the product of a number of factors contributed by auxiliary variables.
In **ratio estimation**, an auxiliary variable, x_i, correlated with y_i is obtained for each unit in the sample. The population total X of the x_i must be known. In practice, x_i is often the value of y_i at some previous time when a complete census was taken. The goal is to obtain increased precision by taking advantage of the correlation between y_i and x_i. The ratio estimate of Y the population total of y_i is YR = (y/x), where y and x are the sample totals of y_i and $x_{i,}$ respectively.
A **record layout** is a description of the data elements on the file (variable names, data types, and length of space on the file) and their physical locations.
The **reference year** is the year for which the data are collected (e.g., the data collected in the 2011-12 Common Core of Data nonfiscal school district survey data for high school dropouts and graduates for the 2010-2011 reference school year).
Rejection region or critical region is the set of values of the statistic in a hypothesis test which lead to rejection of the null hypothesis.
Reliability is the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and thus are inferred to be dependable, and repeatable for an individual test; the degree to which scores are free of errors of measurement for a given group.
Replication method are approximate variance methods that estimate the variance based on the variability of estimates formed from subsamples of the full sample. The subsamples are generated to properly reflect the variability due to the sample design
Required response items include the minimum set of items required for a case to be considered a respondent.
Respondent burden is the estimated total time and financial resources expended by the survey respondent to generate, maintain, retain, and provide survey information.
Response rates calculated using base weights measure the proportion of the sample frame that is represented by the responding units in each study.

A **restricted-use data file** has no direct identifiers for individual respondents and typically includes data that have undergone confidentiality edits. However, restricted-use data files are not required to include variables that have undergone coarsening disclosure risk edits. Thus, a restricted-use data file includes individually identifiable information that is confidential and protected by law. Restricted-use data files are available for use by agency staff and contractors to produce official statistics, and for use by licensed qualified external researchers.
Rural, Distant: Census-defined rural territory that is more than 5 miles but less than or equal to 25 miles from an urbanized area, as well as rural territory that is more than 2.5 miles but less than or equal to 10 miles from an urban cluster.
Rural, Fringe: Census-defined rural territory that is less than or equal to 5 miles from an urbanized area, as well as rural territory that is less than or equal to 2.5 miles from an urban cluster.
Rural, Remote: Census-defined rural territory that is more than 25 miles from an urbanized area and is also more than 10 miles from an urban cluster.

-S-
Sampling error is the error associated with nonobservation, that is, the error that occurs because all members of the frame population are not measured. It is the error associated with the variation in samples drawn from the same frame population. The variance equals the square of the sampling error.
Sampling units are the basic components of a sample frame. Everything covered by a sample frame must belong to one definite sampling unit, or have a measurable probability of belonging to a specific unit. The sampling unit may contain, for example, houses, people, or businesses.
Scaling is the process of creating a scale or scale score. Scaling may enhance test score interpretation by placing scores from different tests or test forms onto a common scale or by producing scale scores designed to support criterion-referenced or norm-referenced score interpretations.
Scoring/rating is the assignment of a numeric value to a response or assessment of an individual.
Section 504 is a federal law designed to protect the rights of individuals with disabilities in programs and activities that receive Federal financial assistance from the U.S. Department of Education (ED). Section 504 provides: "No otherwise qualified individual with a disability in the United States . . . shall, solely by reason of her or his disability, be excluded from the participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance . . . ."
Section 508 is a federal law that requires Federal agencies to ensure that all Electronic and Information Technology (E&IT) procured be accessible by individuals with disabilities. The use of designs or technologies as alternatives to those explicitly prescribed by Accessibility Standards are allowed, provided that they result in substantially equivalent or greater access to, and use of, a product for individuals with disabilities.
Sensitivity analysis is designed to determine how the variation in the output of a model (numerical or otherwise) can be apportioned, qualitatively or quantitatively, to changes in input parameter values and assumptions. This type of analysis is useful in ascertaining the capability of a given model, as well its robustness and reliability.

Simple comparison is a test (such as a t test or a z test), of the difference between two means or proportions.
Simple random sampling (SRS) uses equal probability sampling with no strata or clusters (i.e., every member of the population has an equal chance of being chosen).
Suburb, Large: Territory outside a principal city and inside an urbanized area with population of 250,000 or more.
Suburb, Midsize: Territory outside a principal city and inside an urbanized area with population less than 250,000 and greater than or equal to 100,000.
Suburb, Small: Territory outside a principal city and inside an urbanized area with population less than 100,000.
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 error is the standard deviation of the sampling distribution of a statistic.
Statistical disclosure techniques or disclosure control/limitation techniques are used to prepare microdata files for release. These techniques include perturbation techniques that change the data before dissemination (e.g., confidentiality edits and the removal of direct identifiers) and disclosure control techniques that reduce the amount of detail in the information provided to users of public-use data (e.g., coarsening and the removal of data items).
Statistical inference involves the process of deducing properties of the underlying distribution by analysis of data.
Statistical significance is attained when a statistical procedure applied to a set of observations yields a *p* value that exceeds the level of probability at which it is agreed that the null hypothesis will be rejected.
Strata are created by partitioning the frame and are generally defined to include relatively homogeneous units within strata**.**
Stratification is the division of a population into parts, known as strata, for the purpose of drawing a sample. Stratification may be based on geography or on some other characteristic of the population.
Substitutions are done using matched pairs, in which the alternate member of the pair does not have an independent probability of selection.
A **supplemental area frame** can be created and used to supplement the estimates when undercoverage is a problem in a list frame. This is often done by first, generating a frame of geographic units where all the geographic units are represented, thus 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.
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.
The term **survey year** refers to the year the survey was conducted (e.g., the 2011-12 Common Core of Data nonfiscal school district survey was conducted during the 2011-13 school year).

-T-
The **tail** area of a distribution is the portion of the area under a frequency curve which lies between the start of the distribution and some point lying between the start and the mode. For example, the tail of the sampling distribution of the test statistic contains the rejection region for the hypothesis tested, H0.
The **target population** is the population about which information is desired. It can be any group of potential sample units or persons, businesses, or other entities of interest.
Taylor-series linearization is an approximate variance method in which an estimate is linearized as a first step. The variance of the linearized estimate is then computed using either an exact or approximate variance formula appropriate for the sample design.
Test blueprint is a detailed description for a test, also known as a test specification, that specifies the number or proportion of items, responses, and scoring rubrics and procedures; and the desired psychometric properties of the items and test such as the distribution of item difficulty and discrimination indices.
The **total mean square error** is a measure of the combined overall effect of sampling and nonsampling error on the estimate.
Total nonresponse is the combination of unit nonresponse across all the stages of a data collection (i.e., overall nonresponse) with item nonresponse for a specific item.
Town, Distant: Territory inside an urban cluster that is more than 10 miles and less than or equal to 35 miles from an urbanized area.
Town, Fringe: Territory inside an urban cluster that is less than or equal to 10 miles from an urbanized area.
Town, Remote: Territory inside an urban cluster that is more than 35 miles from an urbanized area.
Weight **trimming** involves restricting the range of the weights to decrease the variance associated with extremely large or small weights.
Type I error is made when the tested hypothesis, H0, is falsely rejected when in fact it is assumed 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 in 5 percent of tests where the null hypothesis is true.
Type II error is made when the tested hypothesis H0, is not rejected when it is false (i.e., a specific alternative hypothesis, H1, is assumed true). The probability of making a Type II error is denoted by beta (b). For example, with a beta level of 0.20, the analyst will conclude that no difference is present in 20 percent of all cases in which the specific hypothesized alternative, H1, is true.

-U-
Undercoverage errors occur when target population units are missed during frame construction.

Unduplication involves identifying cases that occur more than one time in the sample frame (e.g., a frame built from membership lists of organizations with overlapping membership) and removing multiple (i.e., duplicated) entries from the sample frame, so that each case only appears on the frame one time.
Unit nonresponse occurs when a respondent fails to respond to all required response item (i.e., fails to fill out or return a data collection instrument).
A **universe** survey involves the collection of data covering all known units in a population (i.e., a census).
An **Urban Cluster** is a statistical geographic entity delineated by the Census Bureau, consisting of densely settled census tracts and blocks and adjacent densely settled territory that together contain at least 2,500 people. For purposes of delineating Core Based Statistical Areas, only those urban clusters of 10,000 more population are considered.
An **Urbanized Area** is a statistical geographic entity delineated by the Census Bureau, consisting of densely settled census tracts and blocks and adjacent densely settled territory that together contain at least 50,000 people.

-V-
Validation studies are conducted to independently verify that the data collection methodology employed will obtain accurate data for the concept studied.
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.
Variance measures the spread of the distribution of a variable; and the standard deviation is the square root of the variance.
Vertical scaling involves the development of scales that span a broad range of developmental or educational levels for the evaluation of examinee growth over time.
Response to a **voluntary** survey is not required by law.

-W-
A **wave** is a round of data collection in a longitudinal survey (e.g., the base year and each successive followup are each waves of data collection).
Weights are relative values associated with each sample unit that are intended to correct for unequal probabilities of selection for each unit due to sample design. Weights most frequently represent the relative portion of the population that the unit represents. Weights may be adjusted for nonresponse.
A **White** person has origins in any of the original peoples of Europe, the Middle East, or North Africa.