

# Survey Glossary

<http://ccsg.isr.umich.edu/index.php/resources/advanced-glossary>

[[A]] [[X]]

## A

Term	Definition
Accuracy	The degree of closeness an estimate has to the true value.
Adaptation	Changing existing materials (e.g., management plans, contracts, training manuals, questionnaires, etc.) by deliberately altering some content or design component to make the resulting materials more suitable for another socio-cultural context or a particular population. In the context of questionnaire translation, the limits between translation and adaptation are difficult to define because almost every translation includes adaptation to a certain degree. Therefore, sometimes both terms are used in combination ("translation and adaptation") to name the process of making a questionnaire fit for use in another language and culture. For a discussion of adaptation and, for instance, its different forms, see also (Behr & Shishido, forthcoming). See Adaptation for more information.
Adaptive behavior	Interviewer behavior that is tailored to the actual situation encountered.
Adjudication	The translation evaluation step at which a translation is signed off and released for whatever follows next such as pretesting or final fielding (the 'A' in the TRAPD method, see Translation). When all review and refinement procedures are completed, including any revisions after pretesting and copyediting, a final signing off/ is required. Thus, in any translation effort there will be one or more signing-off steps ("ready to go to external assessment," "ready to go to client," "ready to go to fielding agency," for example).
Adjudicator	The person who signs-off on a finalized version of a questionnaire (see Adjudication).
Adjustment Error	Survey error (variance and due to post data collection statistical adjustment.
Advance translation	A translation is made of a source questionnaire to try to find problems in the source text that only become apparent when translation is attempted. The insights are used to modify the source questionnaire or plan for We recommend to carry out the translation using the team approach so as to receive input to the one expected during the final translation phase. Comments made in the course of advance translation typically concern both linguistic / translation-related as well as intercultural issues (see Dorer (2011)).
Anchoring vignettes	A technique used to adjust for noncomparability in self-assessment questions caused by differences in response scale usage across groups. It relies on a set of descriptions (usually brief) of hypothetical people and situations to which self-assessment is calibrated (King, Murray, Salomon, & Tandon, 2004).
Annotation	Information appended to text in the source questionnaire to help clarify the intended meaning of a source text concept, phrase, or term. (See Appendix A-D for further detail and examples of the use of 'Annotations' are also referred to as 'footnotes'.

Anonymity	Recording or storing information without name or identifier, so the respondent cannot be identified in any way by anyone. No one can link an individual person to the responses of that person, including the investigator or the interviewer. Face-to-face interviews are never anonymous since the interviewer knows the address (and likely, the name) of the respondent.
Anonymization	Stripping all information from a survey data file that allows the re-identification of respondents (see confidentiality).
ASCII files	Data files in American Standard Code for Information Interchange (ASCII) format.
Ask different questions (ADQ)	An approach to question design where researchers collect data across populations or countries based on using the most salient population-specific questions on a given construct/research topic. The questions and indicators used in each location are assumed (or better, have been shown) to tap a construct that is germane or shared across populations.
Ask the same questions (ASQ)	An approach to question design whereby researchers collect data across populations/countries by asking a shared set of questions. The most common way to do this is by developing a source questionnaire in one language and then producing whatever other language versions are needed on the basis of translation or translation and Hence the description used in the chapter of "ASQ and translate (ASQT)". is a second way to "ask the same questions" but this procedure is differently organized.
Ask the same questions and translate (ASQT)	An approach to question design whereby researchers collect data across populations/countries by asking a shared set of questions. The most common way to do this is by developing a source questionnaire in one language and then producing whatever other language versions are needed on the basis of translation or translation and is a second way to "ask the same questions" but this procedure is differently organized.
Attitudinal question	A question asking about respondents' opinions, judgments, emotions, and perceptions. These cannot be measured by other means; we are dependent on respondents' answers. Example: Do you think smoking cigarettes is bad for the smoker's health?
Audio computer-assisted self-interviewing (A-CASI)	A mode in which the respondent in which the respondent uses a computer that plays audio recordings of the questions to the respondent, who then enters his/her answers. The computer may or may not display the questions on the screen.
Audit trail	An electronic file in which computer-assisted and Web survey software captures paradata about survey questions and computer user actions, including times spent on questions and in sections of a survey ( and interviewer or respondent actions while proceeding through a survey. The file may contain a record of and function keys pressed, as well as mouse actions.
Auxiliary data	Data from an external source, such as census data, that is incorporated or linked in some way to the data collected by the study. Auxiliary data is sometimes used to supplement collected data, for creating weights, or in techniques.
Base weight	The inverse of the probability of selection.

Behavior codes	“Behavior codes are information about the interviewer and respondent’s verbal behaviors during a survey interview’s question-answer process. They are developed and recorded by human coders, not automatically by computers. To obtain behavior interviews are audio recorded (generally digitally today, but cassette tapes have been used in the past), transcribed, and then coded by a set of at least two coders to identify relevant behaviors” (Olson & Parkhurst, 2013).
Behavior coding	Systematic coding of the interviewer-respondent interaction in order to identify problems and sometimes to estimate the of behaviors that occur during the question-answer process.
Behavioral question	A question asking respondents to report behaviors or actions. Example: Have you ever smoked cigarettes?
Bias	The systematic difference over all conceptual trials between the expected value of the survey estimate of a population parameter and the true value of that parameter in the target population.
Bid	A complete proposal (submitted in competition with other bidders) to execute specified jobs within prescribed time and budget, and not exceeding a proposed amount.
Bilingual glossary	A glossary is a list of words or phrases used in a particular field alongside their definitions. Glossaries are often found at the back of a specialist or academic book as an appendix to the text. A bilingual glossary lists special terms used in a particular field in two languages. A key notion or concept present in one language for a given field may not have a ready single match in a given other language.
Bottom coding	A type of coding in which values that exceed the predetermined minimum value are reassigned to that minimum value or are recoded as missing data.
Bridge language	A language, common to both interviewers and respondents, that is used for data collection but may not be the first language of either person.
Cause and effect diagram	A fishbone-structured diagram for a process, used as a brainstorming tool to help understand or improve the process. The main bone represents the process (e.g., interviewer training), and bones coming off of the main bone are pre-identified factors (e.g., training materials) that may affect the quality of the process. From there potential causes (lack of resources and time) and effects (poor quality materials) can be discussed, and solutions identified. Also known as a fishbone or Ishikawa diagram.
Certification	Objective assessment of performance. Based on pre-established criteria, the interviewer either meets the requirements and may proceed to conduct the study interview or does not meet the requirements and may either be permitted to try again or be dismissed from the study. Certification outcome should be documented and filed at the data collection agency.
Closed-ended question	A survey question format that provides a limited set of predefined answer categories from which respondents must choose. Example: Do you smoke?
Cluster	A grouping of units on the sampling frame that is similar on one or more variables, typically geographic. For example, an interviewer for an in person study will typically only visit only households in a certain geographic area. The geographic area is the cluster

Cluster sampling	A sampling procedure where units of the sampling frame that are similar on one or more variables (typically geographic) are organized into larger groups (i.e. and a sample of groups is selected. The selected groups contain the units to be included in the sample. The sample may include all units in the selected or a sub-sample of units in each selected cluster. The ultimate purpose of this procedure is to reduce interviewer travel costs for in person studies by producing distinct groups of elements where the elements within each group area are geographically close to one another.
Code structure	List of descriptions of variable categories and associated code numbers. Also referred to as code frame, coding frame, or codes.
Codebook	A document that provides question-level metadata that is matched to variables in a dataset. Metadata include the elements of a dictionary, as well as basic study documentation, question text, statements (the characteristics of respondents who were asked the question), the number of respondents who answered the question, and response frequencies or statistics.
Coding	Translating nonnumeric data into numeric fields.
Coefficients of variation	a measure of dispersion of a probability distribution or frequency distribution, that describes the amount of variability relative to the mean.
Cognitive interview	A pretesting method designed to uncover problems in survey by having respondents think out loud while answering a question or retrospectively.
Cohen's kappa	A statistical measure that accounts for degree of chance of agreements between coders.
Comparability	The extent to which differences between survey statistics from different countries, regions, cultures, domains, time periods, etc., can be attributable to differences in population true values (Johnson & Mohler, 2010). In other words, whether the concepts are comparable or not. It is often referred as "equivalence, equivalence, similarity, or some other frame of reference" (Johnson & Mohler, 2010). Improving implies that error due to translation has to be minimized. In terms of questionnaire translation for multi-national, multi-cultural and multi-regional surveys the aim is to achieve the defined statistical level of comparability across all local versions (expressed in minimized translation error).
Comparativist	A person who carries out comparative studies, especially a student of comparative literature or comparative linguistics.
Complex survey data (or designs)	Survey datasets (or designs) based on stratified single or multistage samples with survey weights designed to compensate for unequal probabilities of selection or nonresponse.
Computer assisted recorded interviewing (CARI)	A system for audio recording of interviews (or interview parts) that allow for monitoring interviewer performance in the field/call center and detection of data fraud.
Computer assisted telephone interviewing (CATI)	A telephone interviewing mode in which a computer displays the questions on a screen, the interviewer reads them to the respondent over the phone, and enters the respondent's answers directly into the computer.
Computer- assisted self interviewing (CASI)	A mode in which a computer displays the questions on a screen to the respondent and the respondent then enters his/her answers into the computer.
Computer-Assisted Personal Interviews (CAPI)	A face-to-face interviewing mode in which a computer displays the questions onscreen, the interviewer reads them to the respondent, and enters the respondent's answers directly into the computer.
Concurrent mixed mode	A mixed mode design in which one group of respondents uses one mode and another group of respondents uses another.

Confidentiality	Securing the identity of, as well as any information provided by, the respondent, in order to ensure to that public identification of an individual participating in the study and/or his individual responses does not occur.
Consent (informed consent)	A process by which a sample member voluntarily confirms his or her willingness to participate in a study, after having been informed of all aspects of the study that are relevant to the decision to participate. Informed consent can be obtained with a written form or orally (or implied if the respondent returns a mail survey), depending on the study protocol. In some cases, consent must be given by someone other than the respondent (e.g., an adult when interviewing children).
Consistency	Consistency is achieved when the same term or phrase is used throughout a translation to refer to an object or an entity referred to with one term or phrase in the source text. In many cases, consistency is most important with regard to technical terminology or to standard repeated components of a questionnaire. Reference to "showcard" in a questionnaire should be consistently translated, for example. The translation of instructions which are repeated in the source text should also be repeated (and not varied) in the target text.
Construct validity	The degree to which a survey question adequately measures an intended hypothetical construct. This may be assessed by checking the correlation between observations from that question with observations from other questions expected on theoretical grounds to be related.
Constructed variable	A recoded variable, one created by data producers or archives based on the data originally collected. Examples are age grouped into cohorts, income grouped into 7 categories, Goldthorpe-Index, or the creation of a variable called POVERTY from information collected on the income of respondents.
Contact attempt record	A written record of the time and outcome of each contact attempt to a sample unit.
Contact rate	The proportion of all elements in which some responsible member of the housing unit was reached by the survey.
Content management	The software and procedures used to capture, save, organize, and distribute information in digitalized form.
Context effects	The effect of question context, such as the order or layout of questions, on survey responses.
Contract	A legally binding exchange of promises or an agreement creating and defining the obligations between two or more parties (for example, a survey organization and the coordinating center) written and enforceable by law.
Convenience sample	A sample of elements that are selected because it is convenient to use them, not because they are representative of the target population.
Conversational interviewing	Interviewing style in which interviewers read questions as they are worded but are allowed to use their own words to clarify the meaning of the questions.
Conversion process	Data processing procedures used to create harmonized variables from original input variables.
Cooperation rate	The proportion of all elements interviewed of all eligible units ever contacted.
Coordinating center	A research center that facilitates and organizes cross-cultural or multi-site research activities.
Copyeditor	The person who reviews a text and marks up any changes required to correct style, punctuation, spelling, and grammar errors. In many instances, the copyeditor may also make the corrections needed.
Coverage	The proportion of the target population that is accounted for on the frame.

Coverage bias	The systematic difference between the expected value (over all conceptual trials) of a statistic and the target population value because some elements in the target population do not appear on the frame.
Coverage error,	Survey error (variance and that is introduced when there is not a one-to-one correspondence between frame and population units. Some units in the target population are not included on the frame (undercoverage), some units on the sampling frame are not members of the target population (out-of-scope), more than one unit on the sampling frame corresponds to the same target population unit (overcoverage), and one sampling frame unit corresponds to more than one target population unit.
Coverage rate	The number of elements on the sampling frame divided by the estimated number of elements in the population.
Coversheet	Electronic or printed materials associated with each element that identify information about the element, e.g., the sample address, the unique identification number associated with an element, and the interviewer to whom an element is assigned. The often also contains an introduction to the study, instructions on how to screen sample members and randomly select the respondent, and space to record the date, time, outcome, and notes for every contact attempt.
Crosswalk	A description, usually presented in tabular format, of all the relationships between variables in individual data files and their counterparts in the harmonized file.
Cultural schema	A conceptual structure, shared by members of a cultural group and created from common experiences, by which objects and events can be identified and understood.
Data capture	The process of converting data (e.g., from questionnaires, audio/visual recordings, samples, etc.) to an electronic file.
Data dictionary	A document linking the survey instrument (questionnaire) with the dataset, or more abstract question or variable-level metadata including question identifiers (variable names and labels); response category identifiers (value labels), and data types (e.g., F2.0, specifying that the response is a two-digit integer with zero decimal places).
Data Documentation Initiative (DDI)	An international effort to establish a standard for technical documentation describing social science data. A membership-based Alliance is developing the DDI specification, which is written in XML.
De-identification	Separating personally identifiable information (PII) from the survey data to prevent a breach of confidentiality.
Decentering	An approach to designing questions in two languages in which neither of the languages nor cultures involved is allowed to dominate. A Ping-Pong-like process of formulation and comparison between the two languages is used to develop versions in each language. Any language or cultural obstacles met with are resolved, often by removing or changing wording in one or both languages. The question formulation in both languages then moves on from that modification. Since the process removes culture-specific elements from both versions, decentered questions may be vague and not especially salient for either target population.
Design effect	The effect of the complex survey design on sampling variance measured as the ratio of the sampling under the complex design to the sampling variance computed as a simple random sample of the same sample size.
Differential item functioning (dif)	Item as a result of systematic differences in responses across cultures due to features of the item or measure itself, such as poor translation or ambiguous wording.

Diglossic linguistic contexts	Diglossic linguistic contexts exist in single language communities that use two or more markedly different varieties of a language or two different languages in different contexts. The variety used may be determined by whether the language is written or spoken in a given instance or by the relationships between participants in a discourse. Considerations such as age, gender, social status, and the topic under discussion may all contribute to the form chosen in any given instance.
Direct cost	An expense that can be traced directly to (or identified with) a specific cost center or is directly attributable to a cost object such as a department, process, or product.
Disclosure analysis and avoidance	The process of identifying and protecting the confidentiality of data. It involves limiting the amount of detailed information disseminated and/or masking data via noise addition, data swapping, generation of simulated or synthetic data, etc. For any proposed release of tabulations or the level of risk of disclosure should be evaluated.
Disposition code	A code that indicates the result of a specific contact attempt or the outcome assigned to a sample element at the end of data collection (e.g., ineligible, complete interview).
Document management system	A document management system (DMS) is a computer system (or a set of computer programs) used to track and store electronic documents and/or images of paper documents. The term has some overlap with the concept of Management Systems. It is often viewed as a component of Enterprise Content Management Systems (ECM, see <a href="http://www.aiim.org/What-is-ECM-Enterprise-Content-Management.aspx">http://www.aiim.org/What-is-ECM-Enterprise-Content-Management.aspx</a> ) and related to Digital Asset Management, Document imaging, Workflow systems and Records Management systems.
Double-barreled (questions)	Survey questions that inadvertently ask about two topics at once.
Editing	Altering data recorded by the interviewer or respondent to improve the quality of the data (e.g., checking correcting mistakes, following up on suspicious values, deleting duplicates, etc.). Sometimes this term also includes and the placement of a number into a field where data were missing.
element (Sample element)	A selected unit of the target population that may be eligible or ineligible.
Eligibility rate	The number of eligible sample elements divided by the total number of elements on the frame.
Embedded experiments	Embedded experiments are included within the framework of an actual study.
Ethics review committee or human subjects review board	A group or committee that is given the responsibility by an institution to review that institution's research projects involving human subjects. The primary purpose of the review is to assure the protection of the safety, rights and welfare of the human subjects.
Event history calendar	A conversational interviewing technique designed to improve recollection of complex sequences of personal events by using respondents' own past experiences as memory cues. This technique is sometimes referred to as life history calendar.
Ex-ante	The process of creating harmonized variables at the outset of data collection, based on using the same questionnaire or agreed definitions in the harmonization process.
Ex-post	The process of creating harmonized variables from data that already exists.
Expectation question	A question asking about respondents' expectation about the chances or probabilities that certain things will happen in the future.

Eye-tracking	Eye tracker tracks the time, duration and location of our eyes' fixations and the saccades between fixations. See Li (n.d.).
Fact sheet	A sheet, pamphlet, or brochure that provides important information about the study to assist respondents in making an informed decision about participation. Elements of a fact sheet may include the following: the purpose of the study, sponsorship, uses of the data, role of the respondent, sample selection procedures, benefits and risks of participation, and confidentiality.
Factual judgment question	A question that requires respondents to remember autobiographical events and use that information to make judgments. In principle, such information could be obtained by other means of observation, such as comparing survey data with administrative records, if such records exist. Factual judgment questions can be about a variety of things, such as figure-based facts (e.g., date, age, weight), events (e.g., pregnancy, marriage), and behaviors (e.g., smoking, media consumption).
Factual question	A question that aims to collect information about things for which there is a correct answer. In principle, such information could be obtained by other means of observation, such as comparing survey data with administrative records. Factual questions can be about a variety of things, such as figure-based facts (date, age, weight), events (pregnancy, marriage), and behaviors (smoking or media consumption). Example: Do you smoke?
Fitness for intended use	
Fixed panel design	A longitudinal study which attempts to collect survey data on the same elements at intervals over a period of time. After the initial sample selection, no additions to the sample are made.
Fixed panel plus birth	A longitudinal study in which a panel of individuals is interviewed at intervals over a period of time and additional elements are added to the sample.
Flow chart	A method used to identify the steps or events in a process. It uses basic shapes for starting and ending the process, taking an action, making a decision, and producing data and documentation. These are connected by arrows indicating the flow of the process. A flow chart can help identify points at which to perform assurance activities and produce indicators of that can be used in control.
Focus group	Small group discussions under the guidance of a moderator which can be used to explore topics, develop topics and items for questions. It is often used in qualitative research to test survey questionnaires and survey protocols.
frame (Sampling frame)	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Frequency format	A response format requiring the respondent to select the option that best describes the frequency in which certain behaviors occur.
Full translation (double/parallel translation)	Each translator translates all of the material to be translated. It stands in contrast to translations. This refers to the first step in the TRAPD model, the 'T'.
Functional equivalence	It measures to what degree the hypothetical construct serves "similar functions" within each society or cultural group. See Johnson (1998b) for more information.
Hadamard matrix	TABLE IMAGE

Hard consistency check	A signal warning that there is an inconsistency between the current response and a previous response; the interviewer or respondent cannot continue until the inconsistency is resolved.
Hours Per Interview (HPI)	A measure of study efficiency, calculated as the total number of interviewer hours spent during production (including travel, reluctance handling, listing, completing an interview, and other administrative divided by the total number of interviews.
Imputation variance	That component of overall variability in survey estimates that can be accounted for by imputation.
Imputations	A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.
Inconsistent responses	Inappropriate responses to branched questions. For instance, one question might ask if the respondent attended church last week; a response of "no" should skip the questions about church attendance and code the answers to those questions as "inapplicable." If those questions were coded any other way than "inapplicable," this would be inconsistent with the skip patterns of the survey instrument.
Indirect cost	An expense that is incurred in joint usage and difficult to assign to or is not directly attributable to a specific department, process or product.
Informant	The person who supplies a list of the eligible elements within the selected unit. For example, many in-person surveys select a sample of housing units at the penultimate stage of selection. Interviewers then contact the housing unit with the aim of convincing the member of the housing unit who responded to the contact attempt to provide a list of housing unit members who are eligible for the study. The housing unit member who provides a list of all eligible housing unit members is called the informant. Informants can also be selected respondents as well, if they are eligible for the study and are chosen as the respondent during the within household stage of selection.
Intention question	A question asking respondents to indicate their intention regarding some behavior.
Interactive Voice Response (IVR)	A telephone interviewing method in which respondents listen to recordings of the questions and they respond by using the keypad of the telephone or saying their answers aloud.
Interface design	Aspects of computer-assisted survey design focused on the interviewer's or respondent's experience and interaction with the computer and instrument.
Interpenetrated sample assignment, interpenetration	Randomized assignment of interviewers to subsamples of respondents in order to measure correlated response variance, arising from the fact that response errors of persons interviewed by the same interviewer may be correlated. allows researchers to disentangle the effects interviewers have on respondents from the true differences between respondents.
Interviewer design effect (Deffint)	The extent to which interviewer variance increases the of the sample mean of a simple random sample.
Interviewer effect	Measurement error, both systematic and variable, for which interviewers are responsible.

Interviewer falsification	Intentionally departing from the designed interviewer guidelines that could result in the contamination of the data. Falsification includes: 1) Fabricating all or part of an interview—the recording of data that are not provided by a designated survey respondent, and reporting them as answers of that respondent; 2) Deliberately misreporting disposition codes and falsifying process data (e.g., the recording of a respondent as ineligible for the sample; reporting a fictitious contact attempt); 3) Deliberately miscoding the answer to a question in order to avoid follow-up questions; 4) Deliberately interviewing a nonsampled person in order to reduce effort required to complete an interview; or intentionally misrepresenting the data collection process to the survey management.
Interviewer observations and evaluations	“In interviewer-administered surveys, interviewers have long been asked to make general assessments about how engaged, cooperative, hostile, or attentive the respondent was during the interview. Additionally, interviewers record information about the interview-taking environment, such as whether other individuals were present or whether the respondent used headphones during an ACASI component. Unlike the previous sources of paradata, these interviewer evaluations are questions asked directly of the interviewer and included as a few additional questions in the questionnaire” (Olson & Parkhurst, 2013).
Interviewer variance	That component of overall variability in survey estimates that can be accounted for by the interviewers.
Item	Researchers differ greatly in how they use this term. In survey research, it usually determines one question in a survey questionnaire, that is, each time the respondent is asked to give an answer.
Item nonresponse, item missing data	The lack of information on individual data items for a element where other data items were successfully obtained.
Item Response Theory (IRT)	A theory that guides statistical techniques used to detect survey or test questions that have item or item functioning (see dif). is based on the idea that the probability of a response an individual provides is a function of the person's traits and characteristics of the item.
Keystrokes	“Keystroke files, sometimes called trails or trace files” are recorded” when interviewers or respondents use specific keys during the survey. Keystroke files contain both response timing data and a record of the pressed during the questionnaire administration” (Olson & Parkhurst, 2013).
Latent class analysis (LCA)	Latent Class Analysis (LCA) is a subset of structural equation modeling, used to find groups or subtypes of cases in multivariate categorical data. These subtypes are called “latent classes”
Listing	A procedure used in area probability sample designs to create a complete list of all elements or of elements within a specific set of geographic boundaries.
Loaded questions/words	Questions that are worded in such a way that invite respondents to respond in a particular way.
Longitudinal study	A study where elements are repeatedly measured over time.
Mean Square Error	The total error of a survey estimate; specifically, the sum of the variance and the squared.
Measurement equivalence	Equivalence of the calibration system used in the questionnaire and the translation.
Measurement error	Survey error (variance and due to the measurement process; that is, error introduced by the survey instrument, the interviewer, or the respondent.

Metadata	Information that describes data. The term encompasses a broad spectrum of information about the survey, from study title to sample design, details such as interviewer briefing notes, contextual data and/or information such as legal regulations, customs, and economic indicators. Note that the term 'data' is used here in a technical definition. Typically are descriptive information and data are the numerical values described.
Microdata	Nonaggregated data that concern individual records for sampled units, such as households, respondents, organizations, administrators, schools, classrooms, students, etc. Microdata may come from auxiliary sources (e.g., census or geographical data) as well as surveys. They are contrasted with macrodata, such as variable means and frequencies, gained through the aggregation of microdata.
Minority country	A country with high per capita income (the minority of countries).
Mode	Method of data collection.
Mouse clicks	"Mouse click files record each action the respondent or interviewer takes using the computer's mouse, ranging from the presence or absence of simple single mouse clicks to the position of the mouse cursor at a specified time interval on an x – y coordinate of the survey page" (Olson & Parkhurst, 2013).
Multi-Trait-Multi-Method (MTMM)	A technique that uses the correlations between multiple methods (i.e. modes) and multiple traits (i.e. variables) to assess the of a measurement process.
Non-interview	A sample element is selected, but an interview does not take place (for example, due to or ineligibility).
Noncontact	Sampling units that were potentially eligible but could not be reached.
Nonresponse	The failure to obtain measurement on sampled units or items. See unit and item nonresponse.
Nonresponse bias	The systematic difference between the expected value (over all conceptual trials) of a statistic and the target population value due to differences between respondents and nonrespondents on that statistic of interest.
Nonresponse error,	Survey error (variance and that is introduced when not all sample members participate in the survey (unit or not all survey are answered (item nonreponse) by a element.
Nonresponse follow-up	A supplemental survey of sampled survey nonrespondents. Nonresponse follow-up surveys are designed to assess whether respondent data are biased due to differences between survey respondents and nonrespondents.
Open tendering	A bidding process in which all the bidders are evaluated and then chosen on the basis of cost and technical merit.
Open-ended question	A survey question that allows respondents to formulate the answer in their own words. Unlike a closed question format, it does not provide a limited set of predefined answers. Example: What is your occupation? Please write in the name or title of your occupation
Outcome rate	A rate calculated based on the study's defined final disposition codes that reflect the outcome of specific contact attempts before the unit was finalized. Examples include rates (the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.), rates (the proportion of all units interviewed of all eligible units ever contacted), rates (the proportion of all units in which a housing unit or respondent refuses to do an interview or breaks-off an interview of all potentially eligible units), and rates (the proportion of all units are reached by the survey).
Outlier	An atypical observation which does not appear to follow the distribution of the rest of a dataset.

Overediting	Extensive editing that becomes too costly for the amount of error that is being reduced.
Overlap in the split translations	A compromise solution between split and full translations is to ensure that some overlap exists between materials divided among The material is split up the way cards are dealt in many games, everyone getting a spread of the material. Each translator could then receive the last one or two questions of another translator’s “piece”. This allows the review team members to have an increased sense of differences in translating approaches between translators and their understanding of source text components at the draft production level.
Overrun	The exceeding of costs estimated in a contract.
Paradata	Couper first introduced the term “paradata” into survey research methodology field (Groves & Couper, 1998) and the definition of paradata has vastly expanded since then. Paradata now refers to additional data that can be captured during the process of producing a survey statistic (Kreuter, 2013). As discussed in the 2011 International Workshop (Smith, 2011), two main types of paradata are available. One is process paradata, which is collected during the process of data collection, such as stamps and data. Another type is related with observational information, such as the observed demographic information of respondents and observed neighborhood conditions.
Pareto chart	A bar chart that reflects the types of most errors in a process, by error type in descending order; for example, the five or six most frequent types of help desk calls from interviewers using computer-assisted interviewing.
Performance measurement analysis	A technique used in quality control to determine whether assurance procedures have worked. For example, analysis of routine measures of interviewer or coder performance.
Personally Identifiable Information (PII)	Information that can be used to identify a respondent that minimally includes name, address, telephone number and identification number (such as social security number or driver’s license number), but may include other information including biometric data.
Pilot study	A quantitative miniature version of the survey data collection process that involves all procedures and materials that will be used during data collection. A pilot study is also known as a “dress rehearsal” before the actual data collection begins.
Pledge of confidentiality	An agreement (typically in written or electronic form) to maintain the confidentiality of survey data that is signed by persons who have any form of access to confidential information.
Portable file	A file that is coded in a non-proprietary format such as XML or ASCII and thus can be used by a variety of software and hardware platforms.
Post-survey adjustments	Adjustments to reduce the impact of error on estimates.
Poststratification	A statistical adjustment that assures that sample estimates of totals or percentages (e.g. the estimate of the percentage of men in living in Mexico based on the sample) equal population totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on Census data). The adjustment cells for poststratification are formed in a similar way as in sample selection, but variables can be used that were not on the original frame at the time of selection.

Poststratification adjustment	A statistical adjustment that assures that sample estimates of totals or percentages (e.g., the estimate of the percentage of men in living in Mexico based on the sample) equal population totals or percentages (e.g., the estimate of the percentage of men living in Mexico based on Census data). The adjustment cells for poststratification are formed in a similar way as in sample selection, but variables can be used that were not on the original frame at the time of selection.
Precision	A measure of how close an estimator is expected to be to the true value of a parameter, which is usually expressed in terms of imprecision and related to the variance of the estimator. Less is reflected by a larger variance.
Precoding	When designing the questionnaire and survey instrument, determine coding conventions and formats of survey (especially the questions) based on existing coding frames or prior knowledge of the population.
Prescribed behaviors	Interviewer behaviors that must be carried out exactly as specified.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins.
Primacy	Context effects in which the placement of the at the beginning of a list of options increases the likelihood that it will be selected by the respondent.
Primary Sampling Unit (PSU)	A cluster of elements sampled at the first stage of selection.
Probability proportional to size (PPS)	A sampling method that assures that sample estimates of totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on the sample) equal population totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on Census data). The adjustment cells for postratification are formed in a similar way as strata in sample selection, but variables can be used that were not on the original frame at the time of selection.
Probability sampling	A sampling method where each element on the sampling frame has a known, non-zero chance of selection.
Process analysis	The use of tools such as flowcharts to analyze processes, e.g., respondent tracking, computerized instrument programming and testing, data entry, etc. The aim is and to identify indicators or measures of the of products. analysis also is used to identify improvements that can be made to processes.
Process improvement plan	A plan for improving a process, as a result of process analysis. A improvement plan may result from development of a management plan, or as a result of assurance or control.
Process indicator	An indicator that refers to aspects of data collection (e.g., HPI, refusal rates, etc.).
Processing error	Survey error (variance and that arise during the steps between collecting information from the respondent and having the value used in estimation. errors include all post-collection operations, as well as the printing of questionnaires. Most processing errors occur in data for individual units, although errors can also be introduced in the implementation of systems and estimates. In survey data, processing errors may include errors of transcription, errors of errors of data entry, errors in the assignment of weights, errors in disclosure avoidance, and errors of arithmetic in tabulation.
Progress indicator	An indicator that refers to aspects of reaching the goal (e.g., number of complete interviews).

Proxy interview	An interview with someone (e.g., parent, spouse) other than the person about whom information is being sought. There should be a set of rules specific to each survey that define who can serve as a proxy respondent.
Public use data files	An anonymized data file, stripped of respondent identifiers that is distributed for the public to analyze.
Public-use data file	An anonymized data file, stripped of respondent identifiers that is distributed for the public to analyze.
Quality	The degree to which product characteristics conform to requirements as agreed upon by producers and clients.
Quality assurance	A planned system of procedures, performance checks, quality audits, and corrective actions to ensure that the products produced throughout the lifecycle are of the highest achievable quality. assurance planning involves identification of key indicators of quality used in quality assurance.
Quality audit	The process of the systematic examination of the quality system of an organization by an internal or external quality auditor or team. It assesses whether the management plan has clearly outlined assurance, control, corrective actions to be taken, etc., and whether they have been effectively carried out.
Quality checklist	A checklist for quality identifies all the steps, procedures, and controls specified to ensure required procedures have been followed and their goals met. An example of a Translation Checklist is the ESS Round 7 Translation Quality Checklist (European Social Survey, 2014c).
Quality control	A planned system of process monitoring, verification, and analysis of indicators of quality, and updates to assurance procedures, to ensure that quality assurance works.
Quality management plan	A document that describes the quality system an organization will use, including assurance and control techniques and procedures, and requirements for documenting the results of those procedures, corrective actions taken, and process improvements made.
Quality profile	A comprehensive report prepared by producers of survey data that provides information data users need to assess the quality of the data.
Question-by-question objectives	Text associated with some questions in interviewer-administered surveys that provides information on the objectives of the questions.
Questionnaire adaptation	The deliberate technical or substantive modification of some feature of a question, response scales, or other part of a questionnaire to better fit a new socio-cultural context or particular population (e.g., updating language: "radio" for "wireless", adapting an adult questionnaire for children: "tummy" for "stomach"; or for cultural needs: walk several blocks versus walk 100 yards).
Quota sampling	A non-probability sampling method that sets specific sample size quotas or target sample sizes for subclasses of the population. The sample quotas are generally based on simple demographic characteristics (e.g., quotas for gender, age groups, and geographic region subclasses).
Random route (Random walk)	For each randomly-chosen sampling points (e.g., urban units, small cities, or voting districts), interviewers are assigned with a starting location and provided with instructions on the random walking rules - e.g., which direction to start, on which side of the streets to walk and which crossroads to take. Households are selected by interviewers following the instructions. The routes end when the predefined number of respondents (or households) is achieved (Bauer, 2016). Since the probability of the selected household is unknown, this method is categorized as non-probability sampling methods (Bauer, 2016)..

Random-digit-dialing (RDD)	A method of selecting telephone numbers in which the target population consists of all possible telephone numbers, and all telephone numbers have an equal probability of selection.
Ranking format	A response format where respondents express their preferences by ordering persons, brands, etc. from top to bottom, i.e., generating a rank order of a list of items or entities. Example: Listed below are possible disadvantages related to smoking cigarettes. Please enter the number 1, 2, 3, or 4 alongside each possible disadvantage to indicate your rank ordering of these. 1 stands for the greatest disadvantage, 4 for the least disadvantage. Harmful effects on other people's health. Stale smoke smell in clothes and furnishings. Expense of buying cigarettes Harmful effects on smoker's health
Rating format	A response format requiring the respondent to select one position on an ordered scale of response scales. Example: To what extent do you agree or disagree with the following statement? *It is a good idea to ban smoking in public places. *Strongly agree *Somewhat agree *Neither agree nor disagree *Somewhat disagree *Strongly disagree
Recency	Context effects in which the placement of the at the end of a list of options increases the likelihood that it will be selected by the respondent.
Recontact	To have someone other than the interviewer (often a supervisor) attempt to speak with the sample member after a screener or interview is conducted, in order to verify that it was completed according to the specified protocol.
Refusal rate	The proportion of all units of all potentially eligible sampling units in which a respondent sampling unit refuses to do an interview or breaks off interviews of all potentially eligible sampling units.
Reinterview	The process or action of interviewing the same respondent twice to assess reliability (simple response variance.)
Reliability	The consistency of a measurement, or the degree to which an instrument measures the same way each time it is used under the same condition with the same subjects.
Reluctance aversion techniques	Techniques that can reduce reluctance to participate in potential respondents, thereby increasing the overall response rate.
Repeated panel	A series of fixed panel surveys that may or may not overlap in time. Generally, each panel is designed to represent the same population definition applied at a different point in time.
Replicated question	A question which is repeated (replicated) at a later stage in a study or in a different study. Replication assumes identical question wording. Questions which were used in one study, then translated and used in another are also frequently spoken of as having been "replicated."
Replicates	Systematic probability subsamples of the full sample.
Residency rule	A rule to help interviewers determine which persons to include in the household listing, based on what the reports.
Response distribution	A description of the values and frequencies associated with a particular question.
Response distributions	A description of the values and frequencies associated with a particular question.
Response latency	A method of examining potential problems in responding to particular items, measured by the time between the interviewer asking a question and the response.
Response options	The category, wording, and order of options given with the survey question.
Response rate	The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.

Response scales	The category, wording, and order of options given with the survey question. See Questionnaire Design for more information.
Response styles	Consistent and stable tendencies in response behavior which are not explainable by question content or presentation. These are considered to be a source of biased reporting. For example, extreme response style is the tendency to select the two extreme endpoints of a scale; midpoint response style refers the consistent selection of middle or neutral category of the scale; acquiescent response style is the tendency to agree with or to select the positive responses.
Responsive designs	Responsive design was developed by Groves and Heeringa (2006). Usually the following steps are included: first, researchers pre-identify a set of design features that are of interest (e.g., tradeoff between cost and error); second, researchers identify and monitor these indicators; third, researchers provide intervention or alter the design features based on the pre-identified design decisions rules. Most recently, with the development of technology, real-time responsive design can be achieved in a single survey, such as prompting respondents who show signs of in web surveys.
Restricted tendering	A bidding process in which only bidders prequalified through a screening process may participate in bidding, in which they are evaluated and then chosen on the basis of cost and technical merit.
Restricted-use data file	A file that includes information that can be related to specific individuals and is confidential and/or protected by law. Restricted-use data files are not required to include variables that have undergone coarsening disclosure risk edits. These files are available to researchers under controlled conditions.
Reviewer	Person who participates in the review of translations in order to produce a final version (see Appendix A of Translation).
Rotating panel design	A study where elements are repeatedly measured a set number of times, then replaced by new randomly chosen elements. Typically, the newly-chosen elements are also measured repeatedly for the appropriate number of times.
Sample design	Information on the target and final sample sizes, strata definitions and the sample selection methodology.
Sample element	A selected sampling unit of the target population that may be eligible or ineligible.
Sample management system	A computerized and/or paper-based system used to assign and monitor sample units and record documentation for sample records (e.g., time and outcome of each contact attempt).
Sample person	A person selected from a sampling frame to participate in a particular survey.
Sampling bias	The systematic difference between the expected value (over all conceptual trials) of an unweighted sample estimate and the target population value because some elements on the frame have a higher chance of selection than other elements.
Sampling error	Survey error (variance and due to observing a sample of the population rather than the entire population).
Sampling error computational units (SECUs)	PSUs in 'one PSU per sampling designs that are grouped in pairs, after data collection, for purposes of estimating approximate variances.

Sampling frame	A list or group of materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.
Sampling units	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term sampling units ( refers to the sampling units chosen in the first stage of selection. The term sampling units ( refers to sampling units within the PSUs that are chosen in the second stage of selection.
Sampling variance	A measure of how much a statistic varies around its mean (over all conceptual trials) as a result of the sample design only. This measure does not account for other sources of variable error such as coverage and nonresponse.
Satisficing	To answer survey questions optimally, four stages of cognitive processing are required: (1) interpret the questions comprehensively, (2) retrieve information from memory, (3) form a judgment, and (4) map the judgment to the appropriate response category. However, to lower cognitive burden, instead of seeking to optimize, respondents may skip some steps when the answer survey questions. This behavior is called satisficing.
Satisficing behaviors	Decision-making strategies that entail searching through the available alternatives until an acceptability threshold is met.
Secondary Sampling Unit (SSU)	A cluster of elements sampled at the second stage of selection.
Sequential mixed mode	A mixed mode design in which additional are offered as part of a follow-up program.
Shared Language harmonization	Shared language harmonization can be understood as the procedures and result of trying to harmonize as much as possible different regional varieties of a “shared” language across countries, e.g. in terms of vocabulary and/or structure. An example would be to harmonize translations into Italian between questionnaires used in Italy and in Switzerland.
Silent monitoring	Monitoring without the awareness of the interviewer.
Simple random sampling (SRS)	A procedure where a sample of size $n$ is drawn from a population of size $N$ in such a way that every possible sample of size $n$ has the same probability of being selected.
Social desirability bias	A tendency for respondents to overreport desirable attributes or attitudes and underreport undesirable attributes or attitudes.
Socio-demographic question	A question typically asking about respondent characteristics such as age, marital status, income, employment status, and education. Example: What year and month were you born?
Soft consistency check	A signal warning that there is an inconsistency between the current response and a previous response. The soft consistency check should provide guidance on resolving the inconsistency, but the interviewer or respondent may continue the survey without resolving it.
Source document	The original document from which other (target) documents are translated or adapted as necessary.
Source instrument	The original instrument from which other (target) instruments are translated or adapted as necessary.

Source language	The language in which a questionnaire is available from which a translation is made. This is usually but not always the language in which the questionnaire was designed.
Source questionnaire	The questionnaire taken as the text for translation. The source questionnaire would normally not be intended to be fielded as such but would require local even if fielded in the language: for instance, in the ESS, the source questionnaire is English, but the questionnaires fielded in Ireland and UK differ slightly from the source questionnaire.
Source variables	Original variables chosen as part of the harmonization process.
Split panel	A design that contains a blend of cross-sectional and panel samples at each new wave of data collection.
Split translation	Each translator translates only a part of the total material to be translated in preparation for a review meeting, in contrast to translating the entire text (see translation).
Stand-alone experiment	An experiment conducted as an independent research project.
Standardized interviewing technique	An interviewing technique in which interviewers are trained to read every question exactly as worded, abstain from interpreting questions or responses, and do not offer much clarification.
Statistical process control chart	A statistical chart that compares expected process performance (e.g., number of hours worked by interviewers in a week) against actual performance. For example, interviewers who perform outside upper and lower boundaries on this measure are flagged; if greater variation from expected performance for some interviewers in a certain location can be explained (e.g., a hurricane or a snow storm causing lower than expected hours worked), the process is in control; if not, corrective actions are taken.
Straightlining	Clicking the same answer for each item in a multi-numeric list.
Strata (stratum)	Mutually exclusive, homogenous groupings of population sample elements or of elements that comprise all of the elements on the frame. The groupings are formed prior to selection of the sample.
Stratification	A sampling procedure that divides the sampling frame into mutually exclusive and exhaustive groups (or and places each element on the frame into one of the groups. Independent selections are then made from each one by one, to ensure representation of each subgroup on the frame in the sample.
Substitution	A technique where each nonresponding sample element from the initial sample is replaced by another element of the population, typically not an element selected in the initial sample. increases the nonreponse rate and most likely the nonresponse bias.
Survey lifecycle	The lifecycle of a survey research study, from design to data dissemination.
Survey population	The actual population from which the survey data are collected, given the restrictions from data collection operations.
Survey weight	A statistical adjustment created to compensate for complex survey designs with features including, but not limited to, unequal likelihoods of selection, differences in rates across key subgroups, and deviations from distributions on critical variables found in the population from external sources, such as a national Census.
Systematic sampling	A procedure that selects of every kth element on the sampling frame after a random start.
Tailoring	The practice of adapting interviewer behavior to the respondent's expressed concerns and other cues, in order to provide feedback to the respondent that addresses his or her perceived reasons for not wanting to participate.

Target language	The language a questionnaire is translated into.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.
Target variables	Variables created during the harmonization process.
Task	An activity or group of related activities that is part of a survey process, likely defined within a structured plan, and attempted within a specified period of time.
Taylor Series variance estimation	A commonly used tool in statistics for handling the variance estimation of statistics that are not simple additions of sample values, such as odds ratios. Taylor series handles this by converting a ratio into an approximation that is a function of the sums of the values.
Team translation	Team approaches to survey translation and translation assessment bring together a group of people with different talents and functions in the team so as to ensure the mix of skills and discipline expertise needed to produce an optimal translation version in the survey context. Each stage of the team translation process builds on the foregoing steps and uses the documentation required for the previous step to inform the next. In addition, each phase of translation engages the appropriate personnel for that particular activity and provides them with relevant tools for the work at hand.
Tender	A formal offer specifying activities to be completed within prescribed time and budget.
Timestamps	Timestamps are time and date data recorded with survey data, indicated dates and times of responses, at the question level and questionnaire section level. They also appear in trails, recording times questions are asked, responses recorded, and so on.
Top coding	A type of coding in which values that exceed the predetermined maximum value are reassigned to that maximal value or are recoded as missing data.
Total Survey Error (TSE)	Total survey error provides a conceptual framework for evaluating survey quality. It defines quality as the precise estimation and reduction of the square error (MSE) of statistics of interest.
Tracking	The process of attempting to locate a sample element that changed contact information (e.g. address, telephone number, email address) since the last time the element's contact information was collected.
Transformation algorithms	Changing the values of a variable by using some mathematical operation.
Translatability assessment	Translatability assessment is a recently developed process which aims to identify potential problems in translation and in the initial instrument development stage in the language (Conway, Acquadro, & Patrick, 2014; Sperber et al., 2014). It evaluates to which extent, a question of interest can be meaningfully translated.
Translator	The person who translates text from one language to another (e.g., French to Russian). In survey research, translators might be asked to fulfill other such as reviewing and copyediting.
Trusted digital repository	A repository whose mission is to provide reliable, long-term access to managed digital resources to its designated community, both now and in the future.
Undocumented code number	A code that is not authorized for a particular question. For instance, if a question that records the sex of the respondent has documented codes of "1" for female and "2" for male and "9" for "missing data," a code of "3" would be an "undocumented code."

Unique Identification Number	A unique number that identifies an element (e.g. serial number). That number sticks to the element through the whole survey lifecycle and is published with the public dataset. It does not contain any information about the respondents or their addresses.
Unit nonresponse	An eligible sampling unit that has little or no information because the unit did not participate in the survey.
units (Sampling units)	Elements or clusters of elements considered for selection in some stage of sampling. For a sample with only one stage of selection, the units are the same as the elements. In multi-stage samples (e.g., enumeration areas, then households within selected enumeration areas, and finally adults within selected households), different sampling units exist, while only the last is an element. The term sampling units ( refers to the sampling units chosen in the first stage of selection. The term sampling units ( refers to sampling units within the PSUs that are chosen in the second stage of selection.
Universe statement	A description of the subgroup of respondents to which the survey item applies (e.g., “Female, ≥ 45, Now Working”).
Unwritten language	An unwritten language is one which does not have a standard written form used by the native speakers of the language.
Usability testing	Evaluation of a computer-assisted survey instrument to assess the effect of design on interviewer or respondent performance. Methods of evaluation include review by usability experts and observation of users working with the computer and survey instrument.
Validity	The extent to which a variable measures what it intends to measure.
Variance	A measure of how much a statistic varies around its mean over all conceptual trials.
Vignettes	Brief stories/scenarios describing hypothetical situations or persons and their behaviors to which respondents are asked to react in order to allow the researcher to explore contextual influences on respondent’s response formation processes.
Vocal characteristics	“Analysis of vocal characteristics, also called paralinguistic data (Draisma & Dijkstra, 2004), like codes, examines audio recordings of interviews to identify notable traits of the interviewer’s voice itself, rather than behaviors during the interview. These vocal properties include pitch (higher or lower sounding voices), intonation (rising or falling pitch), speech rate, and loudness” (Olson & Parkhurst, 2013).
Weighting	A post-survey adjustment that may account for differential coverage, sampling, and/or processes.
Within-country language harmonization	The process of harmonizing language versions within one multilingual country, such as harmonizing the Ukrainian and Russian versions within Ukraine.
Word list	Word lists can of course serve various purposes: When regional varieties of a language are to be accommodated, a word list can be created of the words that are required for specific varieties of a language. They can also be incorporated into computer applications of an instrument. A word list can be a useful resource for interviewers. They cannot, however, address challenges faced when regional varieties differ in more radical and structural ways from one another. A word list can also serve similar functions as a glossary.
Working group	Experts working together to oversee the implementation of a particular aspect of the survey lifecycle (e.g., sampling, questionnaire design, training, control, etc.)

XML (eXtensible Markup Language)	XML (Extensible Markup Language) is a flexible way to create common information formats and share both the format and the data on the World Wide Web, intranets, and elsewhere. XML documents are made up of storage units called entities, which contain either parsed or unparsed data. Parsed data are made up of characters, some of which form character data, and some of which form markup. Markup encodes a description of the document's storage layout and logical structure. XML provides a mechanism to impose constraints on the storage layout and logical structure.
----------------------------------	---

From:  
<https://wiki.capstan.be/> - **cApStAn**

Permanent link:  
[https://wiki.capstan.be/doku.php?id=generalreference:survey\\_glossary&rev=1501668815](https://wiki.capstan.be/doku.php?id=generalreference:survey_glossary&rev=1501668815)

Last update: **2017/08/02 13:13**

