Data Analysis: Gathering Data - Questionnaires/Surveys

WHAT IT IS

Return to Table of Contents

Questionnaires and surveys are structured ways of collecting data from a population or a sample of a population. If questionnaires or surveys are mailed out or distributed by other means, they usually are self-administered by the respondent. If questionnaires or surveys are administered in person or over the phone, they often resemble a structured interview.

Survey questions may be closed-ended or open-ended. Questions with closed-ended (fixed) response categories are more easily quantified. Categorizing the responses of open-ended questions may require use of content analysis. (See the module on Content Analysis for more information.)

WHEN TO USE IT

Questionnaires and surveys usually gather descriptive and normative data, though they can gather cause-and-effect data. Generally, they are useful for validating a grounded theory and are not good for exploring new ideas. Other methods should be used when the identification and location of knowledgeable respondents is difficult or if literacy or language barriers exist. The objective of surveys is to collect a limited amount of focused information from a sample of individuals. Up to a point, the larger the sample, the more valid the data. But, increased validity must be weighed against the cost of a larger sample.

HOW TO PREPARE IT

- Define the objectives, uses, and target population of the questionnaire/survey.

- Decide upon the mix of open-ended versus closed-ended questions, noting that open-ended questions can:
  - broaden the scope of possible responses
  - assist in formulating other more specific questions
  - tend to collect qualitative rather than quantitative information
  - sometimes be quantified using content analysis (see the Content Analysis module)
  - involve gathering fairly in-depth data from a few respondents
  - provide a context for deeper understanding of responses
  - often be used during management and performance reviews

  And that closed-ended questions can:
  - limit the scope of responses
  - be binary (i.e. use “yes/no” questions)
  - be scaled (e.g. "strongly agree," "agree," "disagree," or "strongly disagree")
  - yield quantifiable data more easily than open-ended questions
  - involve gathering less in-depth data
  - be more easily administered to large numbers of respondents
  - assist in reviews of internal controls
  - provide strong confirmation of findings using other methods

  (See the module on Interviews for more information.)

- Decide upon a format and scaling strategy for closed-ended questions.
  - — **Yes/no questions** are common in internal control questionnaires. They must be very specific. Such questions yield limited data, and responses can be misleading because of the lack of elaboration.
  - — **Multiple choice questions** are efficient and can provide useful data. They can be hard to design since all important responses must be included, and response categories are easily misinterpreted.
  - — **Ranking questions** rank response options according to their importance, size, or cost. A common ranking scale is the 1-10 scale. While a larger number of possible ranks can yield richer data, both the reliability and usefulness of the data can suffer if too many ranks are used. For these and other reasons, ranking questions are often considered unreliable by many statisticians. The methods used to analyze ranked data can also be complicated.
  - — **Intensity scaled questions** are similar to multiple choice questions. However, they attempt to gather evaluative information from respondents. For example, a response scale for a question rating service quality might include such categories as "excellent," "very good," "good," "fair," and "poor."

  When scaling questions, avoid biasing response categories in favor of one type of response. For example, having three categories indicating a favorable perception and only one indicating an unfavorable perception would bias the question.

- Design the questionnaire/survey instrument. Some important design techniques are:
  - — Include an introductory statement with the survey which briefly summarizes the survey's purpose, motivates respondents to participate, and addresses confidentiality.
  - — Provide clear, concise directions for completing the survey. Directions should address how answers should be indicated, how to deal with items which are not applicable, and what to do with the completed survey.
  - — Include a demographic section which gathers relevant data on respondents. Avoid unnecessary questions.

- Refine the questions, being sure that each question is:
  - — relevant to the audit/evaluation issue under study
  - — targeted to respondents most likely to give meaningful answers
  - — answerable without placing undue burden on the respondent
  - — has a good probability of contributing useful information to the audit/evaluation report

- Assess the appropriateness of the language used in each question. To ensure that questions are clear and can be properly answered, avoid:

- speaking over the respondent's head
- speaking down to the respondent
- using double negatives, abbreviations, jargon, or acronyms
- oversimplifying to the point of being patronizing or demeaning
- using lengthy questions
- asking multiple questions in a single question
- framing questions in vague or indirect terms
- using extreme or inflammatory language which may mislead or insult respondents

- Scale responses. Scales should usually have five or fewer options.

- Review questions to eliminate biased or unfair wording. Common forms of bias which may appear in questions are:
  - implied answer bias in which questions imply the right answer, e.g., "should management directives be followed or not?"
  - unequal choice bias in which possible responses are biased e.g., "who is to blame, staff or careless managers?"
  - loaded terms which engender emotional reactions, e.g., "deadbeat," "incompetent," "wrong," or "wasteful"

- Consider the order of the questions asked. Initial questions should be simple and routine. More complex or sensitive questions should come later. Prioritize questions to allow for shortening the instrument later, if needed.

- Determine whether questions meet the survey objectives (i.e. establish content validity).
  - One method for doing this is calculating the Index of Congruence. This method uses 3-5 independent judges to rate how well a given question meets each stated objective. The higher the congruence, the better the question.
  - It may not be necessary in all cases to compute an Index of Congruence for the survey. However, it is highly recommended that several individuals other than the survey designer review questions for clarity and correspondence to objectives.

- If possible, pilot test the survey with 3-5 persons who are similar to the members of the population who will ultimately be surveyed. Revise questions based on the Index of Congruence and/or pilot testing.

- Devise a sampling plan. (See the Sampling module for more information.) First, decide whether to use random or purposeful sampling.
  - **Random sampling** gives each member of the population an equal probability of being selected as a member of the sample.
  - **Purposeful sampling** does not give each member of the population an equal probability of being selected as a member of the sample.

Random sampling undeniably produces more statistically valid results in that they can be generalized to describe or predict the characteristics of the whole population. Purposeful sampling is more useful if a survey is for exploratory purposes. (The various types of purposeful sampling are listed in the appendix to the Sampling module.)

- Determine an appropriate sample size. In general, the higher the sample size, the lower the margin of error in the survey. (See the module on Sampling for more information on calculating a sample size.)

- Administer the survey. Making firm conclusions or predictions from a survey or questionnaire with a response rate lower than 70 percent is risky. Yet, a tradeoff exists between overall cost and response rate. A mailed survey with no follow up is inexpensive but may have a low response rate.

- Enter response data into an automated data file. Among the software often used for this are SAS, SPSS, dBASE, Minitab, and Quattro Pro.

- Analyze the results. The type of analysis conducted depends on whether the results of the survey are expected to be descriptive, normative, or cause-and-effect.
  - **Descriptive** results are obtained using such techniques as frequency distributions, cross-tabulations, measures of central tendency, and measures of dispersion. Descriptive statistical techniques do not ascertain cause-and-effect relationships. (See the modules on Describing Data and Displaying Data for more information.)
  - **Normative** results are obtained using such techniques as t-tests, z-tests, F-tests, Chi-square, and analysis of variance. Normative statistics do not ascertain cause-and-effect relationships. Rather, they compare two or more populations and may be used to determine both which populations differ and how strong is the difference. Such statistics are an important part of benchmarking. (See the module on Normative Statistics for more information.)
  - **Cause-and-effect** results are obtained using such techniques as correlation, regression analysis, and Chi-square. Inferential statistics measure relationships among different variables in the data and may also be used to infer cause and effect. Inferential statistical methods can be highly technical and difficult to use properly without advanced training. (See the module on Cause-and-Effect Statistics for more information.)

- Determine reliability. Reliability relates to whether the survey yields similar results upon repeated trials. It actively considers if measures used are consistent and free of random error or bias. Measuring reliability is especially important for surveys which are repeated. Among the issues

that can compromise reliability are:

- mood or alertness fluctuations on the part of respondents
- variations in conditions of survey/questionnaire administration
- differences and/or errors in scoring or interpreting results
- random errors

Different types of reliability and methods for establishing them include:

- **Test-retest reliability** refers to administering the same instrument to the same sample within a time period during which the issue is not likely to change, often 30-90 days. This can be expensive and can bias results due to respondent familiarity with the questions.
- **Alternate form reliability** refers to giving a highly similar instrument to the same sample and checking correlations between responses to similar items. Content validity is vital for this to work.
- **Split half reliability** refers to getting good correlations between responses to half of the questions on an objective with the responses to the other half of the questions on the same objective. Questions constituting the two halves must be as alike as possible.
- **Interrater reliability** refers to the reliability of the different persons administering or interpreting the instrument and/or its results and establishes the administrator as a source of error.
- **Internal consistency reliability** refers to having variation around responses to a given item under an objective, similar to variation around a composite score for a given objective, and is tested via the Cronbach Alpha Test.

Reliability is often downplayed during audits/evaluations due to the cost and difficulty of measuring it if a survey or questionnaire is not expected to be used again.

- Determine validity. If a survey or questionnaire has questionable validity, any conclusions drawn from it should be considered skeptically. The general types of validity are:
  - **Internal validity** examines how well the survey or questionnaire measures the sample, i.e. do those who responded represent the sample?
  - **External validity** examines how well the data obtained can be generalized to other samples, i.e. does the sample represent the population?

  Specific types of validity include:
  - **Content validity** refers to how well the survey or questionnaire reflects its objectives and is often established by the Index of Congruence discussed above. This is an important step in pilot testing an instrument.

- **Construct validity** refers to how well questions related to the same objective correlate with each other and is often established through Factor Analysis. Other possible means include opinions of judges, known correlations, criterion group studies, and appeals to logic.
- **Criterion (concurrent) validity** refers to how well results obtained from one data gathering instrument are supported by other surveys or questionnaires and is often established by correlating the results of the different data gathering instruments. This tells the extent to which a behavior surveyed allows one to distinguish between observations on another behavior.
- **Predictive validity** refers to how well the survey or questionnaire actually predicts future behavior and is established by both sample size and observation of behavior.

In the best of all worlds, all types of validity would be addressed, but this is often impractical. Auditors/evaluators will have to be the judge. However, one should report on all types of validity, if at all possible. Among the threats to validity are:

- weak links between attitude, behavior, and perception
- rivalry and/or response bias due to a desire to (dis)please or apprehension over being surveyed or evaluated
- lack of understanding of the question or inability to report on personal behavior
- lack of objectivity or consistency in administration
- too few questions or observations used to assess the behavior
- physical limitations of respondents
- inconsistency between pre- and post-tests and/or influence of pre-test on post-test
- differences in selection processes for different samples used to test the same construct
- the Hawthorne effect
- mortality of the sample
- regression to the mean over time (reduces variability)

- To enhance reliability and validity, consider the following:
  - To make possible a high degree of variability, seek high response rates, preferably around 70 percent of the sample.
  - Lengthening the data gathering instrument or expanding the response scale (up to a point) can provide a better sample of the construct.
  - Pilot test and administer the data-gathering instrument, if possible.
  - Build tests of reliability and validity into your sampling plan.
  - Use random samples wherever possible. Try to avoid purposeful sampling since they usually yield only possible hypotheses, not data from which statistically significant inferences can be made. (See the module on Sampling for more information.)
  - In general, the number of observations (participants) minus the

number of variables (questions on your data gathering instrument) must be 30 or greater. Try to make this number at least 60 to provide greater precision.

- Summarize the results of the data analysis in writing.
- Communicate the results to the appropriate parties, and add the survey and/or question list and write-up to the project working papers.
- If questions will be used again, revise those that did not give useful data.

ADVANTAGES

Surveys and questionnaires can:

- be versatile
- gather a great deal of data
- complement other evaluation techniques
- eliminate the bias that can occur during interviews
- be widely distributed
- be easy to quantify
- facilitate creation of graphs and charts
- expand line staff input

DISADVANTAGES

Surveys and questionnaires can:

- have long turnaround times
- prove difficult and time-consuming to develop and pilot test
- be challenging when good closed-end questions are needed or when open-ended questions are analyzed
- suffer from low response rates which impair ability to generalize results
- prove costly when assessing reliability and validity
- generate different conclusions depending on who interprets responses and results