TRN00-017

Measuring Public Awareness: A Quick Reference Guide

Prepared for the Public Affairs Program—2000 CSEPP National Conference

Innovative Emergency Management
INCORPORATED

Comments and suggestions concerning this seminar should be directed to IEM at the following address:

Dr. Barbara Cochran
Innovative Emergency Management
8555 United Plaza Boulevard, Suite 100
Baton Rouge, LA 70809
Voice (225) 952-8191
Fax (225) 952-8122
barbara.cochran@ieminc.com

TABLE OF CONTENTS

INTRODUCTION **1**

PURPOSE 1

PLANNING A SURVEY **3**

SURVEY RESEARCH DESIGN 3
Qualitative versus Quantitative Techniques *4*
Measurement *5*
Reliability and Validity *11*
Sampling: Who You Should Survey *14*

CONSTRUCTING A SURVEY INSTRUMENT **18**

ADMINISTERING THE SURVEY 18
Mail Surveys *19*
Group Administration of Surveys *19*
Telephone Interviews *19*
Personal Interviews *20*
ASKING THE RIGHT QUESTIONS 20
Format *20*
Physical layout *22*
Content *22*
Question Order *23*
Instructions *24*
RESPONSE EFFECTS 25
Task Effects *25*
Interviewer Effects *25*
Respondent Effects *26*

INTERPRETING SURVEY RESULTS **27**

INTERPRETING SURVEY STATISTICS 27
Statistical Significance *29*
READING CHARTS AND GRAPHS 29

PRESENTING FINDINGS **31**

WHAT IS A PICTURE WORTH? 31

CONCLUSION ..... ERROR! BOOKMARK NOT DEFINED.

REFERENCES ..... ERROR! BOOKMARK NOT DEFINED.

GLOSSARY ..... ERROR! BOOKMARK NOT DEFINED.

INTRODUCTION

Survey Research is a method of obtaining information directly from people by asking questions. This distinguishes survey research from other types of scientific investigation in which observation is the primary method of collecting information. To the casual observer, survey research seems relatively simple, but the very nature of the survey makes the information obtained subject to many biases. Obtaining valid and useful information using the survey method requires training, as well as practical experience.

KEY POINT

- **While survey research may seem relatively simple, skill and planning are required to ensure that respondents provide the information you want.**

Purpose

This guide provides a basic overview of survey techniques and methods for measuring public awareness in CSEPP communities. While this guide is not a comprehensive course in research design and analysis, it does provide information on important design issues in survey research. Information in this document should prepare you to deal with survey contractors and to read and interpret survey results.

The following topics are addressed in this guide:

√ **Planning a Survey**
- Quantitative versus Qualitative Techniques
- Measurement
- Sampling

√ **Constructing a Survey Instrument**
- Administering the Survey
- Asking the Right Question
- Response Effects

√ **Interpreting Survey Results**
- Interpreting Statistics
- Reading Charts and Graphs

√ **Presenting Findings**

In addition, this guide contains a glossary that defines common terms used in survey research and statistical documents such as the *Statistical Abstracts of the United States*. A list of books and papers for further readings is also provided.

PLANNING A SURVEY

Research must be systematic, not haphazard, in order to have confidence in results. Your research design is your overall research plan. It spells out the details of what you are going to do and how you are going to do it. Decisions about exactly how research is going to be carried out must be made before you actually start surveying.

It is important to make your plan explicit. Leave nothing to be misinterpreted by those assisting you in conducting the research. Do not trust your memory, as details are easily forgotten. Vague, under-specified designs yield useless data.

Survey Research Design

There are two basic types of surveys: questionnaires and interviews. Both questionnaires and interviews fall under the category of self-report measures, but questionnaires are self-administered, while interviews are administered by trained interviewers. The quality of information gathered through surveys depends on decisions made during the design phase of the research project. This section discusses basic design issues that must be decided early in the survey research project.

The survey design must reflect the basic purpose of the research. The goal in survey research is to record how often some event occurs[1] and under what conditions. Determine the question you need to answer. The question must be clear and precise. It must be a question that can be answered by objective evidence; in other words, by something you can measure. Designing a survey involves the following:

- deciding whether to use qualitative or quantitative techniques
- determining a measurement method, sampling strategy, and mode of administration
- constructing the questionnaire or interview script

[1] An event may be a particular behavior or a thought, attitude, or opinion.

KEY POINT

- **Surveys are self-report measures. To avoid problems inherent to self-report measures, surveys must be carefully designed.**

KEY POINT

- **While qualitative techniques provide a rich source of information, analyzing the data may be difficult and time consuming.**

- developing a method of training interviewers
- determining analyses necessary to answer the original research question

Qualitative versus Quantitative Techniques

Qualitative Techniques

The goal of qualitative research is to obtain information-rich data. Qualitative techniques include focus groups, field observation, in-depth interviews, and case studies. Qualitative methods are often used as a preliminary step to help with planning research. They can be used for pilot work if the research question needs refinement. Inductive analyses are most often used with qualitative methods. Data is collected first and then grouped into meaningful categories through content analyses, etc.

Inexperienced researchers sometimes choose qualitative research techniques because they seem less complicated. A common misconception is that if you want to know what people think about a particular topic, all you have to do is ask. In this, however, caution should be exercised. Qualitative data can be much more difficult to analyze than quantitative.

The advantages of qualitative research include the fact that it can increase the depth of understanding and that it is flexible and allows for the pursuit of new areas of inquiry.
The disadvantages are that cost can be prohibitive. This can lead to small sample size that affects the reliability and general soundness of the study. There is also the problem that if qualitative research is not extremely well planned it can lead to nothing of value.

Quantitative Techniques

Quantitative techniques require that variables of interest be measured. Certain critical decisions about how the variables will be measured are made during the design of the research. Quantitative research is deductive in nature. The researcher identifies research questions as well as categories of answers during the design phase of the experiment. Designing a good quantitative survey requires a great deal of time and work. It takes a high degree of expertise to ensure that you obtain the information you want.

The main advantages of quantitative research are that it allows greater precision in reporting results and it permits the use of statistical analyses in order to understand results better. Quantitative techniques also allow the use of inferential statistical procedures so that generalizations about the entire target population may be made from a sample survey.

The main disadvantage of using quantitative techniques is that some information may be lost as a result of predefined response categories. Quantitative techniques require careful expert planning to preserve information necessary to answer the research question.

KEY POINT

- **When the purpose of research is to generalize results of a sample survey to the entire population, quantitative research techniques are best.**

Measurement

A critical part of designing quantitative research is to determine exactly how variables of interest will be measured. **Measurement** is assigning numerals to objects, events, or properties of those objects or events according to certain rules. A **numeral** is a symbol that has no value until a level of measurement is determined (see levels of measurement below). **Assigning** is the designation of numerals to represent certain objects or events. **Rules** specify the way numerals are to be assigned.

What to Measure

Researchers may wish to measure many different aspects of any typical research venue. Sometimes we are interested in the proposed cause of some event. Other times we are interested in the result of some event or action. Relating proposed causes to results can give you an idea of which actions or processes are effective in accomplishing their purpose.

Some terms that may be used to describe certain types of measures are listed and defined below:

Input measures attempt to quantify variables which are thought to be *causes* for particular outcomes. These possible causes may also be called **predictor variables** in a statistical analysis.

KEY POINT

- **Measuring abstract concepts like attitudes or opinions involves carefully defining how you will obtain those measures.**

Process measures attempt to quantify processes that cause changes in outcomes.

Output measures attempt to quantify objects that are the result of some input or process.

Outcome measures attempt to quantify events or states that result from some input or process.

Variables are conditions or events that vary over time or across situations. Thus, variables can have more than one value. For example, your age is a variable. It varies over time. In addition, variables can be measured empirically. That is, they can be measured directly. Variables have a definition that is independent of the research environment. Both input and process measures quantify factors that researchers may call **independent variables**. Output and outcome measures are used to quantify what researchers commonly call the **dependent variable.**

Constructs, on the other hand, are abstract. They are formulated to serve as causal or descriptive explanations. Constructs are not directly measurable. Constructs are defined within a particular research environment. Researchers sometimes use constructs (for example, confidence in elected officials) because the research question involves something that is not a directly measurable variable.

Surveys may include items about both variables and constructs. An important note, however, is that both variables and constructs must be operationally defined when they are measured for the purposes of research. An **operational definition** specifies the way a variable or construct will be measured. For example, I could define the variable *years of residence* as the number of months a person has lived in the state divided by twelve. Alternatively, I could define the same variable as *the current year minus the year the resident moved to the state.* It is possible that the two different methods of calculation would yield different answers for a number of respondents, which could be a problem if the variable were important in interpreting the results of the survey. Therefore, it is important to formulate concise operational definitions for the variables and constructs you intend to measure.

Similarity between Measurement and Reality
One of the most important characteristics of a measure is its correspondence with reality, sometimes called isomorphism. Whether a measure can be considered isomorphic depends on the operational definition of the variable or construct being measured. Researchers must endeavor to make the measures relatively realistic.

Discrete and Continuous values: Some variables have the property of taking on a finite set of values. We refer to these as **discrete variables**. Discrete variables cannot be subdivided beyond certain basic units. For example, even though we sometimes hear statements like, "The average American family has 2.3 televisions," we know that there are not thousands of families out there who own a third of a television. When asking survey respondents about discrete variables, there is no need to specify that they must answer in specific units. People will not tell you that they drive half a car, or that there are 3.2 persons living in their house.

Unlike discrete variables, however, **continuous variables** can take on any value. Theoretically, measurements of continuous variables can be broken into an infinite number of subsections for the purpose of assigning values. Time is an example of a continuous variable. If you ask how many hours a week people spend watching television, they could respond in hours, minutes, seconds, or fractions of seconds. Obviously, this could become a problem for researchers. It is important, during the design phase of a survey, to consider categories of measurement for continuous variables.

Since constructs are abstract, they are usually treated the same as continuous variables. When including questions on surveys about continuous variables, such as time, or about constructs, such as trust in elected officials, the researcher must decide during the design of the survey how to categorize these measurements.

How to Measure

There are four different ways to measure things. These are referred to as levels of measurement. The meaning of the numerals assigned to classes of objects or events depend on the

KEY POINT

- **In order to determine the measurement scale you should use, you must first determine the type of statistical analysis required to answer your research question.**

KEY POINT

- **Nominal and ordinal measures are appropriate for some types of research; however, special statistical procedures called non-parametric statistical procedures must be used to analyze the data.**

level of measurement used. Analyses that can be performed also depend on the level of measurement.

Nominal scales are the lowest level of measurement. In nominal measurement, the numerals assigned simply classify things. In other words, the numerals are labels that stand for a particular category. They have no mathematical significance. An example of nominal measurement is zip codes. You cannot add, subtract, or take an average of a set of zip codes.

Sometimes researchers use nominal measures to classify objects or properties of objects, but it is important to remember that many statistical analyses are not appropriate with nominal measures. For example, in a survey you may be asked to respond with "1"if you are male and "2" if you are female. The resulting numerals can only be used to group respondents into categories. The numerals used to represent the categories imply no specific order or ranking. You cannot say that males are better than females because they have a higher ranking. Neither do the numerals represent any specific interval or amount of any characteristic or property. You cannot say that females are twice as good as males because their score is twice as high.

Ordinal scales of measurement convey more information than do nominal scales. Objects measured at the ordinal level are ranked along some dimension, usually from smaller to greater. Examples of ordinal scales include the rankings of football or basketball teams, beauty pageant finishing orders, and Olympic medals. You cannot add, subtract, or take averages with ordinal measures.

Ordinal measures convey information about the relative position of the properties being measured; however, they do not give information about the difference between the categories. Nor is any information conveyed about the absolute amount of the property being measured. Based on the ranking, you cannot say that the difference in ability between the first place football team and the second place team is the same as the difference between the ninth place team and the tenth place team.

Interval scales are the third level of measurement. A score on an interval scale indicates rank order just as an ordinal scale does, but in addition it incorporates the property of equal differences between scores. Adjacent scores on an interval scale are separated by equal values. An example of an interval scale is temperature. The same amount of heat is required to warm an object from 20 degrees to 30 degrees as to warm it from 40 degrees to 50 degrees. Some arithmetic operations can be carried out on scores measured on an interval scale. For example, it makes sense to talk about average temperature.

One disadvantage of an interval scale, however, is that it lacks a true zero point. That is, a score of zero does not infer an absolute lack of the property being measured. A temperature of zero does not mean there is no temperature. The lack of a true zero means that a researcher cannot make statements of a proportional nature. For example, a temperature of 60 degrees is not twice as warm as a temperature of 30 degrees.

Ratio scales are considered the highest level of measurement. Ratio scales have all the properties of the other scales; that is, the scores categorize, they infer rank order, they have equal intervals, and in addition, they have an absolute zero. A score of zero on a ratio scale infers an absolute lack of the property being measured. For example, if my bank balance is zero I have no money in the bank. The amount of money I have is measured on a ratio scale.

Because ratio scales have the property of absolute zero, proportional statements can be made. For example, it makes sense to say that a car traveling at 50 miles per hour is going twice as fast as a car traveling at 25 miles per hour. All mathematical operations are possible with scores measured on a ratio scale. Because of this, a class of statistical procedures called parametric statistics can be carried out on ratio scale scores. These procedures are important in being able to make predictions about a population based on measurements taken from a sample.

While ratio scales are common in the hard sciences, they are not as common in social science or in survey research (e.g., public opinion polls). You can ask people questions that are measured on a ratio scale (for example, “How much time per

KEY POINT

- **Most survey ratings scales are designed to measure at the interval or ratio scale.**

KEY POINT

- **Likert scales and semantic differential scales are frequently used to measure public opinion because the resulting data can be easily analyzed.**

week do you spend watching television?"), but it is much more common to ask people to rate statements of opinion, or to rank statements in order of importance. In other words, the topics survey researchers are most interested in are often related to attitudes, opinions, and knowledge people hold rather than to externally measurable events. For this reason, the survey researcher must consider the type of statistical analyses desired during the design phase, and must ensure that the appropriate level of measurement is used. Using inappropriate analyses for the level of measurement can lead to incorrect conclusions about the meaning of the data collected. For most rating scales used in surveys, an interval level of measurement is assumed.

Ratings Scales

The most commonly used measurement tools in both observation and self-report research are rating scales. There are several basic types of scales. Among the most popular types are numerical scales, forced-choice scales, and graphic scales. These three types and examples of each are discussed below.

Numerical scales are rating scales in which the respondent works with a sequence of defined choices. Choices may be explicitly numbered on the survey, or the respondent may be presented with alternatives such as, "Strongly Agree," "Agree," "Disagree," and "Strongly Disagree." Examples of this type of scale are the **Likert scale** and the **Thurstone scale** (also called the **equal appearing intervals** scale). The problem with this type of scale is that it is subject to a specific type of bias called *response set*. This bias is discussed in greater detail in the section below titled *Response Effects*.

Forced-choice scales were originally developed to overcome response set bias. In this type of survey, respondents are presented with two equally desirable or two equally undesirable choices. A forced-choice survey question may ask, for example, "Which characteristic best describes your city officials: honest or intelligent?" Alternatively, a forced choice survey may present two statements such as, "Television networks should include programming that is suitable for family viewing" and "Television networks should limit airing of programs with offensive content to late night." The respondent is then asked to choose between the two statements. Some respondents resist such survey questions because they

are forced to make what they consider to be unreasonable choices. For example, a respondent may agree with both statements in the previous example.

The researcher using a forced-choice scale is typically interested in an overall pattern showing a preference for a particular type of response, rather than answers to specific questions. Researchers can gain a lot of information about public opinion, however, with a well-constructed forced-choice scale.

Graphic scales present a straight line vertically or horizontally with two opposite choices on either end of the line. Respondents mark the line at some point indicating the degree to which they rate the item relative to the two extreme positions. For example, a question on a graphic rating survey may ask the following: "How would you rate the performance of your mayor in attracting new industry to your city?" The response choices presented may be:
"Excellent _____Poor" or
"Excellent___:___:___:___:___Poor."
The respondent is asked to place a mark on the line indicating his or her opinion relative to the two end points. An example of a graphic scale is the **semantic differential scale**.

As stated above, ratings scales are commonly used in survey research. Constructing a valid rating scale requires a great deal of time, effort, and expertise, and scales that reliably measure abstract constructs such as attitudes can take years to construct.

Reliability and Validity

Using any survey instrument without testing it is poor research and can produce unexpected results. There is no way to predict how respondents will react to your survey questions or ratings scales without testing. At least one pilot study should be conducted for any newly developed scale to ensure its reliability and validity. A **pilot study** is a scaled-down version of the survey primarily conducted to test the survey instrument itself. A survey instrument must be both reliable and valid to be useful in any type of research.

Reliability refers to whether the instrument consistently gives the same answer. Any measurement instrument must be

KEY POINT

- **One way to avoid the biased responses likely with numeric and graphic scales is to use a forced-choice scale.**

KEY POINT

- **Any survey instrument must be pilot tested and must prove both reliable and valid before it can be used in a large-scale research study.**

reliable to be useful. For example, a bathroom scale that displays a different weight every time it's stepped on has very low reliability and is not very useful. Reliability has three components: stability, internal consistency, and equivalency.

Stability refers to consistency over time. This is also called **test-retest reliability**. In other words, a stable instrument yields the same result tomorrow that it yielded yesterday. This is an important characteristic in an instrument, but it can be difficult to establish in a survey instrument since people change over time.

Another important component of reliability is **internal consistency**. This simply refers to the fact that different parts of a scale should be consistent in the measurements they yield. For example, if the fuel gauge on my car indicates that I have half a tank of gas when there are actually eight gallons remaining in the tank, then indicates that I have a quarter tank when I have six gallons remaining, the internal consistency of the scale is low. Sometimes a technique called **split-half reliability** is used to assess the internal consistency of an instrument.

The third component of reliability is **equivalency**, which may also be referred to as cross-test reliability. This means that two equivalent forms of a test should yield the same measurements. A special case of equivalency is used to assess the degree to which different observers agree on ratings. This may be called **inter-coder reliability, inter-rater reliability, or inter-observer reliability**. For example, two judges who rate a movie "PG" because of violent content have high inter-rater reliability.

Validity refers to the ability of an instrument to measure what it is designed to measure. A valid measure is an accurate, useful measure. An example of a valid measurement instrument is a political poll that accurately predicts the outcome of an upcoming election.

Validity seems like a straightforward concept on the surface, but it is not an all or none property. Thus, we usually refer to the degree to which a measure is valid for a particular purpose. There are several types of validity. Among those that most

concern researchers are face validity, predictive validity, concurrent validity, and construct validity. Each of these will be discussed below.

Face validity refers to whether an instrument appears to be measuring what it is said to measure. Face validity is somewhat subjective because different people have different ideas about what is required to accomplish certain measurements. Laboratory experiments can sometimes seem far removed from the real-world events they are intended to predict. This does not mean that they are not valid, however. The extent to which an instrument has face validity is usually of minimal concern to most seasoned researchers.

Predictive validity is the ability of the instrument to predict future outcomes. For example, the Scholastic Aptitude Test measures characteristics related to success in an educational setting. It accurately predicts the likelihood of success in college. We can say that the SAT has high predictive validity.

Concurrent validity is closely related to predictive validity. To establish concurrent validity a researcher must test the new instrument against some established measure of the same thing. Both predictive validity and concurrent validity are sometimes referred to as criterion-related validity.

Construct validity is the extent to which there is evidence that a particular instrument measures a hypothetical construct. The process of demonstrating construct validity can be complicated. It usually involves demonstrating relationships between some set of variables that are predicted to be related to the construct. Whenever an abstract construct is measured it is necessary to demonstrate construct validity.

This is not a comprehensive list of all types of reliability and validity. You will find the definitions of other types of reliability and validity in the glossary.

KEY POINT

- **A measurement instrument may be reliable and yet not be valid. However, an unreliable instrument cannot be considered valid either.**

KEY POINT

- **Being able to extend your survey results to the entire population depends on selecting a sample that is representative of the population.**

Sampling: Who You Should Survey

When every member of the population of interest is questioned the survey is called a census. Because conducting a complete census is usually cost prohibitive, most research projects use sample surveys. A sample is a subset of the population. One of the most critical elements of survey planning is designing an appropriate sampling procedure. Using an inappropriate sampling procedure can render results invalid because results from the sample survey cannot be extended to the entire population. In other words, you cannot say that the same results would be obtained if you surveyed the entire population unless your sampling procedure yields an appropriate sample.

If results of a sample survey are to be generalized to the entire population, the sample must be **representative**. A representative sample is a subset of the population whose characteristics, in general, correspond to those of the population as a whole. Obtaining a representative sample depends on starting with the appropriate sampling frame. The **sampling frame** is the list from which the sample is selected. Random-digit-dialing computer programs are sometimes used to obtain survey samples. This technique is convenient but excludes members of the population who don't have a telephone. If it is important to include segments of the population who may not have telephones, using random-digit dialing does not produce an appropriate sampling frame. In addition to choosing the appropriate sampling frame, you must select an appropriate sampling technique. When you are interested in generalizing your results to all members of the population it is best to use a probability sampling technique.

Probability sampling techniques

Simple random sampling is a process in which each member of the population has an equal chance of being selected. Individuals are selected from the population by some purely random method such as drawing names from a hat or selection from a random number table. This is considered to be the most reliable way to get a representative sample. Of course, using this procedure depends on having all members of the population included in your sampling frame.

Two different types of simple random sampling are sampling without replacement and sampling with replacement. **Sampling without replacement** involves removing population members from the list when they are selected for a sample so that they cannot be selected again. **Sampling with replacement** involves returning members to the list so that they can be selected again. Each of these techniques is appropriate for particular research designs.

Systematic random sampling involves selecting every n^{th} member of the population from the sampling frame. Systematic random sampling may be appropriate if the list being used as the sampling frame has no regular pattern. If names on the list are grouped or ordered in some way, this procedure may not yield a random sample.

A stratified sampling technique is used when adequate representation of certain subsets of the population must be ensured. These subsets are called strata or segments.

Proportionate stratified sampling involves randomly selecting individuals from desired segments of the population. The number selected from each segment is based on that segment's proportion in the population. Inexperienced researchers sometimes think this technique is more likely to yield a representative sample than simple random sampling. Stratified sampling, however, is not more reliable than simple random sampling and can be extremely complicated and expensive to use in survey research because of non-response or people selected who refuse to participate.

Cluster sampling is a technique in which the population is divided into clusters, sometimes geographical areas. Each cluster or area is sampled using simple random sampling. This is considered a better overall alternative to stratified sampling because it is usually less costly, and social and economic strata are many times concentrated in geographical areas.

Multistage sampling involves first dividing the population into clusters then randomly selecting clusters or geographical areas to be sampled. From the selected clusters, individuals are chosen using simple random sampling.

KEY POINT

- **Simple random sampling is considered the best technique for obtaining a representative sample.**

KEY POINT

- **Non-probability sampling has particular applications but should not be used when the goal is to make predictions about the population based on a sample survey.**

Non-probability sampling

Non-probability sampling is sometimes used for special purposes such as pilot studies (testing the instrument itself) or marketing studies in which some particular target segment of population is of interest. However, use of these techniques is not advisable for samples intended to measure public attitudes, opinions, or general awareness because they do not yield representative samples.

Availability sampling is sampling from a collection of readily accessible subjects such as a group of students enrolled in a college class or shoppers in a mall. It is also known as a convenience sample. Availability samples are commonly used in pilot studies.

A volunteer sample is composed of people who volunteer. Research has shown that volunteers have particular characteristics that distinguish them from the general population. They tend to have a higher education and higher occupational status, a greater need for approval, and a higher intelligence level, among other characteristics. Thus, the results of volunteer surveys do not generalize well.

Purposive sample includes people who are selected on the basis of specific characteristics or qualities. Purposive samples are often used in advertising studies.

Quota sample includes people who are selected to meet a predetermined percentage because they have certain characteristics.

Sample Size

In order to generalize your findings your sample must be large enough to represent the population. There are no fixed rules for determining an appropriate sample size. A larger sample is usually more representative, but also costs more. An appropriate sample is large enough to represent the population but small enough to be cost-effective.

An important consideration in determining sample size is the estimated frequency of the phenomenon you intend to measure. If you want to measure something you expect to find fairly frequently in the population, a smaller sample will probably

include enough examples of the phenomenon to measure it. If you are looking for something relatively rare, you may need a larger sample. A special type of statistical procedure called a **power analysis** can help determine the required sample size for a given effect.

Sample Weighting

Sometimes after the survey is complete and the data analysis is begun a researcher discovers that certain segments of the population are grossly underrepresented in the sample. Rather than re-sampling or re-surveying, a statistical technique called sample weighting can be used to overcome the problems and correct the results of the survey. Sample weighting is not a substitute for using an appropriate sampling design. It is a corrective measure that can be used when samples are not representative because of uncontrollable factors such as low response rates among certain segments of the population.

KEY POINT

- **Sample weighting can be used to statistically correct results if non-response by certain segments of the population threaten the generalizability of the sample survey results.**

KEY POINT

- **Two important factors to consider when determining how a survey will be administered are probable response rates and cost of administering the survey.**

CONSTRUCTING A SURVEY INSTRUMENT

After determining the general design of the study, researchers must construct the survey instrument itself. Decisions about the general design will affect the form of the survey instrument; however, there are other considerations that are just as important. Determining how to best ask your questions is a key factor in getting the information you need. Among the factors that must be considered are how the survey will be administered and how the content, order, and formatting of your questions may affect responses. These topics as well as some typical response effects are discussed in this section.

Administering the Survey

There are two main modes of administering self-report surveys: questionnaires and interviews. Questionnaires usually take the form of written lists of questions that are self-administered by the respondents, whereas interviews are administered by a trained interviewer. Each method has its advantages and disadvantages. After briefly discussing the pros and cons of questionnaires and interviews in general, we will elaborate on some frequently used techniques for administering surveys.

Questionnaires are convenient because they can be administered to large numbers of people fairly economically. They also provide a relative level of anonymity to respondents. This may encourage people to reveal information they would not otherwise divulge. One disadvantage of self-administered questionnaires is that response rates can be somewhat disappointing. Another problem is that results can be biased because of individual differences that influence the likelihood of responding to a mail survey.

Interviews, on the other hand, have advantages as well. They provide an opportunity to establish rapport with the respondent and to encourage respondents to supply missing information. They also give the researcher greater control over the situation

by allowing the researcher to probe for additional information. The primary disadvantage of interviews is that they are more costly. Interviewers must be recruited and trained and the cost of compensating them for large studies can be cost-prohibitive. The four primary ways to administer questionnaires and interviews are discussed below.

Mail Surveys

Mail surveys are questionnaires that are self-administered and distributed through the mail. Mail surveys are cost-effective in terms of the number you can distribute, but the response rate is typically low. This is because the researcher has very little control over response. One technique for increasing response rate is an appeal that makes respondents feel obligated, like sending them a gift or making them feel privileged to be included. In addition to low response rates, mail surveys are subject to biased results because people with strong opinions or negative opinions are more likely to return mail surveys than are people with more moderate views.

Group Administration of Surveys

Group administration of survey questionnaires can be cost-effective if you use existing groups. This approach may increase response rate but may provide a biased sample, depending on the types of groups that are available to be surveyed. As with mail surveys, there is very little opportunity to interact with individual respondents because discussions between respondents and the researcher may influence the answers of others in the group. This produces the risk of contamination of data for the entire group.

Telephone Interviews

Telephone interviews are currently the most common way to administer surveys. Conducting interviews on the telephone is less costly than face-to-face interviews. Many people can be interviewed in a relatively short period of time with a minimum number of interviewers. However, interviewers must still be recruited and trained, making telephone interviews more costly than mail surveys.

KEY POINT

- **While questionnaires are initially less costly, interviews usually produce better response rates and thus more information in the long run.**

KEY POINT

- **Surveys with open-ended questions can be very difficult to administer, as well as difficult to analyze**.

The recent increase in telemarketing also poses a problem for telephone interviews. People get so many unwelcome calls they may be reluctant to participate. This may affect response rates; therefore, it is more important than ever that interviewers be well-trained.

Personal Interviews

The best response rates are usually achieved through personal (face-to-face) interviews. However, this method of collecting data is expensive and time-consuming. In addition, the personal contact with the interviewer is likely to produce biased responses. Bias may take the form of reluctance to discuss certain topics or **demand characteristics** in which the respondent is concerned with giving socially acceptable answers because of the personal contact.

Asking the Right Questions

The information you gather in a survey depends not only on what you ask, but also on how you ask the questions. Ambiguous wording or poor placement of questions can bias your results. Decisions about the format, content, and order of the questions on your survey may affect your result. Each of these topics is discussed below.

Format

The first decision you must make about the format of your questions is whether to use open-ended or closed-ended questions. **Open-ended questions** require respondents to generate their own answers. This allows respondents the opportunity to express any opinions they have. Skilled interviewers using open-ended questions may also use information supplied by the respondents to probe for more information or greater detail. Transcribing responses to open-ended questions can prove difficult, however, because interviewers must accurately record everything respondents say.

Quantifying responses to open-ended questions is not impossible, but it does require careful planning, and analysis is very work-intensive. The biggest problem involves

categorizing responses and assigning scores to these categories. Open-ended questions can be useful in small studies such as pilot-studies; however, the time involved in collecting and analyzing responses is considered a major disadvantage of the open-ended format.

Closed-ended questions require respondents to choose from supplied answers. Closed-ended formats are popular because the answers are easily quantified. The major disadvantage to closed-ended questions is that some information may be lost because researchers fail to include some important responses. When closed-ended questions are used, researchers must also decide how possible responses should be presented. This decision is usually made when a measurement scale is selected. (See the section titled *Measurement* for information on measurement scales.) General forms that answers can take are described below.

- **Dichotomous** answers such as yes/no or agree/disagree are easy to tabulate but are not sensitive to degrees of opinion. Thus the biggest problem with dichotomous response categories is the loss of data.

- **Multiple-choice** answers allow respondents to choose from several options. For example, the choices may be "all of the time," "most of the time," "some of the time," "rarely," and "never." Multiple-choice questions must include all possible answers, and answers must be mutually exclusive. In other words, the possibility of a respondent selecting two of the answers for one question must be eliminated.

- **Rating scales** are widely used in survey research because they allow respondents to express the strength of their beliefs in a way that is quantifiable (see information on numerical rating scales above).

- **Rank ordering** is a technique that allows respondents to express answers by placing a list in some order. This technique may be used if you are interested in perceptions of relative importance of some object or activity.

KEY POINT

- **Closed-ended questions are favored by most survey researchers because the resulting data is more easily analyzed.**

KEY POINT

- **The content of survey questions should always reflect the original purpose of the research.**

- The **Checklist** allows respondents to choose multiple items from a list. This type of response is used when you want to identify the most desirable alternatives from a list of possibilities.

- **Forced-Choice** questions list statements in pairs, or give pairs of answers for each statement. Respondents are required to select one statement or one response from each pair. Statements or responses paired together on the questionnaire usually express opposing opinions. Forced-choice questionnaires are usually relatively long with numerous question pairs on the same topic. Researchers using this technique are looking for particular patterns of responses that indicate attitudes or opinions of interest.

- **Fill-in-the-blank** questions are not used frequently by survey researchers. They can be useful, however, if you are trying to determine language or ideas that have meaning for most of the population (for example, to use in an advertising campaign).

Physical layout
Physical layout of the questions is the second important aspect of formatting. The layout should be clean and neat. Avoid layouts that are difficult to read. For example, do not string possible answers together on a single line unless there is space to clearly separate choices. Failing to allow enough space for respondents to indicate their answers can make scoring a questionnaire difficult. It also may cause loss of data because of questionnaires that cannot be scored. Also avoid layouts in which possible responses are not clearly matched with questions.

Content
In addition to the format of questions, the content or wording of the questions is an important consideration. Listed below are some general guidelines for wording questions.

1. Remember the purpose of the research. Ask only questions that directly relate to the topic of your study.

2. Make information you want to convey to the respondent clear. Questions must clearly and unambiguously communicate the desired information to the respondent.
3. Make clear what information you want from the respondent. Questions should be worded to allow accurate transmission of respondents' answers to researchers.
4. Avoid leading questions. Word your questions in such a way that you do not lead respondents toward a particular answer.
5. Avoid biased wording or terms. Do not use wording that makes assumptions or that could cause respondents to choose a particular answer.
6. Avoid potentially embarrassing questions. Asking for information that may be considered too personal has the potential to inhibit or bias responses to other questions.
7. Do not ask for highly detailed information. Respondents may become frustrated and fail to complete the questionnaire.
8. Avoid double-barreled questions. Do not ask two or more questions in the same sentence. This is confusing to respondents and presents scoring problems as well.
9. Keep questions short. Be concise. Most respondents can better understand brief questions.
10. Limit the length of the questionnaire. Whether your survey is a self-administered questionnaire or an interview, if it takes too much time many people will not be willing to complete it.

Question Order

Besides the wording and format of the questions, the order in which you ask the questions has an effect on responses. In general, initial questions should be simple and easy to answer. If you ask the hardest questions first, people may get discouraged and quit. You should also place the demographic data and sensitive items at the end of the questionnaire. Respondents are more inclined to give this type of information after they've seen the nature of the questions in the main body of the survey.

KEY POINT

- **The order of the survey questions should be carefully planned to avoid eliciting responses that are biased by preceding questions.**

KEY POINT

- **Instructions on questionnaires and interview scripts must be clear and precise.**

Questions should be organized in a logical sequence, proceeding from the general to the specific. Questions on similar topics should be grouped together. Both these techniques aid respondents' memory for information. This is called a **memory priming effect**.

You must be careful about using memory priming, however, as it can be a double-edged sword. The priming effect can be beneficial by helping people remember information, but it can also cause **question contamination.** Questions become contaminated when proceeding items bias respondents towards a particular answer or a particular end of the scale. For example, if you ask several questions about corruption in politics then ask respondents to rate corruption in the campaign of a particular politician, you are likely to bias the responses toward greater corruption.

Instructions

In addition to the questions themselves, your questionnaire or interview script will contain instructions. Adequate instructions are critical. Wording of instructions can be just as important as wording of survey questions.

Self-administered questionnaires must have clear and precise instructions for recording responses. It is also necessary to include explicit information on how to return the questionnaire.

Instructions for interviewers must include question scripts with directions for how the questions should be read. For example, if you want the interviewer to stress a particular word, it should be indicated on the script. You should also include predetermined probe points and clear directions about recording answers. Interviewers must be trained to read the interview script in the same way to each participant. Survey results can be rendered invalid by interviewers who adlib excessively or who fail to accurately record responses.

Response Effects

Response effects are factors that bias individual respondents in favor of particular answers or particular ends of a scale. Response effects produce systematic biases in survey results that may render them invalid. Response effects may be characteristics of the survey instrument, the interviewer, or the respondent. Care should be taken to avoid response effects whenever possible. The three main categories of response effects are discussed below.

Task Effects

Task effects are characteristics of the questionnaire or the interview script that bias responses. Many of these factors, such as order effects and question content, are discussed previously in this section.

Interviewer Effects

Interviewer effects are characteristics of the interviewer that affect the way respondents answer questions. Do not confuse interviewer effects with experimenter effects**. Experimenter effects**, also called **researcher bias**, occur when beliefs, attitudes, or knowledge causes the experimenter to behave in a way that influences participants. Interviewer effects occur when particular characteristics of the interviewer like age or race bias respondents in favor of particular answers.

Research on interviewer effects has shown that there are only a few factors that consistently cause such problems: primarily age, experience, and training. To ensure that survey responses are not biased you should select interviewers that are experienced in survey research and provide adequate training. You should also select interviewers who are similar in age to survey respondents. For example, it may not be wise to use college-aged interviewers for a survey of the business community.

KEY POINT

- **To avoid response effects researchers must carefully construct the survey instrument, eliminating all factors known to bias responses; they must also carefully select and thoroughly train interviewers.**

KEY POINT

- **Some researchers try to avoid bias cause by non-response by offering incentives for returning mail questionnaires.**

Respondent Effects

Respondent effects bias specific individuals in favor of particular responses. Research has revealed many individual differences that can cause biased responses. One type of respondent effect is the **response set**: participants with a particular response set are more inclined to respond in the same way to all questions. For example, some individuals tend to give positive responses most of the time; this is called an **acquiescent response set**. Others are unduly influenced by **social desirability** and so are reluctant to give responses they feel are socially unacceptable or that make them feel inferior in some way. There are methods for screening respondents for these biases, but the cost of doing so may be prohibitive. Some researchers believe that these biases are part of natural variability in the population due to individual differences and should be treated as such.

One of the largest biasing effects related to the characteristics of the respondents themselves is **non-response bias**. Non-response can bias results because people who refuse to respond may be systematically different from those who willingly participate in research. Sometimes this effect is referred to as the **volunteer sample effect**. To avoid this bias in your survey you must use techniques to ensure an adequate number of responses from those who may be reluctant to participate under certain conditions. For example training interviewers to be sensitive and offer to call back at a more convenient time when the person says they are busy may encourage reluctant participants to grant an interview.

INTERPRETING SURVEY RESULTS

The goal of any research project is to reach some clear-cut conclusions. In short, the objective is always to answer the original research question. The first steps to understanding your research results are taken during the design of your survey. Decisions about how the data will be analyzed must be made early because they influence the way your survey instrument is constructed and the way your sample is selected. The types of statistical analyses performed depend on your research questions. Questions you must answer to determine appropriate analyses include the following:

- Are you interested in describing some event or state of a population or sample?
- Are you looking for relationships among variables?
- Are you interested in comparing groups of respondents?
- Are you interested in looking at changes over time?

You must also consider how large your survey sample will be, how many variables you are measuring, and how you will take your measurements.

KEY POINT

- **Answering the original research question is the primary objective in data analysis.**

Interpreting Survey Statistics

There are many ways to analyze survey statistics. This section outlines several basic types of analyses and gives examples of each.

Descriptive statistical procedures are used to classify and summarize data. If the descriptive measure is calculated from population data it is called a **parameter**. If it is calculated from a sample it is called a **statistic**. Descriptive measures include the following:

- **Frequencies** are statistics that are simply a count of the variable of interest.

KEY POINT

- **If survey data does not meet the theoretical assumptions of the procedure being used, results of any statistical analysis can be misleading.**

- **Proportions or percentages** describe frequencies in the context of the entire sample.
- **Averages**, in particular the mean, mode, and median, which are measures of central tendency.
- **Variability** measures, such as the variance and standard deviation, which are measures of variation.

Correlational analyses are used to quantify the relationship between two variables. The type of correlation procedure used depends on the level of measurement of the variables. Some examples of correlational measures are:

- **Pearson Product Moment correlation**, also called Pearson's *r,* which is used with scores measured on an interval or ratio scale.
- **Point-Biserial correlation**, which is used with one interval or ratio level variable and one nominal variable.

Analyses of differences are used to test whether the difference between scores of two groups or two different conditions of testing is statistically significant. Examples of these analyses are listed below.

- ***t* tests** are used to compare one score of one group with a known standard, or to compare the scores of two groups.
- **Analysis of variance (ANOVA)** is used when comparing the scores of more than two groups.
- **Chi-Square** is used to compare observed rates with expected rates of certain types of responses to survey questions.

Analyses of change include special forms of *t* tests and ANOVAs that can be used to measure change over time. Examples of these procedures are:

- ***t* test** for dependent samples compares scores from the same sample at two different times.

- **Repeated measures ANOVA** compares scores from one or more samples collected at more than one time.

An important thing to remember about statistical analysis is that each procedure is based on a set of assumptions about the data being analyzed. If the assumptions are not met, the results of the analysis are considered unreliable. For example, using a procedure that assumes a random sample from the population is inappropriate if your sampling procedure was non-random.

Statistical Significance
Researchers and analysts frequently use the term "statistically significant findings." When research findings reach the popular press the word "statistically" is usually omitted and the word "significant" is emphasized. This paraphrasing leads to a misunderstanding about the "significance" of research results. The real-world significance of a particular research finding and the finding that results are "statistically significant" are very different matters. When a researcher finds through statistical analysis that results are significant, the finding really means that the result is not likely to be due to a **sampling error**. That is, the result would not be expected to show up by chance in a random sample. In other words, what the researcher has measured in the sample is likely to be present in the entire population. That doesn't mean that the finding is so important that it will change the world. It simply means that the researcher can extend the findings from the sample to the population.

KEY POINT

- **Statistical significance means that the finding is likely to be a relatively accurate measure of some characteristic of the population, rather than the result of a sampling error. It does not imply real-world significance or importance.**

Reading Charts and Graphs

Visual representations are generally used to summarize data and to emphasize relationships or trends. After we have designed the survey and collected and analyzed the data, tables, charts, and graphs can help us understand what our data means.

Tables are used to summarize data. They can describe respondents and show relationships and changes. Tables usually have a title that summarizes the purpose and content.

KEY POINT

- **Do not be deceived by graphics that show effects that are disproportionate to the actual effects in the data.**

Tables display columns and rows of numbers, percentages, scores, and statistical test results. The columns are typically used to display categories such as segments of the population or other groups or conditions whose scores you may want to compare. Rows show the different scores.

Charts and graphs usually have a title that summarizes the purpose or content. They also have a legend or key that allows you to interpret visual elements.

Pie Charts are used to show the proportion of the whole that each response category occupies. Pie charts are circles that represent the whole segmented into wedges that represent the proportion of each category.

Bar Graphs are used to show comparisons of groups or categories of scores. The groups or categories are on the horizontal axis and the scores are displayed on the vertical axis. An exception to this is the horizontal bar graph. In this type of graph the groups or categories are on the vertical axis and the scores are on the horizontal axis.

Histograms are a special type of bar graph for displaying frequencies of occurrence. Typically, the variables we have measured or the groups we have tested are on the horizontal axis and the frequencies are on the vertical. Horizontal histograms are also used frequently.

Line Graphs are used to show changes over time or across conditions. Time or conditions of surveying are usually displayed on the horizontal axis and scores are represented on the vertical axis.

Visual displays of data can be deceptive. A graph may display a difference in the scores of two groups, but that does not mean that the difference is significant. Scaling of graphs can also be misleading. Be sure to look at the labels and legends before jumping to the conclusion that the distances on the graph are truly representative of differences in the data.

PRESENTING FINDINGS

Possibly the most effective way to present complex survey results is in a graphical format. Graphs, charts, and tables can summarize data clearly and concisely. Diagrams or pictures can be used to describe a survey design or to represent the time line of the study. Pictures can be worth a thousand words, but they can also be misleading.

What Is a Picture Worth?

Visual elements of a graph or chart convey as much meaning as the text, or perhaps more. When designing visual presentations of survey results you must be careful that the picture you present is a true representation of the data. This section includes tips for producing effective graphics and examples of good and bad graphs.

Guidelines for developing graphics:

1. Tables, graphs, and diagrams should always have a title that summarizes their purpose and content.
2. All information the viewer needs to understand the meaning of the graphic should be included in the title, labels, and legends. In other words, the graphic should be "self- contained."
3. Diagrams should illustrate points and tables and graphs should summarize data clearly and succinctly. Graphics should add to the understanding of the information.
4. Data presented in tables and graphs should be logically organized (for example, from smallest values to largest values, or vice versa).
5. Use extra lines and grid marks sparingly. The appearance should be clean and neat.

KEY POINT

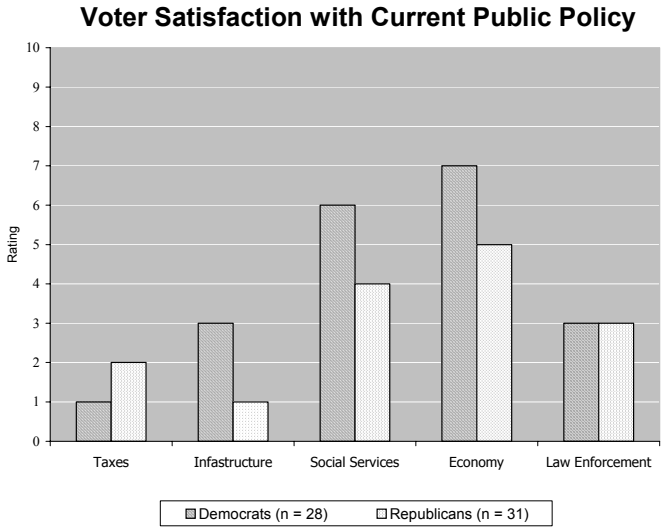
- **The integrity of your survey presentation can be compromised by graphics that do not accurately portray results.**

KEY POINT

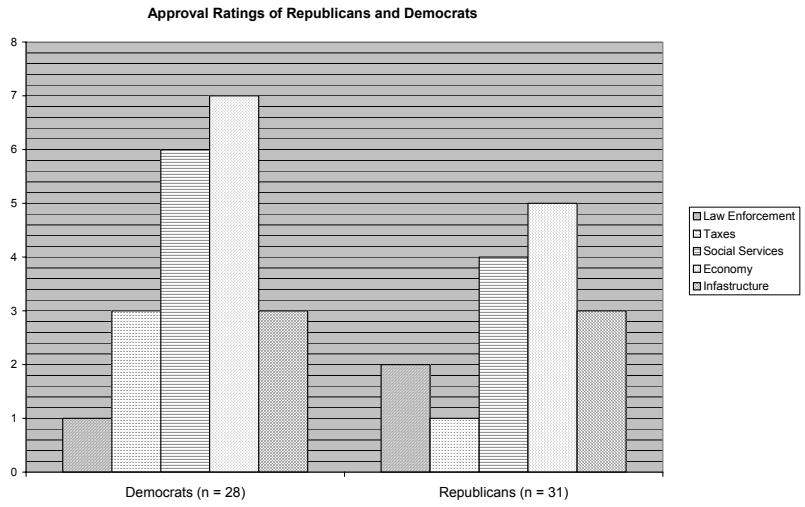
- **Keep in mind that the purpose of a graphic is to increase understanding of the procedure you used or the results of your survey. Design graphics to emphasize the information you wish to convey.**

6. Emphasize the data. Lighten graph elements that do not contain information, like organizational elements such as vertical and horizontal lines.
7. Be very careful that you do not deceive the viewer by making differences appear more dramatic that they really are. Differences portrayed in the graphics should be proportional to differences in the data.

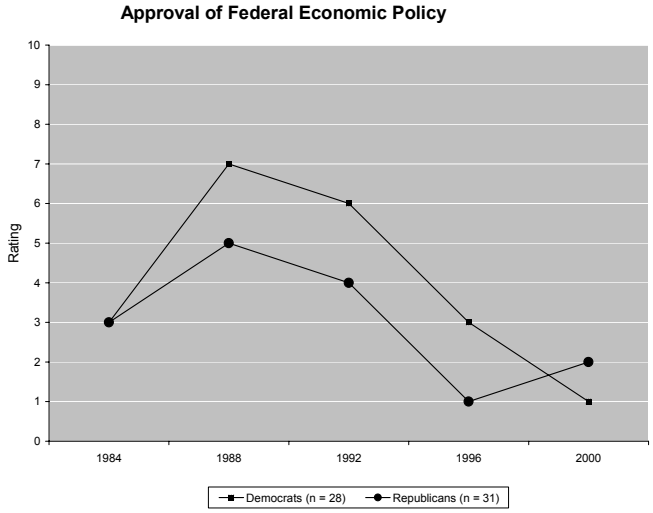
Below are some examples of good graphics and bad graphics. As discussed above, good graphics are clear, precise and logically organized. Bad graphics leave the reader or listener confused about what you are trying to communicate.



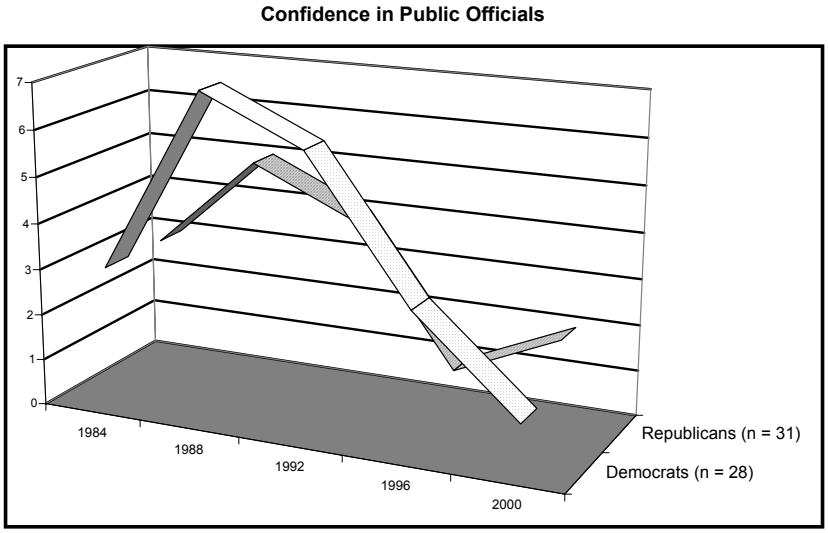
The graph above is an example of a good bar graph. The title clearly states what the graph displays. The graph itself is clean and easy to read. The grid lines and labels do not overwhelm the data. The scale appropriately displays the magnitude of the ratings and the relationship of the ratings of Democrats and Republicans. The sample size for each group is clearly stated. All information the reader needs to understand the data is included in the graphic.



This is an example of a bad graph. The title is confusing and the way the data are organized does not allow you to easily understand the point of the display. The graphic also looks cluttered because of the extra grid lines and patterned bars.



This is an example of a good line graph. Again, the purpose of the graph is clear from the title and the organization of the data. The graph stands alone. The reader does not need to read anything or look up extra information in the research report to understand what the graphic means.



This is an example of a bad line graph. There is no purpose for the three dimensional nature of the graph. The title and labels do not give enough information to interpret the data.

Keep in mind that the purpose of using a graphic is to make difficult data easier to understand. You should be sure to include all the information needed to understand what the graph is about. You should also eliminate graph elements that do not communicate information. Doing these two things can make the difference between a graph the aids understanding of your survey and a graph that leaves your audience confused.

CONCLUSION

Surveys are frequently used to obtain information that would be difficult to get in other ways. Surveys involve the use of a predetermined set of questions, generally in the form of a questionnaire or an interview. Planning a good survey involves making various decisions, such as what form the questions will take and how the responses will be scored and analyzed.

Most surveys involve some form of sampling. The general goal of sampling is to identify a group of individuals whose responses can be generalized to the population of interest. How well responses can be generalized depends on the representativeness of the sample, that is, how well the characteristics of the sample correspond to those of the population.

The construction of a good questionnaire is often time consuming. Issues requiring attention include the wording of individual questions and the format of the questionnaire itself. Another factor that can affect responses is question order. Information you gather in a survey depends on what you ask, how you ask, and in what order you ask questions.

Administering the survey can be accomplished through mail questionnaire, group-administered questionnaire, personal interview, or telephone interview. Each of these methods has advantages and disadvantages. Knowing which method to choose depends on the nature of the survey itself and on related factors, such as budget constraints.

The analysis of survey begins during the design phase of the study. It is important to design your analyses to answer your original research question. The design of your analyses will affect the design of your questionnaire or interview as well as the way you select your sample and administer your survey. You cannot wait until you have finished collecting data to think about how to analyze your results.

Conducting a successful survey is hard work. It requires planning, skill, and practical experience. The information gained from a well done survey, however, is worth the effort.

REFERENCES

Fink, A. & Kosecoff, J. *How to Conduct Surveys: A Step by Step Guide* (2nd ed.) Thousand Oaks, CA: Sage Publications. 1998.

Handbook of Survey Research. Eds. Rossi, P. H., Wright, J. D., and Anderson, A. B. San Diego, CA: Academic Press. 1983.

Kish, Leslie. *Survey Sampling*. New York. John Wiley & Sons, Inc.1995.

Mahler, I. (1953). Attitudes toward socialized medicine. *Journal of Social Psychology*, 38, 273-282.

Rosnow, R. L. & Rosenthal, R. *Beginning Behavioral Research: A conceptual Primer*. New York: Macmillian Publishing Company. 1993.

Stevens, James. *Applied Multivariate Statistics for the Social* Sciences. Hillsdale, NJ: Lawrence Erlbaum Associated, Publishers. 1986.

Tufte, Edward R. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press. 1983.

U. S. Bureau of the Census. *Statistical Abstracts of the United States* (116th ed.). Washington, DC. 1996.

Wimmer, R. D. & Dominick, J. R. Mass Media Research: An Introduction. (5th ed.). Belmont, CA.: Wadsworth Publishing Company. 1997.

GLOSSARY

In addition to the terms in bold print introduced in other sections of this guide, you may encounter the following terms when dealing with survey contractors. You may also see some of these terms in statistical documents such as *Statistical Abstracts of the United States*.

📖 **Acquiescent response set**: the tendency of individuals to go along with any request or attitudinal statement.

📖 **Alpha (α):** probability of a Type I error; synonymous with significance level and p value.

📖 **Analysis of variance (ANOVA)**: a statistical procedure used to decompose sources of variation in two or more independent variables.

📖 **Archival records**: the relatively permanent repositories of information that can be used in secondary observation.

📖 **Area probability sampling**: a type of survey sampling in which the subclasses are geographic areas; a type of cluster sampling.

📖 **Artifact**: a variable that creates an alternative explanation of results (a confounding variable).

📖 **Available sample**: a sample selected on the basis of accessibility.

📖 **Bar chart**: a chart or graph that displays scores or frequencies by category or group.

📖 **Beta (β):** probability of a Type II error.

📖 **Bias:** net systematic error.

📖 **Census**: a survey in which the sample comprises every element of a population.

📖 **Central tendency**: a single value that is chosen to represent a typical score in a distribution of scores. Frequently used measures of central tendency are the mean, the mode, and the median.

📖 **Checklist question**: a type of survey question in which the respondent is given a list of items and is asked to mark those that apply.

📖 **Chi-square statistic**: a measurement of observed versus expected frequencies. Often referred to as crosstabs.

📖 **Closed-ended question**: a question the respondent must answer by making a selection from a prepared set of options.

📖 **Cluster analysis**: a multivariate statistic that classifies phenomena into groups or segments.

📖 **Cluster sample**: a sample taken from a cluster or a sample placed into groups or categories

📖 **Codebook**: a menu or list of responses used in coding open-ended questions.

📖 **Coefficient of determination**: the proportion of variance shared by two variables.

📖 **Cohort analysis**: a study of a specific population as it changes over time.

📖 **Concept**: a term that expresses an abstract idea formed by generalization.

📖 **Confidence interval**: an interval within which there is a stated probability that the statistic (for example the mean) will fall.

📖 **Confidence level**: probability of rejecting a null hypothesis that is true. Also called alpha level.

📖 **Confounding**: when the effect of one variable is mixed up with the effect of another variable so that the researcher cannot distinguish between them.

📖 **Construct**: a combination of concepts that is created to describe a specific situation or state of being (for example, authoritarianism is a construct).

📖 **Construct validation**: the procedure by which a means is devised for the measurement of a construct and is then related to the participants' performance in a variety of other spheres as the construct implies.

📖 **Construct validity**: the degree to which the test is truly measuring the construct of interest.

📖 **Content validity**: A type of test validity that addresses whether the test adequately samples the relevant material.

📖 **Continuous variable**: a variable that can take on any value over a range of values and can be meaningfully broken into subparts (for example, height).

📖 **Control group**: participants who do not receive experimental treatment and thus serve as a basis for comparison in an experiment.

📖 **Convenience sample**: a non-probability sample consisting of respondents or subjects who are available, such as college students in a classroom.

📖 **Convergent validity**: validity supported by a substantial correlation of conceptually similar measures.

📖 **Correlation**: degree of relation between variables.

📖 **Correlation coefficient**: an index of the degree of association between two variables, typically Person *r,* or related product-moment correlation.

📖 **Correspondence with reality**: the extent to which a theory or hypothesis agrees with accepted truths based on reliable findings.

📖 **Criterion variable**: in survey research, the variable presumed to be the effects variable. In experimental research, the dependent variable.

📖 **Criterion validity**: the extent to which a measure correlates with one or more outcome criteria.

📖 **Cross-lagged correlation**: a type of longitudinal study in which information about two variables is gathered from the same sample at two different times. The correlations between variables at the same point in time are compared with the correlations at different points in time.

📖 **Cross-tabulation analysis (crosstabs)**: a measurement of observed versus expected frequencies.

📖 **Data archives**: data storage facilities where researchers can deposit data for other researchers to use.

📖 **Demand characteristic**: the premise that participants' awareness of the research condition may affect their performance. Also known as the Hawthorn effect.

📖 **Descriptive research**: research in which the objective is to map out a situation or set of events.

📖 **Descriptive statistics**: statistical methods and techniques designed to reduce data sets to allow for easier interpretation.

📖 **Descriptive survey**: a survey that attempts to picture or document current conditions or attitudes.

📖 **Design-specific results**: research results that are based on, or specific to, the research design used.

📖 **Discrete variable**: a variable that can be conceptually subdivided into a finite number of indivisible parts (for example, the number of children in a family).

📖 **Discriminate analysis**: a multivariate statistic used to classify groups according to variable similarities or to analyze the statistical significance of a weighted linear combination of variables.

📖 **Discriminate validity**: validity supported by a lack of correlation between conceptually unrelated measures.

📖 **Dispersion**: the amount of variability in a set of scores.

📖 **Disproportionate stratified sampling**: over-representing a specific stratum (segment) or characteristic in a sample.

📖 **Distribution**: a collection of scores or measurements.

📖 **Double-barreled question:** a single question that in reality requires two separate responses (for example, "Do you like the price and style of this item?").

📖 **Double-blind experiment**: a research study in which experimenters and others do not know whether a given participant belongs to the experimental group or the control group.

📖 **Dummy variable**: the variable created when a variable at the nominal level is transformed into a form more appropriate for higher order statistics.

📖 **Ecological validity**: the extent to which an experimental situation reflects the outside world it is intended to represent.

📖 **Effective reliability**: the composite reliability of two or more judges' ratings.

📖 **Empirical validity**: another name for criterion validity.

📖 **Equal-appearing intervals method**: an attitude scaling technique in which values are obtained for items on the assumption the underlying intervals are equidistant; also called a *Thurstone* scale.

📖 **Equivalency**: the internal consistency of a measure.

📖 **Equivalent-forms reliability**: the extent to which different forms of a test are intercorrelated.

📖 **Error**: deviation of a score from the mean of the group or condition.

📖 **Error variance**: the variability in scores created by unknown factors.

📖 **Evaluation apprehension**: a fear of being measured or tested, which may result in invalid data.

📖 **Exhaustivity**: a state of a category system such that every unit of analysis can be placed into an existing slot.

📖 **Expected frequencies**: in a Chi-square analysis, counts expected under specified row and column conditions if certain hypotheses were true.

📖 **Experimental design**: a blueprint or set of plans for conducting laboratory research.

📖 **Experimental-expectancy effect**: artifact that results when the hypothesis held by the experimenter leads unintentionally to behavior toward participants that, in turn, increases the likelihood that the hypothesis will be confirmed.

📖 **External validity**: the degree to which the results of a research study are generalizable to other situations.

📖 **Face validity**: the extent to which a test seems on its surface to be measuring what it purports to measure.

📖 **Factor analysis**: a multivariate statistical procedure used primarily for data reduction, construct development, and the investigation of variable relationships.

📖 **Factor score**: a composite or summary score produced by factor analysis.

📖 **Field observation**: a study of a phenomenon in a natural setting.

📖 **Filter question**: a question designed to screen out certain individuals from further participation in a study; also called screener question.

📖 **Flesch reading ease formula**: an early readability formula based on the number of words per sentence and the number of syllables per word.

📖 **Focus group**: an interview conducted with 6-12 participants simultaneously and a moderator who leads a discussion about a specific topic.

📖 **Fog Index**: a readability scale based on sentence length and the number of syllables per word.

📖 **Follow-back panel**: a research technique in which a current cross-section sample is selected and matched with archival data.

📖 **Forced-choice question**: a question that requires a participant to choose between two specified responses.

📖 **Frequency curve**: a graphical display of frequency data in the form of a smooth, unbroken curve.

📖 **Frequency distribution**: a collection of scores, ordered according to magnitude, and their respective frequencies.

📖 **Frequency polygon**: a series of lines connecting points that represent the frequencies of scores.

📖 **Good subjects**: research participants who seek to provide responses that will validate the experimenter's hypothesis.

📖 **Gross incidence**: the percent of qualified respondents reached of all contacts made.

📖 **Group administration**: conducting measurements with several participants simultaneously.

📖 **Halo effect**: a response set in which the bias results from the judge's overextending a favorable impression of someone, based on some central trait, to the person's other characteristics.

📖 **Histogram**: a bar chart that illustrates frequencies and scores.

📖 **Homogeneity**: equality of control and experimental groups prior to an experiment: also called point of prior equivalency.

📖 **Hypothesis:** a tentative generalization concerning the relationship between two or more variables that predicts an experimental outcome.

📖 **Identity matrix:** a square matrix whose elements equal zero except for those along the principal diagonal, which equal one.

📖 **Incidence:** the percentage of a population that possesses the desired characteristics for a particular research study.

📖 **Independent variable:** the variable that is systematically varied by the researcher.

📖 **Index number:** a measure of difference or change, usually expressed in percentage or proportions.

📖 **Inter-coder reliability**: the degree of agreement between or among independent coders or judges.

📖 **Inter-correlation matrix**: a matrix composed of correlations between pairs of variables.

📖 **Internal consistency**: the level of consistency of performance among items within a scale.

📖 **Internal validity**: a property of a research study such that results are based on expected conditions rather than on extraneous variables.

📖 **Interval level**: a measurement system in which the intervals between adjacent points on a scale are equal (for example, a thermometer).

📖 **Isomorphism**: similarity of form or structure.

📖 **Leading question**: a question that suggests a certain response or makes an implicit assumption (for example, "How long have you been an alcoholic?").

📖 **Likert scale**: a measurement scale where respondents strongly agree, agree, are neutral, disagree, or strongly disagree with the statements.

📖 **Longitudinal research**: research in which the same participants are studied over a period of time.

📖 **Mail survey**: the mailing of self-administered questionnaires to a sample of people: the researcher must rely on the recipients to mail back their responses.

📖 **Marker variable**: a variable that highlights or defines the construct under study.

📖 **Mean**: the arithmetic average of a set of scores.

📖 **Measurement**: a procedure whereby a researcher assigns numerals to objects, events, or properties according to certain rules.

📖 **Measurement error**: an inconsistency produced by the instruments used in a research study.

📖 **Median**: the midpoint of a distribution of scores.

📖 **Mode**: the score that occurs most often in a frequency distribution.

📖 **Mortality**: in longitudinal research, the percent of the original sample that drops out of the research project before the end.

📖 **Multistage sampling**: a form of cluster sampling in which samples are selected in stages.

📖 **Multivariate analysis of variance (MANOVA)**: an extension of analysis of variance used to study more than one dependent variable.

📖 **Multivaraite statistics**: statistical methods that investigate the relationship between one or more independent variables and more than one dependent variable.

📖 **Mutually exclusive**: a category system in which a unit of analysis can be placed in one and only one category.

📖 **Need for social approval**: the tendency of individuals to respond in ways that elicit a favorable evaluation.

📖 **Nominal level**: the level of measurement at which arbitrary numerals or other symbols are used to classify persons, objects, or characteristics**.**

📖 **Non-parametric statistics**: statistical procedures used with variables measured at the nominal or ordinal level.

📖 **Non-probability sample**: a sample selected without regard to the laws of mathematical probability.

📖 **Non-response bias**: error that is due to non-response or non-participation.

📖 **Normal curve**: a symmetrical, bell-shaped curve that possesses specific mathematical characteristics.

📖 **Normal distribution**: a mathematical model of how measurements are distributed. A graph of a normal distribution is a continuous, symmetrical, bell-shaped curve.

📖 **Null hypothesis**: the denial or negation of a research hypothesis.

📖 **Observed frequencies**: counts obtained in specific rows and columns in a *Chi-square* analysis.

📖 **Open-ended question**: a question to which respondents are asked to generate an answer or answers with no prompting from the item itself (for example, "What is your favorite type of television program?").

📖 **Operational definition**: the meaning of a variable in terms of the operations necessary to measure it or in the experimental methods involved in its determination.

📖 **Ordinal level**: the level of measurement at which items are ranked along a continuum.

📖 **Panel study**: a research technique whereby the same sample of respondents is measured at different points in time.

📖 **Parameter**: a characteristic or property of a population.

📖 **Parametric statistics**: statistical procedures appropriate with variables measured at the interval or ratio level.

📖 **Parsimony principle**: the premise that the simplest method is the most preferable.

📖 **Partial correlation**: a method used to control a confounding or spurious variable that may affect the relationship between independent and dependent variables.

📖 **Percentile**: location of a score in a distribution defining the point below which a given percentage of the cases fall (for example, a score at the 90th percentile falls at a point where 90% of the scores fall at or below that score).

📖 **Periodicity**: any form of bias resulting from the use of a non-random list of participants or items in selecting a sample.

📖 **Pilot study**: a trial run of a study conducted on a small scale to determine whether the research design and methodology are relevant and effective.

📖 **Point estimates**: estimates of particular characteristics of the population that may be made from a sample (for example, the mean of the sample is a point estimate for the mean of the population).

📖 **Population**: a group or class of objects, people, or units in which a researcher is interested. Also known as the target universe.

📖 **Population distribution**: the frequency distribution of all the variables of interest as determined by a census of the population.

📖 **Population values**: the true score of the population on some variable as opposed to estimates from sample measurements.

📖 **Population parameters**: statistics computed from population values as opposed to estimates calculated from samples.

📖 **Power**: in statistical hypothesis testing, the probability of rejecting the null hypothesis when an alternative is true.

📖 **Predictive validity**: the extent to which a test can predict future outcomes.

📖 **Predictor variable**: the variable used to predict another variable. Also the possible cause of some effect.

📖 **Prestige bias**: the tendency of a respondent to give answers that will make him or her seem more educated, successful, financially stable, or otherwise prestigious.

📖 **Probability**: the mathematical chance of an event's occurring.

📖 **Probability level**: a predetermined value at which researchers test their data for statistical significance.

📖 **Probability sample**: a sample selected according to the laws of mathematical probability.

📖 **Proposition**: a statement of the form "if A then B," which links two or more concepts.

📖 **Proprietary data**: research data gathered by a private organization that are available to the general public only if released by that organization.

📖 **Protocol**: a document containing the procedures to be used in a field study.

📖 **Purposive sample**: a sample deliberately chosen to be representative of a population.

📖 **Qualitative research method**: a description or analysis of a phenomenon that does not depend on the measurement of variables.

📖 **Quantitative research method**: a description or analysis of a phenomenon that involves specific measurements of variables.

📖 **Quota sample:** a sample selected to represent certain characteristics of interest.

📖 ***r*:** Pearson's product-moment correlation. Defines the relationship of two variables.

📖 ***r^2*:** Pearson's correlation squared; also called the proportion of variance accounted for, or coefficient of determination.

📖 **Random digit dialing:** a method of selecting telephone numbers that ensures that all telephone households have an equal chance of being selected.

📖 **Random error:** the effect of uncontrolled variables that cannot be specifically identified. Theoretically self-canceling, in that the average of the errors will probably equal zero.

📖 **Random sample:** a subgroup or subset of a population selected in such a way that each unit in the population has an equal chance of being selected.

📖 **Range:** a measure of dispersion based on the difference between the highest and lowest scores in a distribution.

📖 **Ratio level:** a level of measurement that has all the properties of the three lower level scales and also has a true zero point.

📖 **Rating scales:** the common name for a variety of measuring instruments on which the observer or judge gives a numerical value (either explicitly or implicitly) to certain judgements or assessments.

📖 **Reactivity:** a participant's awareness of being measured or observed and its possible impact on that participant's behavior.

📖 **Readability:** the total of all elements in a piece of printed material that affect the degree to which people understand the piece.

📖 **Redundancy index:** a mathematical procedure used in canonical correlation to aid in interpreting relationships between variable sets.

📖 **Reliability:** the extent to which observations or measures are consistent or stable over time.

📖 **Repeated-measures design**: a research design wherein numerous measurements are made on the same participants.

📖 **Replication:** an independent verification of a research study.

📖 **Research question:** the question that serves as the focus of a research study. The goal of the research is to answer this question.

📖 **Residuals:** effects left over when all sources of known variation are subtracted from scores or means.

📖 **Sample:** a subgroup or subset of a population or universe.

📖 **Sample distribution:** the frequency distribution of all the variables of interest as determined from a sample.

📖 **Sample selection bias:** systematic error resulting from the nature of the sampling units.

📖 **Sample plan:** a design, scheme of action, or procedure that specifies how the participants are to be selected in a survey study. Also called a sample design.

📖 **Sample-specific results**: research results that are based on, or specific to, the research sample used.

📖 **Sampling distribution**: a probability distribution of all possible values of a statistic that would occur if all possible samples of a fixed size from a given population were taken.

📖 **Sampling error**: the degree to which measurements obtained from a sample differ from the measurements that would be obtained form the population.

📖 **Sample estimates**: values or statistics computed from samples that serve as estimates for population values.

📖 **Sampling frame**: a list of the members of a particular population from which a sample is drawn.

📖 **Sampling interval**: a random interval used for selection of participants or units in the systematic sampling method.

📖 **Sampling rate**: the ratio of the number of people chosen in the sample to the total number in the population.

📖 **Scale**: a form of measurement such as 10-point scales, Likert, Guttman, or semantic differential.

📖 **Scattergram**: a graphic technique for portraying the relationship between two variables.

📖 **Screener**: a short survey or portion of a survey designed to select only appropriate respondents for a research project.

📖 **Semantic differential**: a rating scale consisting of seven spaces between two bipolar adjectives (for example, "good-------bad").

📖 **Sigma (Σ):** the Greek capital letter symbolizing "the sum of."

📖 **Simple random sampling:** a sampling plan in which the participants are selected individually on the basis of a randomized procedure (e.g., a table of random numbers).

📖 **Skewness**: the degree of departure of a curve from the normal distribution (curves can be positively or negatively skewed).

📖 **SMOG Grading**: a measure of readability based on the number of syllables per word.

📖 **Stability**: the degree of consistency of the results of a measure at different points in time.

📖 **Standard deviation**: the square root of the variance (a mathematical index of dispersion).

📖 **Standard error**, or coefficient of variation: an estimate of the amount of error present in a measurement.

📖 **Standard score**: a measure that has been standardized in relation to a distribution's mean and standard deviation.

📖 **Statistics**: a science that uses mathematical methods to collect, organize, summarize, and analyze data.

📖 **Statistical significance**: a sample measurement that reflects a real population value rather than a sampling error.

📖 **Stratified sample**: sample selected after the population has been divided into categories.

📖 **Structured interview**: an interview in which standardized questions are asked in a predetermined order.

📖 **Summary statistics**: statistics that summarize a great deal of numerical information about a distribution, such as the mean and the standard deviation.

📖 **Summed ratings method**: a method of attitude scaling that uses item analysis to select the best items; developed by Rensis Liekrt.

📖 **Systematic random sampling**: a procedure to select every n^{th} member of the population for a study, such as every 10th person in a telephone directory.

📖 **Systematic variance**: a regular increase or decrease in all scores or data in a research study by a known factor.

📖 **Telephone survey**: a research method in which survey data are collected over the telephone by trained interviewers who ask questions and record responses.

📖 **Test-retest reliability**: the degree of consistency of a test or measurement, or the characteristic it is designed to measure, from one administration to another.

📖 **Theory**: a set of related propositions that presents a systematic view of phenomena by specifying relationships among concepts.

📖 ***Thurstone* attitude scale**: the Equal-appearing intervals method.

📖 **Total survey area (TSA)**: a region in which an audience survey is conducted.

📖 **Trend study**: a longitudinal study in which a topic is restudied using different groups of respondents.

📖 **Triangulation**: using a combined quantitative and qualitative approach to solve a problem.

📖 ***t*-test**: a statistic used to determine whether the difference between group means is statistically significance.

📖 **Type I error**: rejection of the null hypothesis when it should be accepted.

📖 **Type II error**: acceptance of the null hypothesis when it should be rejected.

📖 **Unstructured interview**: an interview in which the interviewer asks broad and general questions but retains control over the discussion.

📖 **Validity**: the degree to which a test actually measures what it purports to measure.

📖 **Variable**: a phenomenon or event that can be measured or manipulated.

📖 **Variance**: a mathematical index of the degree to which scores deviate from the mean.

📖 **Volunteer sample**: a group of people who go out of their way to participate in a survey or experiment (for example, by responding to a newspaper advertisement).

📖 **Volunteer bias**: systematic error resulting when participants who volunteer respond differently from how individuals in the general population would respond.

📖

Weighting: a mathematical procedure used to adjust a sample to meet the characteristics of a given population; also called sample balancing.