Health Surveys and Social Science: A Primer for Applied Survey Projects

February 20, 1998

Lois M. Haggard, Ph.D.
Bureau of Surveillance and Analysis
Utah Department of Health

Preface

In this document I have compiled information about social science research methods and survey research methods. It is meant to be used by professionals who find themselves in a position to use applied social science, such as a survey or some similar type of social science measurement tool. It is assumed that the reader will be employing a contractor for the actual sampling and data collection process, as there are many details on these topics that are beyond the scope of this document.

Survey research is extremely labor-intensive, and therefore expensive. As a result, it is rare that the available resources will allow the best possible methodology to be used. In this context, what separates 'good' research from 'bad' research is primarily the decisions and trade-offs that are made along the way about how those resources will be applied. It is my hope that the information in this document will help the reader to move more confidently through the many decisions that will confront them in the conduct of applied social science.

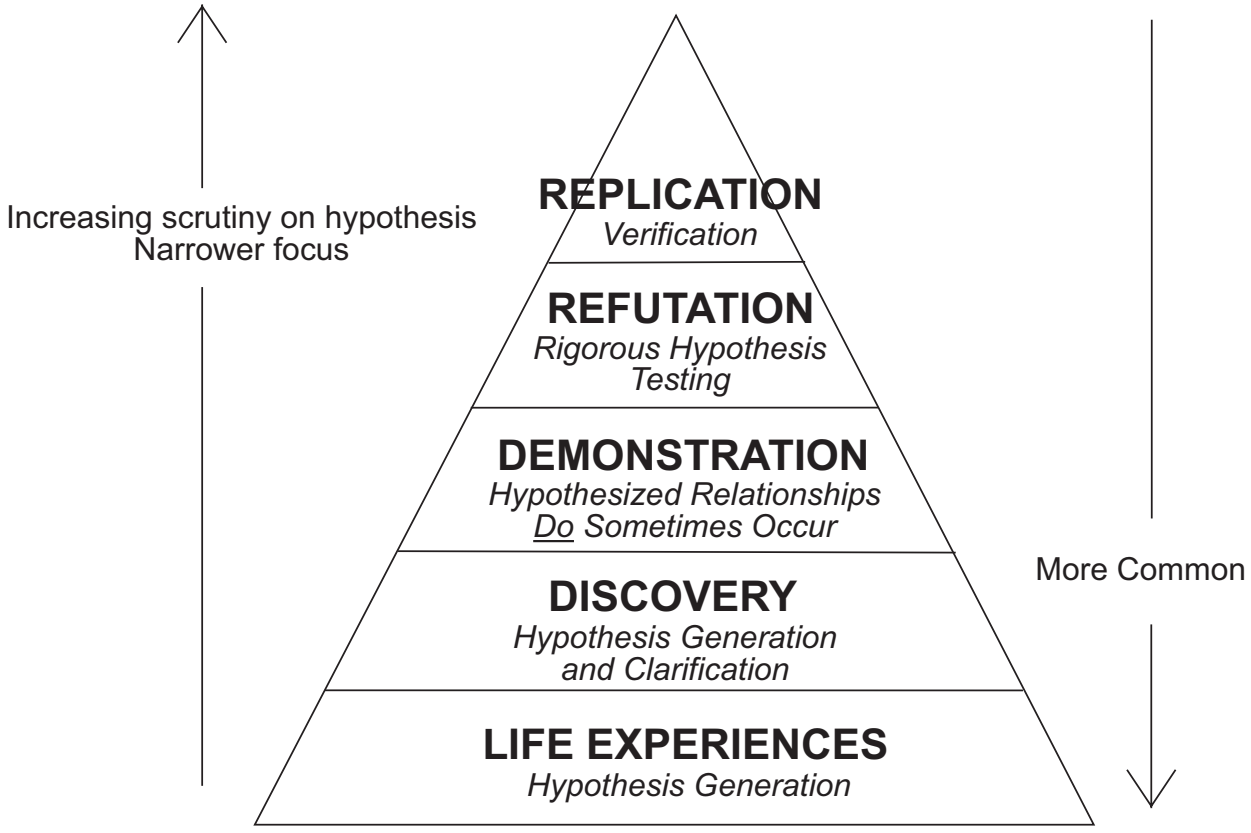
I'm sure this attempt will probably fall short of it's mark, as survey research methods is an extremely broad topic in itself. At the very least, perhaps, it should provide a framework that can be used to categorize and understand new knowledge as it is gained in future experiences.

In writing this document, I have relied on other written material, in some places quite extensively. One textbook in particular has been quite useful. It is Charles M. Judd, et al., (1991). Research Methods in Social Relations, 6th Ed. Fort Worth: Holt, Rinehart and Winston, Inc. It is paraphrased extensively in sections I. and II. Stephen B. Hulley and Steven R. Cummings (1988) Designing Clinical Research: An Epidemiologic Approach. Baltimore: Williams and Wilkins, was also helpful in conceptualizing measurement error.

Table of Contents

Section I. The Big Picture: An Overview of the Nature of Scientific Inquiry on Social Issues

The Functions of Social Research



Informal Observation (Life Experiences)

Our life experiences are the basis for our naive hypotheses and theories. A naive hypothesis is one that has not been critically examined using scientific methods and systematic observation to support and modify it.

Research for Purposes of Discovery

More systematic than informal observation, the discovery process has the following characteristics:

- It is an **inductive** process. Gather the details and try to detect patterns and themes.
- It is meant to **broaden the researcher's perspective** and to remove any "blinders" the researcher may have because of his or her lack of personal experience.
- Phenomena are **systematically** observed.
- Observations **inform and refine** existing hypotheses, and **generate** new hypotheses. The discovery process provides a "reality check" for the researcher's naive hypotheses.

- Research conducted for this function can be either **quantitative or qualitative**.
- This process frequently uses **descriptive statistics**, although they are not necessary to the process. Any use of inferential statistics is purely exploratory.

Researchers who skip this process are in danger of producing theories that are biased and out of touch with the "real world" (and perhaps the "truth"). Better to be humble and open to the fact that there are equally viable alternative ways of viewing the issue, and that you may want to explore what these are. This may be especially relevant for cross-cultural.

Bonus Factoid: Inferential Statistics

Scientists make a distinction between descriptive and inferential statistics. Descriptive statistics involves computation of a value that tells us something about some phenomenon, such as the lap time in a race, or the percentage of male students in a class.

Inferential statistics try to make an inference about the value of some statistic in a population, based on the value of that statistic in a sample. Probability theory is used to suggest the probability that the sample accurately represents the value in the population.

Research for Purposes of Demonstration

Attempts to demonstrate that hypothesized relationships between variables truly exist and can be observed in at least some populations and at some points in time. The demonstration process has the following characteristics:

- It is a **deductive** process. The researcher structures the data gathering with a purpose in mind.
- Phenomena and relationships are **systematically** observed and measured. Measurement becomes more standardized, less flexible and responsive to individual participants.
- Typically uses **quantitative** measures.
- Observations **inform and refine** existing hypotheses, and **generate** new hypotheses.
- This process frequently uses **inferential statistics**. Inferential statistics go beyond describing the sample, and make inferences about population characteristics.
- **Representative sampling** is very desirable. Representativeness makes for a more credible demonstration.

Although relationships between variables are observed and measured, cause cannot be implied. "**Association does not imply causation.**"

Research for Purposes of Refutation

The scientific method requires that hypotheses be refutable. A theory or hypothesis cannot be "proved," it can only be tested in situations where refutation is possible. You may observe only white swans, but that does not prove there are no black swans. You may look at 100 million swans -- and if they are all white, you still cannot say that all swans are white. The black one may still be out there somewhere. The refutation process has the following characteristics:

- It is a **deductive** process.
- Observations **test existing hypotheses** about the **cause and effect** nature of theorized relationships.
- It is meant to put an hypothesis to a **sensitive test** -- one where refutation is possible.
- Hypothesis testing uses confusing double negatives, such as "failed to reject the null hypothesis," in order to communicate precisely what was gained by the research.
- **Measurement is tightly controlled** in an attempt to rule out all possibility of external influences that can later lead to alternative explanations for the findings.
- This process relies almost exclusively on **inferential statistics**.
- This process is resource-intensive. As a result, it tends to have a **narrower focus**. Generally testing one or a few hypotheses at a time.

Bonus Factoid: Four criteria for a causal relationship:

1. Independent variable (IV) must precede dependent variable (DV) in time.

2. Covariation must be demonstrated (usually statistical association).

3. Nonspuriousness (rival hypotheses, alternative explanations must be ruled out).

4. Plausible theoretical explanation (preferably parsimonious and *a priori* -- that is, before examining the data).

Research for Purposes of Replication

The replication process is generally one in which a previously tested hypothesis is subjected to further testing, either in a different study population, in different settings, or by different researchers. A theory that has withstood attempts to replicate it broadly is termed a "robust" theory. Although replication increases our confidence in the theory, it still does not prove it. No matter how many people fail to see a black swan, the lack of existence of a black swan is never fully proved.

Section II. The Nature of Social Theories and Hypotheses

All research, whether basic or applied, relies on theories and hypotheses.

Any research project will be more productive and informative to the extent that the implicit theories and hypotheses are made explicit.

What is a theory?

A social theory is a set of interrelated logical propositions that attempt to explain social phenomena. A social theory:

- Contains social constructs of interest, and attempts to account for or explain them in some way.
- Describes relations between these constructs. Often these relationships are causal. These hypothesized relations are part of a theory.
- Specifies observable (measurable) indicators that can be used to conduct empirical research.

Example: Prochaska's Transtheoretical Model of Behavior Change.

Individuals do not go directly from old behaviors to new behaviors, but progress through a sequence of stages: precontemplation, contemplation, preparation, action, and maintenance. Progression through the stages is often not linear, because many individuals regress and cycle back through the earlier stages. Individuals may recycle through the stages several times before they succeed in their efforts to change (Grimley, et al., 1995). Progression from one stage to the next depends on the decisional balance of that progression. The decisional balance is the net effect of the individual's perceived gains versus his or her perceived costs or losses from performing the behavior. Gains and losses may take the form of utilitarian gains and losses, approval or disapproval from significant others, and self-approval or self-disapproval (Prochaska et al., 1994). Furthermore, there are predictable processes, such as consciousness-raising, self-liberation, helping relationships, and reinforcement management, that assist in developing and maintaining desired health behaviors. These processes are variably dominant depending on the individual's stage of change. For instance, the precontemplation stage involves primarily consciousness-raising, while the maintenance stage involves counter conditioning, helping relationships, reinforcement management, and stimulus control.

What is an hypothesis?

An hypothesis is a statement or a proposition about the relationship between two

variables.

The most productive social theories...

- offer plausible explanations for relevant and socially significant issues,
- are consistent with existing theories on the topic, and
- provide new insights or shed new light on previously unforeseen implications.

Example: stimulus control and cigarette smoking

> Persons who successfully practice stimulus control will be better able to stay quit from cigarette smoking compared with those who do not practice stimulus control.

Hypotheses are not generally needed in descriptive studies, that is, studies that describe merely how characteristics are distributed in a population. However, it can often be a useful exercise to identify and examine any expectations you may have for what you will find.

What is a construct?

Constructs are generally rich theoretical concepts or ideas. Examples include social status, power, health, self-efficacy, verbal ability, stimulus control, stage of change, etc. They are abstractions that help us think about, categorize, and communicate ideas about social phenomena. Constructs are the focus of social scientific thought and theory.

Example: constructs used in stimulus control and cigarette smoking hypothesis

successful practice

stimulus control

cigarette smoking

What is a variable?

A variable is one way to represent a construct. It is not the same as a construct -- one construct may be represented by more than one variable. For instance, the construct of "anxiety" may be measured by increase in heart rate, blood pressure and skin conductance, and by a person's verbal reports of his or her experience ("I feel nervous." "I'm afraid."). We work with variables because they are more concrete than constructs, and suggest ways we may measure the presence or absence of a construct in a social situation. The steps we take to define measurement of a variable are called operational definitions.

Example: variables used to measure stimulus control

Construct: stimulus control

Variables: - keeping cigarettes out of the house

- avoiding situations (such as bars and parties) where smoking cues are present or cigarettes are available

- asking others to not smoke around you

What is an operational definition?

An operational definition specifies how to measure a variable so that a quantitative score may be assigned to a person or a situation. For instance, creating a one-to five scale is a means of operationalizing, or creating an operational definition for, a variable. Other examples of operationalizations include reading a thermometer to measure temperature, using a standard test to measure verbal ability, using a survey questionnaire to measure attitudes, or using an APGAR score to indicate a newborn's general health.

Example:

Variables:

- *keeping cigarettes out of the house*

  On how many of the past 30 days did you have cigarettes available to you at home? (quantified as a number from 0 to 30)

  How much effort did you put into keeping cigarettes out of sight, out of reach, or otherwise unavailable to you during the past 30 days? Would you say a lot of effort, a fair among of effort, some effort, or no effort? (quantified as a number from 1 to 4)

- *avoiding situations (such as bars and parties) where smoking cues are present or cigarettes are available*

  On how many of the past 30 days were you in situations outside your home where others were smoking cigarettes? (quantified as a number from 0 to 30)

  Most people who are trying to quit smoking have certain situations where the urge to smoke is particularly strong. What is the one situation that is most powerful at triggering your desire to smoke a cigarette? (open ended)

  On how many of the past 30 days were you in (the situation described above)? (quantified as a number from 0 to 30)

  During the past 30 days, how much effort did you put into avoiding situations that trigger a strong desire to smoke? Would you say a lot of effort, a fair amount of effort, some effort, or no effort? (quantified as a number from 1 to 4)

- *asking others to not smoke around you*

  Do you have any friends, relatives, or family members who smoke cigarettes? (yes, no)

  [If Yes] Have you asked them to not smoke while they are with you? (yes, no)

  [If Yes] During the past 30 days, did any of them smoke cigarettes while they were with you? (yes, no)

[If Yes] On how many occasions during the past 30 days did this happen? (quantified as a number from 0 to 97)

Bonus Factoid: Empirical Research:

A measurement taken of the current value of some variable is called an *observation*. Any research that relies on measurement and observation is referred to as *empirical research*.

Sometimes a scientist will say, 'It's an empirical question. By this, she means that the question can be answered or informed by measurement or observation of real-world phenomena, as opposed to philosophical questions, rhetorical questions, or theological questions, for instance, which are informed by logical or faith-based dialogue.

Section III. Sampling Basics

Good sample design requires conceptualizing a series of sample definitions that maximize the representativeness, or applicability, of the study and minimize implementation costs and errors. These decisions are typically expressed as a set of definitions (below), each with a slightly different emphasis.

- Population of Inference. What is the population of interest? All Utahns? All teenagers in Salt Lake County? All nursing homes?
- Sampling Frame. What criteria will define inclusion on the sampling list? In a statewide telephone survey of adults in Utah, the sampling frame typically includes all households with telephones. The frame could also be an existing list.
- Sampling Pool. Which units (households, persons) will actually have the potential of being selected? For instance, the pool of phone numbers generated by random digit dialing in phone surveys.
- Sample. Which members of the sampling pool actually participated in the survey? e.g., all respondents who completed the data collection instrument.

Example:

In a (hypothetical) study of current clinical practices among Medicaid physicians in Utah, the **population of inference** was defined as all Utah physicians who deliver care to Medicaid patients.

The **sampling frame** was the list of physicians who had received payments from the Utah Medicaid program in the past six months. This list was generated from the accounts payable records of the Medicaid program.

The Medicaid program knew that they did not have to contact all the physicians on the list in order to get good estimates of current clinical practices. After all, sampling is, by definition, the practice of selecting a subset of the population for measurement purposes. The program randomly selected a **sampling pool** of 300 Utah physicians from the sampling frame. The object was to contact as many of those physicians as possible.

Near the end of the data collection period, the survey firm indicated that they had contacted all the physicians who were easy to reach and willing to participate. It seemed as though all the physicians left in the sampling pool whom they were still trying to contact were either never in their offices, or down-right surly about doing a survey. But they had contacted only 50% of the physicians in the sampling pool. The Medicaid program was not satisfied, as a 50% response rate would yield a **sample** of only 150 physicians. Not only was the sample small, but also it was probably not representative of all Medicaid physicians in Utah (busy and surly docs might have different clinical practices than accessible and kindly docs). They negotiated an

agreement with the data collection firm to contact a **sample** of at least 200 physicians out of the 300 who were originally in the sampling pool. A 67% response rate was less than they had originally hoped for, but they also knew the difficulties involved in completing surveys with practicing physicians.

Random selection is another sampling concept that requires mention. Random selection does not mean haphazard selection. One way to randomly select persons from a list is to generate a list of random numbers (your spreadsheet program can do this) that is as long as your list of persons. Then match each person with a random number (the first person with the first number on the list, the second with the second, etc.). Now sort the list in the order of the random number (ascending or descending order of the numbers). If you start selecting from the top of the list, and select sequentially down the list, you will be selecting a random sample of persons from the list.

Systematic sampling is another familiar process in which the researcher selects every n^{th} (e.g., every 10^{th}) member of the list. This is not random sampling. But it is acceptable if the list does not include any cyclical order (e.g., if you started with Monday, and picked every seventh day from now until Christmas, your sample would include only Mondays). Systematic sampling is improved if the first selection on the list is made randomly (i.e., select your first member by applying the above random selection process to the first n^{th} members on the list) and the remaining members are selected systematically.

Section IV. Qualitative and Quantitative Social Research

Qualitative Research

Qualitative research relies almost entirely on open-ended explorations of people's words and thoughts; it can also deal with actions. It can take on many forms. Some common forms are key informant interviews, focus groups, participant observation, ethnography, and critical incident reports. Field workers enter the worlds of the people they study instead of bringing those people to a laboratory or asking them to answer a structured interview or questionnaire. Although it is systematic in many ways, the field worker is flexible, and is guided by the viewpoints of those being studied. The investigator is attentive to the larger picture rather than the details, and must place any information in its appropriate context in order to represent it accurately. Finally, the researcher is nonjudgmental, and must avoid his or her biases or make them explicit so that their probable effects may be assessed. The objective is to gather rich, contextual information in the voices of the participants to gain an understanding from their points of view.

Quantitative Research

Quantitative research is structured so that a participant's thoughts and experiences can be given a numeric value. In order to do this, the researcher must narrow the focus of the investigation. It is more focused on details rather than on the big picture. The narrow focus invariably imposes researcher biases onto the phenomenon being studied. That is, you don't get a report of the whole picture, just that portion of the picture that the researcher has deemed relevant. Quantitative research generally tries to gather data from a sample that is rather rigidly representative of some population of inference. It generally uses some form of structured interview, such as a paper-and-pencil questionnaire or a telephone interview, although it can also take other forms, such as behavior coding, or applying numeric values to aspects of events. The objective is to gather specific information to make precise and accurate statements about a construct or set of constructs in a population.

A Comparison of Selected Features of Qualitative and Quantitative Research Methods

	Qualitative Methods	Quantitative Methods
Research Function	Typically applied to the discovery process in social research.	Can be applied to any of the functions of social research.
Sampling	Broad cross-section of population is usually desirable. Representative probability samples are not used.	Typically a representative probability sample is desirable.
Sample Size	Generally smaller, often fewer than 100 observations.	Generally larger to increase sampling precision (decrease sampling error), often 600 or more.
Data Collection Strategies	Flexible, responsive to context, topic directed by researcher, but content largely directed by the participant.	Researcher imposes structure and scope. Variation from other than the narrow focus of interest is intentionally minimized.
Data Collection Mechanisms	Key-informant interviews, focus groups, participant observation, field notes	Survey interviews, questionnaires, standardized instruments
Nature of Data Collected	Largely verbal. "Participants' voices" produce rich and varied information.	Numeric at collection, or converted to numeric later, produces relatively sterile and focused information.
Pace of Interview	Slower, fewer questions, more time spent fleshing out various aspects of each question	Shorter and faster-paced, more focused interviews, much less time spent on each question.
Data Analysis	Identify categories or themes	Create computerized data set and analyze with computer software, often using statistical procedures.
Data Reporting	Prose, quotations, verbal	Counts, percentages, and other statistics

V. Developing a Research Question

Whether a project seeks to conduct basic or applied research[1] it starts with a research question that describes what the researcher would like to know. Toward the beginning of a project, the research question is rather fluid (and is often a little too vague). As the project progresses, the research question changes to accommodate new information, narrow the scope of the project, become more specific or concrete, or to improve in other ways. As a result of this process of improvement, the research question becomes less fluid and more stable over time.

The following steps are activities that should take place in the development of a research question and study plan. These steps should take place regardless of whether the research is being conducted for the purposes of discovery, demonstration, refutation, or replication. Note that many studies have multiple research questions, and the activities outlined below pertain to a single research question.

Also remember to get input from your peers at each step along the way. Collegiality, or the sharing of information and mutual critiquing of work, is critical to productive research. We are sometimes too shy to ask, or afraid of negative feedback. Hopefully you will have sympathetic and clear-thinking individuals near at hand with which to discuss your ideas as they develop and take shape.

Steps in Developing a Research Question

1. Use one or two sentences to state the *general problem or issue* you would like to address with the study. You may also want to include any reasons you may have for addressing the problem or issue. This often helps a research group to stay focused on the primary issue at hand.
2. Again, in one or two sentences, list just one *question* you have with regard to the problem or social issue. What is it you want to know?
3. Check the research literature for written information on the topic. Are there published articles or books that offer relevant perspectives or study methodologies?
4. In your scanning of the literature, do you find existing theories, models, or approaches that can help to structure your thinking about the issue? Do you agree or disagree with the existing literature? Don't assume that everyone knows more about your topic than you do. Be discriminating and skeptical. If a theory or approach doesn't make sense to you, there's probably a reason for it.
5. Do you have any hypotheses with regard to this question? Remember, you don't really need hypotheses when your research project is for purposes of discovery. But even with

[1] Basic research is conducted for the purposes of advancing knowledge in a particular field, without an immediate application. It derives from a desire to better develop or refine a scientific theory. Applied research is conducted to better understand a phenomenon in a population, without a particular desire to advance knowledge in a field. It derives from a desire to inform relatively immediate decisions or actions.

a project for discovery purposes you may have ideas about what you expect you'll find with respect to the research question. Or perhaps you want to document the level of some variable. Do you have any expectations about what this level will be? Or you may expect to find that two variables are related. For instance, younger drivers are more likely to be involved in motor vehicle accidents than older drivers (motor vehicle accidents are related to age).

6. Identify the *constructs* you have employed in stating your hypotheses. In the example just above, the constructs employed were 'motor vehicle accidents' and 'age.'

7. What is your *population of inference*? All persons in the U.S.? Persons with valid driver's licenses? Licensed drivers in Utah?

8. Now that you have identified your constructs and your population of inference, you can provide *operational definitions* for each of the constructs. For instance, a motor vehicle accident may be defined as an accident that was reported to the police. Or you may only want to look at accidents where someone was injured. You may also want to measure a construct with more than one variable. For instance, you could get self-reports from people about their involvement in accidents. This would include accidents not reported to the police. Then you may also want to get police records as another view of the construct. Age is typically defined as a person's age on their most recent birthday.

9. What *sampling frame* do you propose to use? The sampling frame is merely the method of contacting the study participants (see Section III). It will depend on both the population of inference and the operationalization of your variables. For instance, one sampling frame for licensed drivers in Utah is the driver's license division. Or, if your study is more interested in the property damage aspects of motor vehicle accidents, you may choose the motor vehicle registration records as a sampling frame.

10. Develop a one to two page study plan that specifically describes how the participants will be selected and how the variables will be measured.

At this point, the research team members may start designing the data collection instrument, such as a survey questionnaire. As the study plan develops, the process of development involves getting further feedback, gaining better familiarity with the research literature, revising your study question and hypotheses, pretesting, pilot testing, and so forth until the data collection begins. Changes in the research question after data collection begins are costly in terms of staff time and money -- so make sure you cover all your bases before you begin data collection.

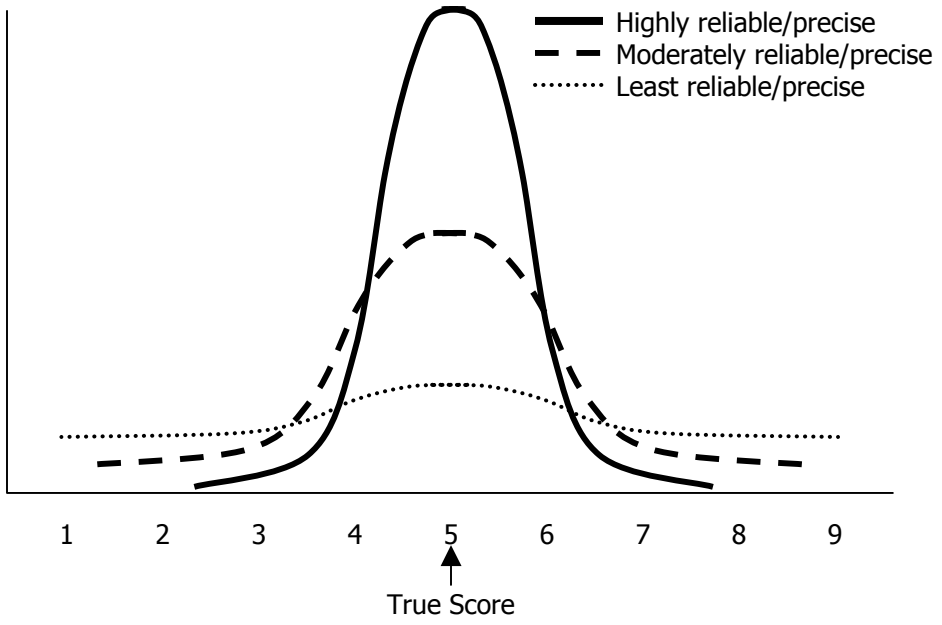
Section VI. Measurement Quality: Precision and Accuracy

	Random Error / Precision	Systematic Error / Accuracy
At the individual level	Lack of reliability in measurement of construct	Bias, lack of measurement validity
At the aggregate level	Small sample lacks precision in estimate of some variable in the population	Bias, lack of external validity, or ability to accurately generalize to the population of inference

Precision

Measurement Reliability. At the individual level, there is always a certain amount of noise in any measurement. Some measurements have more noise than others. A measurement that is relatively free of random error is said to be reliable. A reliable measure will obtain the same result over repeated observations. True score theory is used to demonstrate the constructs of precision and accuracy. A person's 'true' score is the score that should be obtained by a measurement tool if that tool is totally accurate and precise -- it is a score with no error.

Distribution of scores that could be obtained using three different measures over multiple administrations of each measure for the same individual, given no change in his or her true score.



Sampling Precision. Precision at the aggregate level refers to the extent to which the value of a measure gathered from a sample represents the true value in the population. A sample's precision is good to the extent that the sampling error is small. In order for the sampling error to be small, the sample size must either be large, or be close to the population size (i.e., a sample of almost all members of a population).

The natural dispersion or variance of the measure in the population also is considered in calculating the sampling error. Sampling error can be calculated, or estimated using a table such as the one below. Note that the sampling error refers to the sampling error for a single item, not the entire survey. Often, researchers will report the survey's sampling error as the sampling error that would be obtained for a 50/50 percentage distribution, given the sample size. Sampling error is typically reported as 'plus or minus __%.'

Estimated Sampling Errors for Percentages for Various Sample Sizes and Percentage Distributions for Samples Drawn from Infinite Populations

	Percentage Distribution				
Sample Size	50/50	60/40	70/30	80/20	90/10
100	10	9.8	9.2	8	6
200	7.1	6.9	6.5	5.7	4.2
300	5.8	5.7	5.3	4.6	3.5
400	5	4.9	4.6	4	3
500	4.5	4.4	4.1	3.6	2.7
600	4.1	4	3.7	3.3	2.4
700	3.8	3.7	3.5	3	2.3
800	3.5	3.5	3.2	2.8	2.1
900	3.3	3.3	3.1	2.7	2.
1,000	3.2	3.1	2.9	2.5	1.9
1,100	3	3	2.8	2.4	1.8

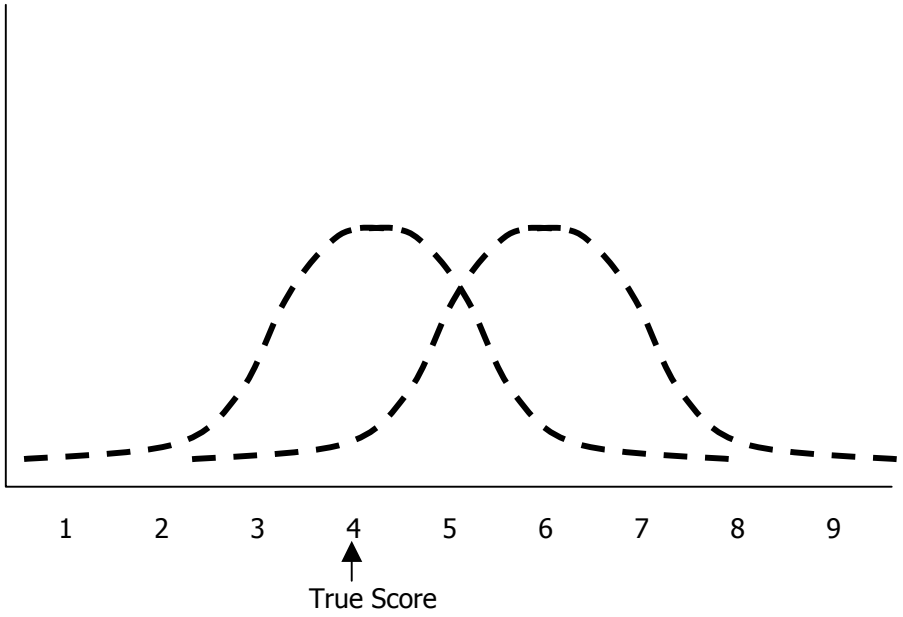
Bonus Factoid: 'So <u>that's</u> what they mean by "margin of error"?

- Sample size and the amount of variation or total dispersion in a score are the only pieces of information considered in the commonly cited "margin of error."
- The correct term, 'sampling error,' is more descriptive and less misleading.
- Sampling error is only a small portion of the overall error that could be present in a given empirical observation. Other sources of error are rarely if ever cited because they are not easily quantifiable. A good researcher will acknowledge these potential sources of error in his or her research measures.

<u>Accuracy</u>

<u>Measurement Bias</u>. At the individual level, a measure is not accurate to the extent that there is bias in the measure. A biased measure is one that is consistently, or systematically, off the mark. Like a bathroom scale with a weak spring, it will consistently give a measurement that deviates from the true score in the same direction. Biased measures may either be reliable or unreliable. A measure can also be invalid if it is measuring some construct other than the one it is intended to measure.

Distribution of scores that could be obtained using two different measures, over multiple administrations of each measure, for the same individual, given no change in his or her true score.



Sampling Bias. Accuracy at the aggregate level is related to bias in the way the sample was selected. This is also sometimes called external validity, and affects the researcher's ability to accurately generalize the sample data to the population of inference.

Lack of external validity will also occur when the study sample has not been randomly selected from the population of inference. If random selection is not used to select the sample members, there will always be bias in the sample. This is true even if the sample demographics match the population demographics. You cannot always assess the nature or extent of the bias, so it is important to use random selection procedures if you want a representative sample.

Sources of Error in Surveys

	Random Error (threatens precision)	Bias/Systematic Error (threatens accuracy)
Coverage	respondent happened to be on vacation when the interviewer called	people with telephones generally have more money than those without
Nonresponse	respondent's dinner guests just arrived 30 minutes early, had to end the interview	self selection: slothful respondent refused to participate in exercise survey
Sampling	small sample size leads to large sampling error	interviewer conducts survey in shopping mall with respondents who "look willing"
Interviewer	interviewer skipped lunch one day and is grumpy on the phone, short on social skills	interviewers 1 and 2 consistently emphasize different words in the same question
Respondent	one respondent skipped lunch and is grumpy on the phone, rates all candidates negatively, another had a two martini lunch, was more generous	respondent embarrassed to report his or her actual weight, reports that it is lower than actual
Instrument	vaguely-worded questions have different meanings to different respondents	a question about tax burden is placed immediately before an open-ended question asking, "What is the most important problem facing Utah today?"
Mode	rare	a personal interview about drug abuse expects high school students to talk openly about drug use with interviewers who look like their mothers

Adapted from: Groves, Robert M. (1989) Survey Errors and Survey Costs Beverly Hills: Sage Publications, p. 17.

A Closer Look at Measurement Issues in Survey Research

The Respondent as a Source of Measurement Error

In survey research, the closest element to the truth is the respondent him or herself. However, contrary to what one might think, even the respondent is limited in his or her ability to know, and to express the answer.

Social Desirability. "Have you ever shoplifted anything?" Even the most honest of respondents would be tempted to answer this question with the socially desirable response, "No," regardless of the true answer to the question. There are specific ways that this question may be asked in order to put the respondent at ease, engender respect for accurate data, and avoid socially desirable responses. There are also measurement scales that may be added to a questionnaire (such as the Crowne-Marlowe Social Desirability Scale) in cases where social desirability cannot be avoided. Such a scale does not prohibit socially desirable responding, but it allows the researcher to measure the presence of it.

Response Acquiescence (yea-saying) and Nay-Saying. Sometimes questions must be worded in ways that make it difficult for a respondent to answer "yes" or "no" throughout an entire questionnaire and still report logically consistent answers to the survey questions. The format of the response options may also alternate so that a "yes" answer or a high number has a different meaning on some questions than on others.

Intentional Misleading. "Could you afford to pay an additional $50 in tuition next semester?" A good survey designer would not expect a truthful answer to this question. Some respondents may give false information as a joke, or in an effort to purposely foil a sponsor's attempt to measure a variable. Price sensitivity is a common example in which respondents often try to bias the results of a survey.

Motivation to Answer the Questions. Some respondents may agree to do the survey, but then lose interest and end up providing answers that are inconsistent or not well thought out. A well-designed survey instrument will provide motivation for the respondent, and keep the respondent's interest along the way.

Lack of Cognitive Access. Even if a respondent is motivated, he or she may not have ready access to the information that he or she is being asked to provide.

- Memory - length of recall and salience of event. For instance, a survey that inquires as to the health of all household members over the past year may easily elicit major salient health events, such as surgery or extended illness, but it may miss many health events, such as colds and hay fever, that do not spring readily to the respondent's mind. Alternative strategies include asking only about the past week, or what the current status is, or administering a panel survey in which the same respondents are called at regular intervals over the course of an entire year.
- Degree of detail and complexity of the task. This is perhaps the most common mistake made

in inquiries of a public policy nature. Take, for instance, "Do you support or oppose proposition 3?" This question may assume more knowledge than the respondent has. A good survey researcher might explain proposition 3, and then ask the question, or ask the question on a different level. He or she might want to anticipate how proposition 3 might affect people's day-to-day lives, and then ask whether the respondents support or oppose these effects?

- Respondent selection. Choosing the right respondent is also important. You want to select a respondent who will be likely to have the information you want to collect. Some surveys will require a nonrandom selection procedure. For instance, the survey design may include procedures to ascertain who in the household is the most knowledgeable on a certain topic. Others may require the selection of a named individual who is the member of a certain group (club members, license holders, etc.).
- Creating an answer that did not exist prior to the question. For instance, "How do you feel about left-handed widgets?" The respondent may never have heard of left-handed widgets, but instead of exposing his or her ignorance, the respondent will often just give an answer. A less threatening screening question (e.g., "Do you own a left-handed widget?") is often a good way to ensure that all the data for a particular question are valid.

Random and situational events. Hunger, fatigue, misunderstanding the question, etc. A good survey organization makes it easy for a respondent to reschedule an interview at a more convenient time. This is more costly than coercing the respondent to complete the interview at the time of initial contact, but it provides for better quality data, and more ethical treatment of human subjects.

Interviewer Bias

The bottom line in minimizing the interviewer's contribution to survey measurement error is to standardize all interviewer behaviors. Interviewers must be carefully trained to behave identically from one respondent to the next, and from one interviewer to the next. For instance, interviewers must provide identical clarifying remarks in answering respondents' questions, give identical probes for more complete information, and select a respondent within a household using identical protocols. A good survey research organization will have strict protocols for every survey that minimize the impact an interviewer may have on data. Interviewers are supervised at all times to ensure that interviewer protocols are being followed as they were intended.

Respondent Selection and Eliciting Cooperation. The interviewer should have no control over respondent selection, other than to follow a selection protocol designed by the researcher in conjunction with the client. Selection should generally be random, or some other representative design. Allowing interviewers to have control over respondent selection interferes with this goal.

- Nonrandom selection, especially difficult to control with face-to-face surveys in public places, e.g., what if the fifth person to exit the door after 12:30 PM is a big burly, mean-looking guy in a black leather jacket? Interviewers must have adequate training to

understand the need for adherence to the protocol. The survey organization must also be able to anticipate the potential for awkward situations, and provide adequate training for interviewers so that they are safe, comfortable, and quality-oriented at all times.

- Eliciting cooperation must be standardized so that every interviewer behaves identically. Differential presentation of the request, giving up too easily, failing to convert refusals, etc., all result in biased survey data.

Data Collection, Data Recording. Some errors occur at the time the data are being recorded.

- Interviewer characteristics. Some interviewers may have mannerisms of speech (e.g., foreign accent, quick speaking), while others may be uncomfortable talking about sensitive topics. A good survey manager must be prepared to hire interviewers on their phone skills, terminate interviewers who do not perform to the standard, and to assign interviewers to data collection projects in such a way as to provide a good fit between the interviewer and the project.
- Variable question administration.
  - Interviewers should read the entire question, as worded -- including all response options (if response options are to be read), even if the respondent answers before the interviewer has finished reading all of them. This skill is apparently a difficult one, as it is one of the first skills to fade over time, and must be constantly renewed.
  - Rewording and paraphrasing of survey questions. It is impossible to write a question that is perfect for everyone. This is true for respondents and interviewers alike. Untrained interviewers will reword questions to suit their own understanding and style. This has disastrous consequences -- especially for surveys that are replications of other survey administrations. This is another crucial reason for fully supervised interviews. The survey lab should have a tap-in phone in another room that allows the data collection supervisor to listen-in unobtrusively to any call being made in the calling facility.
  - Paraverbal communication. Paraverbal communication refers to all that is communicated with the voice that is not part of the spoken language. One example of this is the emphasis an interviewer may place on key words in a survey question. For instance,

    "Of all the MAJOR luxury items you have bought in the past five years, which have you been most pleased with?"

    is likely to elicit a different answer than the question,

    "Of all the major luxury items YOU have bought in the past five years, which have you been most pleased with?"
  - Neutral probes. Survey respondents do not always give complete answers, or answers that meet the objectives of the survey question. Interviewers need to be trained to respond with a neutral, or nondirective "probe." A nondirective probe for a numerically answered question might be to read the entire list of possible answers again for the respondent (certainly not a truncated list). A nondirective probe for an open-ended question might be, "How do you mean that?" or, "Tell me more about that?" Examples of directive (poor) probes include, "So, do you strongly disagree?"

and, "So, does that mean your answer is 'zero'?"

- Helping a respondent to understand the question. Interviewers must have standardized answers to any questions the researcher can anticipate. If they do not have a standard instruction, they should not attempt to explain the meaning of a question. "Whatever that means to you," is the only appropriate response an interviewer should give to a respondent where no instructions have been given.

- Recording the answer correctly. This is especially tricky for answers to open-ended questions. Interviewers must first record the answer verbatim, without paraphrasing the answer in their own terms. Next, if the question asks for the "primary reason," or the "most important," or the "first thing that comes to mind," the interviewer must record only one answer. If all reasons or things are asked for in the survey question, then the interviewer must enter all answers, in the order they were given.

Interviewer Expectations. If an interviewer expects a certain answer to a question, s/he will be more likely to receive it. The researcher must be careful to keep potentially biasing information from the interviewing staff in some instances.

Interviewer Supervision. It is absolutely critical that a survey research organization provides a field supervisor to tap-in on interviews unobtrusively, and ensures that survey interviewers are using appropriate interviewing techniques. Survey interviewers will get lazy, and should never be allowed to perform interviews from their own homes, unless some tap-in system can be arranged. Interviewer training is on going, and a system of providing feedback to individual interviewers on their job performance is essential.

The Data Collection Instrument: Construction of the Questionnaire

Perhaps the most difficult source of error to define is the survey questionnaire itself. Writing rules for questionnaire construction is like writing the rules of English grammar. There are certain rules, but every rule has an exception that may be appropriately applied in certain circumstances. Unfortunately, there is no substitute for experience.

Language and Properties of Words. Virtually every survey uses verbal communication. Care should be taken to avoid ambiguous words, words that would commonly be misunderstood, jargon, unclear response categories, and other forms of imprecision. For instance, "for how long?" is less precise than "for how many years?"

Question Structure.

- Length of the question. Normal reading speed is faster than reading aloud, and retention of written material is better than it is for spoken material. Many telephone surveys use questions that are too long for respondents to understand over the phone.

- Open- versus closed-ended questions. There are inherent trade-offs between open- and closed-ended questions. It is often more precise and easier to analyze "closed" response categories (pre-coded, numeric responses). Sometimes, however, it is difficult to anticipate how survey answers will need to be categorized. In these cases, an open-ended question

format could be appropriate. Open-ended questions allow respondents to answer in their own words. Although these questions provide very rich and interesting data, they require much more time and expense to code and analyze.

Three common problems with questionnaire items

1. Double-Barreled Questions. Make sure each question addresses only one construct.

   Bad: In general, would you say that reports are accurate and on time?

   Better: In general, would you say that reports are accurate?
   In general, would you say that reports are on time?

2. Questions too vague. First drafts of questions are often too vague in what they are asking the participant to report. Be as clear and specific as you can.

   Bad: How often do you use the library?
   How much time do you spend studying in the library?

   Better: In a typical week, about how many times do you check-out or use any library materials, such as books, periodicals, reference materials, and so forth?

   1 NEVER
   2 LESS THAN ONCE A WEEK
   3 ONCE OR TWICE A WEEK
   4 THREE OR FOUR TIMES A WEEK
   5 FIVE OR MORE TIMES A WEEK

   On average, how many hours a week do you spend studying in the library?

   1 ONE HOUR OR LESS A WEEK
   2 TWO TO FOUR HOURS A WEEK
   3 FIVE TO SEVEN HOURS A WEEK
   4 EIGHT OR MORE HOURS A WEEK

3. Experimenter tone biased, demanding, or threatening.

   Bad: What don't you like about the immunization clinic?
   Can you tell me what the risk factors are for contracting HIV, the AIDS virus?

   Better: Is there anything you do NOT like about the immunization clinic?
   [IF YES] What would that be? Anything else?

   For each of the following behaviors, please tell me how risky you believe it is for contracting HIV, the AIDS virus.

   How about having sex with multiple partners. Would you say that this is not at all risky, somewhat risky, moderately risky, or very risky?
   How about injecting drug use? [REPEAT RESPONSES AS NECESSARY]
   How about having unprotected sex, that is, sex without a condom?
   How about having sex in a monogamous relationship, that is, a relationship where two partners have sex only with each other, and no one else?

- Number and order of response categories. A paper questionnaire can employ a greater number of response categories to a single question than a telephone survey can. Respondents to a paper questionnaire can study response options set before them. A phone survey respondent, however, must rely on their ability to remember all responses as they are read. It is generally poor policy to allow respondents to answer in their own words, and then have interviewers "code" their response on the spot into one of several response categories. This would be akin to having 30 separate data coders, each coding 1/30th of the responses. In cases where a large number of categories are anticipated for a phone survey, a viable option is to record verbatim responses (open-ended), and code them later under more controlled conditions.
- Explicit "Don't Know" and "No Answer" categories. Virtually all survey items should contain "don't know" and "refused" response options, although they are seldom read to the respondent. "Don't know" responses should generally be probed at least once in order to ensure it is a valid answer. Sometimes, however, a researcher wants to measure knowledge or awareness, in which case a "don't know" response is clearly a meaningful response; a response one may not want to probe.
- Double-barreled and other ambiguous question wordings must be avoided. For example, the question, "Which of the following actions is the most profitable and desirable for your community to undertake?" is double-barreled. What is most desirable may not necessarily be most profitable. To answer the question, the respondent must focus on one or the other. The researcher's dilemma is in interpreting the data; on which one did they focus?

Question Order

- Context effects. For respondents, and for all of us, thoughts are interrelated in memory. If we are made to think about our financial status, this may tend to evoke thoughts about our job, our bills, our career plans, and so on. Because of this tendency, early survey questions can affect how a respondent answers other survey questions. Inasmuch as one survey question must follow another, context effects are always present. Their effect must be anticipated for every survey.
- Position of sensitive questions. There are many approaches to positioning of sensitive questions. Put them early and get them over with. Put them at the end after professional rapport has been established with a respondent. Put them toward the front so a respondent will know that the survey is important. Although most survey researchers put personal and demographic questions toward the end of a survey, there is no hard and fast rule. Every survey instrument must be designed with the research objectives in mind.

Reliability and "Triangulation." ***One should always use multiple indicators of important variables whenever possible.*** For instance, for a survey assessing the effects of job stress on heart disease, job stress should be measured with multiple questions, perhaps tapping various dimensions of the phenomenon. Heart disease also should be measured in a number of ways, such as blood pressure, and respondent's personal history of heart disease; family history should also be measured so that it could be controlled for statistically. Similarly, when coding open-ended information, two or more coders should code all information, and agree on their

categorizations at least 90% of the time for the coding to be considered "reliable."

Question Validity. A survey question is considered "valid" to the extent that it measures what it is intended to measure. Some researchers rely on what is called "face validity," or, on the face of it, does it seem that the question is measuring the construct? The value of face validity should not be overlooked, however, time and again questions have been found to work counter to what was predicted. A survey researcher must have other validity tools in his or her toolkit as well, such as construct validity, content validity, convergent validity, and discriminant validity. Several of these also suggest particular statistical techniques that he or she may use to judge a question's validity.

Mode of Data Collection

The mode of data collection refers to the mode of communication that was used to ask the survey questions and record the survey answers. The technology surrounding telephone survey research has made great advances in recent years, but there are still important advantages and disadvantages that must be considered with each mode of data collection.

Face-to-Face, In-Person Interviews

- Advantages:

  May increase response rate. It is simply more difficult for a respondent to say "no" to a person's face than it is to a telephone or, simpler still, an envelope in the mailbox.

  Allows for longer interviews. Some researchers believe that longer interviews may be accomplished in person, because the respondent takes longer to become bored. The most important determinant of tolerance for long interviews, however, is always how interesting the survey material is to a respondent.

- Disadvantages:

  Respondent selection is difficult. Thoughtful protocols must be developed for true random selection. Area surveys, such as the Census Bureau Population Survey, take months of preparation and analysis before a respondent is even contacted. Representativeness of the general population cannot be obtained with "shopping center" surveys.

  Sensitive questions become biased. It has been demonstrated that sensitive questions are more problematic face-to-face. There is a greater tendency for social desirability bias.

  Interviewer bias has more channels of communication. Telephone surveys tend to mute the effects of the personality of the interviewer, while face-to-face surveys tend to exaggerate them and to allow the interviewer to influence the respondent with body language and facial expression. The interviewer is rarely aware of his or her influence.

  Much higher costs. Face-to-face surveys are not as cost efficient. The largest discrepancy occurs for surveys of randomly selected members of a geographic area. Such a telephone survey can often be accomplished for under $15 per respondent, whereas a comparable, well-done, face-to-face survey (an "area" survey) will often cost

in excess of $200 per respondent.

Telephone Surveys

- Advantages:

  Respondent selection is more easily standardized. There are a relatively small number of contingencies that must be planned for in selecting a respondent over the phone, and interviewers may be supervised to ensure that protocols are being followed.

  Faster paced. There is a lower tendency to digress over the phone, and it is easier to keep the respondent on task.

  Decrease social desirability. It is easier for a respondent to relate sensitive information over the phone than it is face-to-face.

  Truncate open-ended questions. There is a tendency for respondents to summarize and truncate open-ended information over the phone. This is usually an advantage, as it causes the respondent to make decisions as to the priority of potential pieces of information, rather than leaving that task to the researcher. It can be a disadvantage, however, if the object is to obtain a broad and inclusive statement.

  Good cooperation rates (approaching 70% for well-done surveys). Telephone surveys elicit a much higher cooperation rate than a comparable mail survey with one and even two mailings.

- Disadvantages:

  Respondent must have a phone. A phone survey would be a poor choice for coverage of populations such as homeless, low income, elderly, etc., who are less likely to own a telephone.

Mail Surveys

- Advantages:

  Relatively inexpensive. A typical mail survey is less expensive than a telephone survey, however, considering the advantages of a phone survey, the price is usually well worth it. In addition, the typical mail survey generally has weak follow-up. A well-done mail survey is about the same price as a telephone survey, and uses first-class postage, business-reply return envelopes, address-correction requested and follow-up on envelopes returned marked "undeliverable." It also includes tracking of every envelope mailed, with at least one reminder postcard mailed soon after the questionnaire, and a follow-up questionnaire mailed to all nonrespondents after a few weeks have passed. In addition, in order to assess the impact that nonresponse may have had on the data, a phone survey to nonrespondents to administer at least a set of demographic questions can be made.

  May be completed at respondents' convenience. Some respondents, such as busy executives, may be difficult to pin down for a telephone or in-person interview.

- Disadvantages:

  Low response rate (30% to 55%, for a "typical" mail survey). As was already mentioned, the typical survey, although cheaper, may have severe methodological flaws. All costs are not measured in dollars. One could invest $12,000 wisely, or spend $9,000 foolishly.

  Researcher does not know who filled it out. Will the busy business executive delegate the survey to her secretary to fill out?

  Can be confusing for respondents. Respondents to self-administered questionnaires do not have an interviewer present to head-off confusion.

Inadequate Coverage

The researcher must decide whether the population of interest is well served by the sampling method. For instance, a telephone methodology is generally a good one for statewide samples of random Utah adults. Ninety-five percent of all Utah households have telephones. On the other hand, only about 50% of Native American households on reservation lands have telephones. Many households on reservations and in rural areas are not occupied year-round so a telephone survey would not provide adequate coverage for research dealing with Native Americans in Utah.

Nonresponse

It is often difficult to discern the actual measurement error contributed by nonresponse because no data are collected from nonrespondents, so it is difficult to make comparisons empirically. A good survey organization will use a standard protocol in an attempt to maximize survey response. For instance, if an interview is not obtained on the first call, another attempt is made at a later time or date. The timing of the subsequent call may depend on the outcome of the previous call -- such as calling back a busy signal in fifteen minutes and calling back a "ring, no-answer" in two hours and on different days of the week. It is relatively standard for a survey organization to make at least nine callback attempts to contact a target respondent. Nine calls is about the number required to reach a respondent, and to avoid diminishing returns on the investment of making more attempts to contact a respondent. Surveys that have strict requirements for rate of response (e.g., requiring an 80% response rate) will require additional calls.

If you are conducting a mail survey, or an in-person survey, there are also standard methods for eliciting cooperation, such as sending out reminder cards to mail survey participants. Whatever the survey mode of administration, you should be aware that the process of maximizing response rate is almost always labor intensive and expensive. Be prepared to demand the appropriate level of service, and also to pay for it.

Section VII. Developing and Executing a Research Methodology that is Appropriate to Your Research Question

Many people overlook one or more of these steps, only to find in the end that the information they have gained from the research is not totally relevant or sufficient in some way to address their original needs. The following steps are intended to identify decisions that should be made during the course of planning and executing a research project. If each of these steps is thoroughly addressed, the research project should provide more useful, actionable, and interpretable information.

1. Develop a research team.
2. Decide what resources you have to devote to the project.
3. Come to a consensus about the function of the research: Discovery, demonstration, refutation, or replication.
4. Make some basic decisions about the general focus and scope of your project. What is it "about?" What is the issue you are researching? Why are you interested in this issue?
5. Review existing research on the topic.
6. Develop your research question. Do you have any general theoretical framework that you can identify or make explicit? What are the key constructs? What hypotheses are you making? (See Section V).
7. Decide whether your project will use qualitative or quantitative research methods, or both. This has implications for how you will operationalize your constructs.
8. Make basic sampling decisions: What is your population of inference? Do you have any thoughts about how to identify a sampling frame? Your sample will also have implications for how you operationalize your constructs.
9. What is the best way to operationalize your constructs, given the nature of inquiry (qualitative or quantitative) and the sample you will be conducting?
10. If you have decided to conduct a survey -- what will be the mode of administration, self-administered (e.g., mail), telephone, in-person interviews, or something else?
11. Make a stab at writing your questionnaire items. You can get more help with this step and others after you've hired a contractor. But remember, this is tedious work -- and it is precisely because it is tedious that you will want to stay involved in this phase. You cannot assume that your survey provider will pay attention to the details. You should be prepared to do so. Even if you have a <u>great</u> survey provider, your end product will be better if there are two people who are attending to the details.
12. Once you have a draft survey instrument, examine every question, and ask yourself, "What information will this question provide, and how will I use this information?" You should know exactly how each piece of information applies to your research question. If

your reason is "I just want to know," re-examine that question. You should know the purpose of each piece of data.

13. Circulate the draft questionnaire to experts and colleagues for their suggestions.

14. Pretest the questionnaire. Don't leave this to your survey contractor alone. You should also pretest the questionnaire yourself by administering the questions to someone who is not familiar with the project so you can hear how they interpret the questions. This will help you to see whether your questions are addressing their intended constructs adequately. It will also help you to anticipate questions that participants might have, and to judge the length of the interview and whether it is too burdensome or costly to administer.

15. Select a contractor. You can do this at an earlier step. Don't do it before step 6. You will want a contractor who is adept at helping people through steps 3-6, but I think people are better off if they have made an attempt to get through steps 1-6 before they bring the contractor on board. Once he or she is on board, you can ask them to review your decisions thus far and recommend changes. You may also want a contractor to do a more thorough job at a literature search for existing research on the topic. Also, remember that to accomplish getting a contractor on board at about step 6 in a timely fashion, you will probably need to start the bureaucratic wheels in motion before that point.

16. Have your contractor help you through any questions you have about your measurement tools' quality in terms of item reliability, validity, precision and accuracy. Understand that typically these will be largely unknown before the data have been collected -- but you can think through the issues outlined above and deal with those factors that are known or that can be anticipated.

17. Let your contractor collect the data -- but keep in touch. Make sure that he or she is interpreting your research question correctly, and interpreting your questionnaire items correctly. Keep in touch with them about what experiences their interviewers are having, and what questions the interviewers have had. How have these questions been handled? You may even want to consider sitting in on the first evening of interviewing so you can see how the respondents are interpreting your survey questions.

Section VIII. Other Topics

Ethical Issues in Social Research

At the heart of the matter is that we should "do no harm," but this is sometimes difficult to define. Ethical dilemmas occur when the benefits of the research conflict to some degree with what is in the best interest of the participants. But there are a few established rules of conduct.

1. The basic purposes and premises of the research should be such that they are making a positive contribution to society. In keeping with this perspective, the researcher should publish his or her results so that society may benefit broadly from the exercise.

2. Is there any chance of a negative effect upon the dignity or welfare of the participants? Welfare of the participant includes psychological stress, confusion, fear, or any other negative emotional or physical states or consequences. If there is any question, seek counsel from an established review board that routinely deals with questions regarding the conduct of research with human subjects. Don't trust your own objectivity. You are invested in the research and may be biased in your decision-making.

3. Participants should know they are participating in the research project, that is, they should be consenting participants.

4. Participants' confidentiality, and if possible, anonymity, should be protected. Do not store participants' names or phone numbers, or any other identifying information, alongside their responses, either on paper documents, or in computer files. Identify individuals with an ID number. If a participants' identity needs to be known for some reason, keep a cross-reference file with the participants' names and data set ID numbers. Keep this information protected. Destroy it after the need for it has passed.

5. If, for any reason, participants' confidentiality cannot be maintained, then informed consent must be obtained from each participant specifically with respect to the nature of the disclosure.

6. Disclosure of the survey methods will allow others to make appropriate interpretations of the results. The American Association for Public Opinion Research has published the following guidelines for disclosure of survey methods. They can be adapted for other types of research methodologies.

   The following items should be disclosed:

   - Who sponsored the survey, and who conducted it.
   - The exact wording of the questions asked, including the text of any preceding instruction or explanation to the interviewer or respondents that might reasonably be expected to affect the response.
   - The definition of the population under study, and a description of the sampling frame used to identify this population.

- A description of the sample selection procedure, giving a clear indication of the method by which the respondents were selected by the researcher, or whether the respondents were entirely self-selected.
- Size of all samples and, if applicable, completion rates and information on eligibility criteria and screening procedures.
- A discussion of the precision of the findings, including, if appropriate, estimates of sampling error, and a description of any weighting or estimating procedures used.
- Which results are based on parts of the sample, rather than on the total sample.
- Method, location, and dates of data collection.

Costs of Social Research

Costs vary depending on sample size, mode of data collection (mail, phone, in-person), difficulty in reaching participants, difficulty in eliciting participation, the amount of service requested from the provider (e.g., questionnaire construction versus administering an existing questionnaire), and other factors.

Mail -- thought of as less expensive, but is very close to phone survey cost if it is done correctly.

Phone -- usually pretty efficient, good value if it is appropriate to the research question. Expect to pay about $20 per completed interview for the data collection (assuming RDD sample and about 15 minutes per interview).

In-Person -- costs 10 to 20 times the cost of phone survey for the same sample size.

Writing an RFP and Evaluating the Proposals

You want to give the offerors enough information on which to base a quality proposal, but not so much information that you let them know exactly what you want in too much detail. It is a little like writing an exam -- you don't want everyone to fail, or everyone to pass with honors -- you want to write a test that will be a sensitive measure of the offerors' abilities. That is, you want the RFP to discriminate between good offerors and poor ones.

Using Existing Data

Data sources that are already in existence should be evaluated on the basis of all the sampling and measurement constructs presented here. e.g., Who is the population of inference? How well does the sample represent the population? How well was the process of data collection standardized? How were the variables operationalized? Are the data reliable? Are they valid?

In addition, make sure that the data actually inform your research question, instead of forming it. That is, you should have a research question first, based on your needs and interests,

then evaluate the data set regarding how well it addresses that question. People often look at a data set first, and then try to think up research questions that the data can answer.

Special Topics - RDD and CATI

There are two topics that are not sources of measurement error, but are current methodologies commonly used in telephone survey research to reduce measurement error. The two topics are Random-digit dialing (RDD), and computer-assisted telephone interviewing (CATI).

(RDD) Methodology. The phone survey has become such a mainstream survey research tool that a separate discussion of RDD methods seems required. There are many occasions for which a random sample of individuals is necessary, and a random-digit dial methodology is a sound and relatively inexpensive way to obtain such a sample.

Contrary to popular belief, the phone book is not a good source of phone numbers for a survey for one reason: It does not contain unlisted phone numbers. In Utah, about 30% of all households have unlisted phone numbers (this figure is even higher in most other states). Wealthy households, very low-income households, and households headed by single females are more likely to have unlisted phone numbers.

Some firms still use telephone directories, despite their limitations. Many private survey research firms purchase lists of telephone numbers from companies that sell phone numbers for various geographic areas, but some of these firms are not especially forthcoming with their number-generation methodologies.

Sound random-digit dial methodologies do not rotate through all the extensions in a given prefix area. Nor do they simply have a computer program randomly generate the last four digits of the phone number for each known telephone prefix. A given prefix contains many unassigned phone numbers, disconnected phone numbers, and business phone numbers. To complicate matters, telephone prefixes contain differing quantities of phone numbers. There are 10,000 possible suffixes for a given telephone prefix (0000-9999). One prefix may be almost totally saturated, while another may only have 100 or so numbers currently assigned to it. It is important for the RDD methodology to take this into account and ensure that every household in a geographic area has an equal probability of selection, and furthermore, that the number of calls to businesses, nonassigned, and disconnected phone numbers is kept to a minimum. There are special methodologies, such as Waksberg/Mitofsky and others that deal with these issues effectively.

Computer-Assisted Telephone Interviewing (CATI). CATI is becoming a standard in the field of telephone survey interviewing. It allows for strict quality control, and increases interviewer efficiency. Although CATI methods influence almost all aspects of the data collection process, they have the greatest influence over interviewer error and questionnaire construction.

- Interviewer error. A CATI system will not allow an interviewer to enter an invalid code. It may also be programmed to question or not accept illogical answers (such as a respondent

reporting they have never been married, and including their spouse among the members of their household). CATI may also be programmed to go to special "help" screens if an interviewer needs to clarify a question, or obtain more information about a certain question or section of a questionnaire.

- Questionnaire construction. One of the most fundamental and important uses of a CATI system is the ability to use more complex "branching" structures in questionnaires. It is virtually impossible to construct a questionnaire that does not contain at least one question that is asked or not asked, contingent on the answer to a previous question. This is called a branching structure. For instance, if the respondent has children under age 5, the interviewer should ask the branch of questions about childcare, if the respondent has children between age 5 and 18, the interviewer should ask the branch of questions about public education, and if the respondent has no children under age 18, skip around these branches, and return to the main questionnaire. Recent advancements in educational testing will allow a standardized intelligence test to be administered on a CATI system. If the respondent answers question 1 correctly, he or she will go to a more difficult question than a respondent who answered question 1 incorrectly. Such a questionnaire structure allows the researcher or tester to obtain more precise measurements. It also contains fewer questions than paper-and-pencil tests, which allows the researcher to test intelligence without the confounding effect of respondent fatigue.

Bibliography

Grimley, Diane M.; Prochaska, James O.; Velicer, Wayne F.; and Prochaska, Gabrielle E. (1995) Health Education Quarterly. Vol 22 (1), 20-25.

Prochaska, James O.; Velicer, Wayne F.; Rossi, Joseph S.; Goldstein, Michael G.; Marcus, Bess H.; Rakowski, William; Fiore, Christine; Harlow, Lisa L.; Redding, Colleen A;, Rosenbloom, Dena; and Rossi, Susan R. (1994) Health Psychology. Vol 13 (1), 39-46.

Selected Survey Research References

Aday, LuAnn (1989) Designing and Conducting Health Surveys: A Comprehensive Guide. San Francisco: Jossey-Bass, 422 pp., ISBN 1-55542-173-3.

Aquilino, William S. (1994). Interview mode effects in surveys of drug and alcohol use: A field experiment. Public Opinion Quarterly, 58, 210-240.

Asher, H.B. (1974) Some consequences of measurement error in survey data. American Journal of Political Science, 18: 469-485.

Babbie, Earl (1990) Survey Research Methods, 2nd Ed.. Belmont, CA: Wadsworth Publishing Company.

Biemer, Paul P. Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman (1991) Measurement Errors in Surveys. New York: John Wiley and Sons.

Billet, J. and Loosveldt, G. (1988) Improvement of the quality of responses to factual survey questions by interviewer training. Public Opinion Quarterly, 52 (2), 190-211.

Brick, J. Michael, Waksberg, Joseph, and Scott Keeter (1994) Evaluating the use of data on interruptions in telephone service for nontelephone households. 1994 Proceedings of the American Statistical Association, Section on Survey Research Methods, Vol.1, 19-28.

Cochran, W. (1977) Sampling Techniques. New York: John Wiley and Sons. ISBN: 0-471-16240-x.

Cohen, G., & Java, R. (1995). Memory for medical history: Accuracy of recall. Applied Cognitive Psychology, 9, 273-288.

Converse, Jean M., and Stanley Presser (1986) Survey Questions: Handcrafting the Standardized Questionnaire. Sage Series: Quantitative applications in the Social Sciences. Newbury Park: Sage Publications.

Delgado, Jane L. & Estrada, Leo (1993) Improving data collection strategies. Public Health Reports, 108, 540-545.

DeVellis, Robert F. (1991) Scale Development: Theory and Applications. Beverly Hills: Sage Publications.

Dillman, Don A. (1978) Mail and Telephone Surveys: A Total Design Method. New York: John Wiley and Sons.

Dillman, Don A. and Sangster, Roberta L. Mail Surveys: A Comprehensive Bibliography, 1974-1989. CPL Bibliography no 272. Chicago: Council of Planning Librarians, 1991, 26 pp.

Duncan, George T., Javine, Thomas B. & deWolf, Virginia A. (Eds.) (1993). Private lives and public policies: confidentiality and accessibility of government statistics. Washington, DC: National Academy Press.

Feinberg, Stephen E. (1994). Conflicts between the needs for access to statistical information and demands for confidentiality. Journal of Official Statistics, 10, 115-132.

Foddy, William. (1993). Constructing questions for interviews and questionnaires. New York: Cambridge Univ. Press.

Fowler, Floyd J. Jr. (1984, 1988) Survey Research Methods. Beverly Hills: Sage Publications.

Fowler, Floyd J., Jr., Berwick, Donald M., Roman, Anthony, & Massagli, Michael P. (1994). Measuring public priorities for insurable health care. Medical Care, 32, 625-639.

Fowler, Floyd J., and Mangione, Thomas W. (1990) Standardized Survey Interviewing: Minimizing Interviewer-Related Error. Beverly Hills: Sage Publishing.

Frey, James (1983) Survey Research by Telephone. Beverly Hills: Sage Publishing.

Groves, Robert M. (1989) Survey Errors and Survey Costs. New York: John Wiley and Sons.

Groves, Robert M., Paul P. Biemer, Lars E. Lyberg, James T. Massey, William L. Nichols II, and Joseph Waksberg (1988) Telephone Survey Methodology. New York: John Wiley and Sons.

Guess, Harry A., Jacobsen, Steven J., Firman, Cynthia J., Oesterling, Joseph E., Chute, Christopher G., Panser, Laurel A., & Lieber, Michael M. (1995). The role of community-based longitudinal studies in evaluating treatment effects: Example: Benign prostatic hyperplasia. Medical Care, 33 (April supplement), AS26-AS35.

Hahn, Robert A. (1992) Inconsistencies in Coding Race and Ethnicity between Birth and Death in US Infants. Journal of the American Medical Association, (January 8), pp. 259-63.

Hahn, Robert A. (1992) The State and Federal health Statistics on Racial and Ethnic Groups. Journal of the American Medical Association, (January 8) pp. 268-71.

Hermann, Douglas. (1995). Reporting current, past, and changed health status: What we know about distortion. Medical Care, 33 (April supplement) AS89-AS94.

James, Jeannine M. and Richard Bolstein (1990) The effect of monetary incentives and follow-up mailings on the response rate and response quality in mail surveys. Public Opinion Quarterly, 54, 346-361.

Charles M. Judd, et al., (1991). Research Methods in Social Relations, 6th Ed. Fort Worth: Holt, Rinehart and Winston, Inc.

Kalton, Graham (1983) Introduction to Survey Sampling, Vol 35, Quantitative Applications in the Social Sciences. Newbury Park: Sage Publications.

Kessler, Ronald C. & Mroczek, Daniel K. (1995). Measuring the effects of medical interventions. Medical Care, 33 (April supplement), AS109-AS119.

Kish, Leslie (1965) Survey Sampling. New York: John Wiley and Sons.

Kristal, Alan R., White, Emily, Davis, Julie T., Corycell [sic], Gayle, Raghunathan, Trivellore, Kinne, Suran & Lin, Ting-Kwong. (1993). Effects of enhanced calling efforts on response rates, estimates of health behavior and costs in a telephone health survey using random-digit dialing. Public Health Reports, 108, 372-379.

Lavrakas, Paul J. (1987, 1988) Telephone Survey Methods: Sampling, Selection and Supervision. Beverly Hills: Sage Publications, Inc.

Lee, Eun Sul, Ronald N. Forthofer, and Ronald J. Lorimor (1989) Analyzing Complex Survey Data, Volume 71, Quantitative Applications in the Social Sciences. Newbury Park: Sage Publications.

Lessler, Judith T. (1995). Choosing questions that people can understand and answer. Medical Care, 33 (April supplement), AS203-AS208.

Lessler, Judith T. And Kalsbeek, William D. (1992). Nonsampling Errors in Surveys. New York: Wiley, 425 pp. $69.95.

Levy, P.S.; and Lemeshow, S. Sampling of Populations: Methods and Applications. New York: Wiley, 1991, 415 pp., $49.95.

Lin, I-Fen, & Schaeffer, Nora Cate. (1995). Using survey participants to estimate the impact of nonparticipation. Public Opinion Quarterly, 59, 236-258.

Locander, W.B., and Burton, J.P. (1976) The effect of question form on gathering income data by telephone. Journal of Marketing Research, 13, 189-192.

Mancuso, Carol A., & Charlson, Mary E. (1995). Does recollection error threaten the validity of cross-sectional studies of effectiveness? Medical Care, 33 (April supplement), AS77-AS88.

Martin, J.K., and Bason, J.J. (1993) The effects of interview length, type of sample, sample size and conversion efforts on refusal rates in 23 telephone surveys. Proceedings, Institute of Management Sciences, 23, 166-168.

Matell, M.S. and Jacoby, J. (1971) Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and Validity. Educational and Psychological Measurement, 31, 657-674.

Miller, P.V. (1984) Alternative question forms for attitude scale questions in telephone interviews. Public Opinion Quarterly, 48 (4), 666-668.

Morel, J.G. (1989) Logistic regression under complex survey designs. Survey Methodology, 15

(December), 203-23.

Oldendick, Robert W. (1993) The effect of answering machines on the representativeness of samples in telephone surveys. Journal of Official Statistics, 9, 663-672.

Oskenberg, Lois; Cannell, Charles; and Kalton, Graham (1991) New strategies of pretesting survey questions. Journal of Official Statistics, 7 (3), 349-65.

Parsons, J. A. , Warnecke, R.B., Czaja, R.F., Barnsley, J., and Kaluzny, A. (1994). Factors associated with response rates in a national survey of primary care physicians. Evaluation Review, 18, 756-766.

Piazza, Thomas. (1993). Meeting the challenge of answering machines. Public Opinion Quarterly, 57, 219-231.

Presser, S. And Blair, J. (1994) Survey pretesting: Do different methods produce different results? Sociological Methodology.

Psaty, B. M., Cheadle, A. Koepsell, T.D., Diehr, P, et al. (1994) Race- and ethnicity-specific characteristics of participants lost to follow-up in a telephone cohort. American Journal of Epidemiology, 140, 161-171.

Rogers, Theresa F. (1976) Interviews by telephone and in-person: Quality of responses and field performance. Public Opinion Quarterly, 40, 51-65.

Rossi, Pater H., James D. Wright, and Andy B. Anderson (Eds) (1983) Handbook of Survey Research. Orlando: Academic Press, Inc.

Rowland, Michael L., & Forthofer, Ronald N. (1993). Adjusting for nonresponse bias in a health examination survey. Public Health Reports, 108, 380-386.

Schabenberger, O. And Gregoire, T.G. (1994) Competitors to genuine πps sample designs: A comparison. Survey Methodology, 20, 185-192.

Schechter, S. (1994, August) Proceedings of the 1993 NCHS conference on the cognitive aspects of self-reported health status. (NCHS Cognitive Methods Staff Working Paper Series, No. 2).

Smith, Tom W. (1984) Nonattitudes: A review and evaluation. IN C.F. Turner, and E Martin (Eds) Surveying Objective Phenomena. New York: Russel Sage Foundation, pp 215-255.

Smith, Tom W. (1990) Phone home? An analysis of household telephone ownership.

International Journal of Public Opinion Research, 2, (Winter), 369-90.

Smith, Tom W. (1995). Some aspects of measuring education. Social Science Research, 24, 215-242.

Sobel, Jeffrey; De Forge, Bruce R.; Ferentz, Kevin S.; Muncie, Herbert L. Jr.; Valente, Carmine M.; and Levine, David M. (1990) Physician response to multiple questionnaire mailing. Evaluation Review, 14 (December), 711-22.

Stewart, David W. and Prem N. Shamdasani (1990) Focus Groups: Theory and Practice. Newbury Park: Sage Publications.

Sudman, Seymour (1976) Applied Sampling. San Diego: Academic Press, Inc.

Sullivan, Lisa M., Dukes, Kimberly A., Harris, Lisa, Dittus, Robert S., Greenfield, Sheldon, & Kaplan, Sherrie H. (1995). A comparison of various methods of collecting self-reported health outcomes data among low-income and minority patients. Medical Care, 33 (April supplement), AS183-AS194.

Thornberry, Owen T., Jr., and Massey, James T. (1988) Trends in United States telephone coverage across time and subgroups. In Robert M. Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey, William L. Nicholls II, and Joseph Waksberg (Eds.) Telephone Survey Methodology. New York: John Wiley & Sons.

U.S. Department of Commerce (1978) Glossary of Nonsampling Error Terms, Washington D.C., Statistical Policy Working Paper 4.

Waksberg, Joseph (1978) Sampling methods for random digit dialing. Journal of the American Statistical Association, 73, 40-46.

Weiss, C.H. (1968) Validity of welfare mothers' interview responses. Public Opinion Quarterly, 32 (4), 622-626.

Werking, George S. And Clayton, Richard L. (1991) Enhancing data quality through the use of mixed mode collection. Survey Methodology, 17 (June), 3-14.

Willia, G. (1994, March) Cognitive interviewing and questionnaire design: A training manual. (NCHS Cognitive Methods Staff Working Paper Series, No. 7).

Yammarino, Francis J.; Skinner, Steven J.; and Childers, Terry L. (1991). Understanding mail survey response behavior: A meta-analysis. Public Opinion Quarterly, 55 (Winter), 613-39.

Annual index of questions asked: American Public Opinion Index. Boston, MA: Opinion Research Service.