

Writing EFFECTIVE Survey Items

University Testing and Assessment Center

Oklahoma State University

Lunchtime Workshop

October 7, 2009

ITLE Center

Created By Dr. Katye Perry

I. INTRODUCTION

A. Some Questions:

- 1) How many of you have ever RECEIVED a survey?
- 2) How many of you have ever COMPLETED a survey you received?
- 3) WHY did you complete the survey?
- 4) How many of you have ever RECEIVED a survey you DID NOT complete?
- 5) WHY did you not complete the survey?

B. Four Interrelated Issues

- 1) The Problem
- 2) The People
- 3) The Purpose
- 4) The Method

II. THE PROBLEM

A. Why are you conducting the study?

B. What is the problem?:

[What will the information from the survey tell you about the problem?]

NOTE: Every item on your survey should be related to the problem

III. The People

A. The Population

Who can best address the problem that directs this study?

How large is the population?

B. Census or Sample?

Census: The entire population

Sample: Representative group selected from within the population

- 1) Straight random
- 2) Stratified proportional/even sample
- 3) Cluster
- 4) Systematic

**NOTE: Know your population, its size, demographics and culture, etc.
Representativeness is key.**

IV. The Purpose

A. What specific questions do you want answered?

NOTE: In what way do these questions relate to the problem?

V. The Method

A. How are you going to ask and get the respondents to answer your questions?

B. How do you approach the respondents?

NOTE: What would keep respondents from participating in your study?

C. Prequestionnaire/Survey Cover letter

- 1) Purpose of survey
- 2) Why you are asking them
- 3) What they will get from their participation
- 4) How will their identity be addressed: confidentiality vs. anonymity
- 5) How long it will take to complete the survey
- 6) How long do they have before they have to return the survey
- 7) Symbols of credibility

C. Personal contacts/Advocates

NOTE: Information within the cover letter and advocates provide assurances that the respondents and the information they provide are valued and protected.

VI. CATEGORIES OF FOCUS OF QUESTIONS

A. Behavior

How a person **RESPONDS** to issue, event, etc.

B. Attitude

How a person **PERCEIVES/FEELS** about a particular issue, topics, etc.

C. Knowledge

How much a person **KNOWS** about a particular issue, topic, etc.

VII. NON-THREATENING QUESTIONS

These are questions people are at ease in discussing (Mertens, 2005) and are typically presented in closed (e.g. multiple choice, true-false, fill-in-and or/open (fill-in-the-blank or short answer) formats. (See Checklist ?? for suggestions for writing non-threatening questions)

VIII. THREATENING QUESTIONS

These are questions people are NOT at ease in discussing and may feel defensive when answering the questions (Mertens, 2005). These questions can be presented in closed, open, short answer formats.

The wording of these questions is critical. For example think of querying a group regarding the subject of domestic violence, abortion, gun control, public option for health care reform, etc. (See Checklist__)

VIII. VALIDATION

A. Validity

Determining if the survey measures what it is intended to measure

- Content Validity

B. Reliability

Determining if the survey consistently measures what it is intended to measure

- Test-Retest with a Pilot Group
- Scorer/Rater Consistency
- Interrater Reliability
- Intrarater Reliability

IX. RESPONSE RATES

A. Increasing response rates

- 1) Research findings for mail and web surveys
- 2) Checklist for increasing response survey response rates

B. Acceptable response rates
(See Mertens, pgs 197-198)

REFERENCES

- Cook, C., Heath, F., & Thompson, R. (2000). A meta-analysis of response rates in Web or Internet-based surveys. *Educational and Psychological Measurement*, 60, 821-836. In Mertens, D.M. (2005). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (2nd ed.). Thousand Oaks: Sage
- Dillman, D.A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: John Wiley & Sons, Inc.
- Dillman, D.A., Tortora, R.D., Conrad, J., & Bowker, D. (2001). *Principles for constructing Web surveys*. Working paper. Retrieved April 13, 2002, from <http://survey.sesrc.wsu.edu/dillman/papers.htm>.
- Kittleson, M. (1997). Determining effective follow-up of e-mail surveys. *American Journal of Health Behavior*, 21, 193-196. In Mertens, D.M. (2005). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (2nd ed.). Thousand Oaks: Sage
- Litwin, M.S. (1995). *How to measure survey reliability and validity*. Thousand Oaks: Sage
- Mertens, D.M. (2005). *Research and evaluation in education and psychology: Integrating diversity quantitative, qualitative, and mixed methods* (2nd ed.). Thousand Oaks: Sage
- Ritter, L.A. & Sue, V.M.(Fall, 2007). Using online surveys in evaluation. *New Directions for Evaluation*, (115).

Sudman, S. & Bradburn, N. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.

Yammarino, R.S., Skinner, S.S., & Childers, T.L. (1991). Understanding mail survey Response behavior: A meta-analysis, *Public Opinion Quarterly*, 55, 613-639. In
Mertens, D.M. (2005). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (2nd ed.). Thousand Oaks: Sage

CHECKLIST FOR ASKING NON-THREATENING QUESTIONS

1. Decide whether the question is or is not threatening.
2. When asking a closed question (sic e.g. multiple choice) about behavior, make sure that all reasonable alternative answers are included. Omitted alternatives and answers lumped under an *Other* category will be underreported.
3. Aided-recall procedures may be helpful if the major problem is underreporting of behavior (e.g. Provide example of behaviors associated with the question).
4. Make the question as specific as possible. More reliable information is obtained when-instead of asking about usual behavior-you ask about behavior in an exact time period. Looser questions that require less time may be used, however, if the goal is to group respondents into categories rather than to measure a behavior very precisely.
5. The time period of the question should be related to the saliency of the topic. Periods of a year (sometimes even longer) can be used for highly salient topics, such as purchase of a new house, birth of a child, or a serious auto accident. Periods of a month or less should be used for items with low saliency, such as purchases of clothing and minor household appliances; periods that are too short, however, should be avoided, since forward telescoping (remembering the event as having occurred more recently than it did) can cause substantial overstatements in reports of behavior.
6. The use of records (where available, household observations, and bounded recall (sic e.g. reminder of answers previously provided by respondents on same question(s)) will reduce or eliminate telescoping and also improve the reporting of detailed information.
7. Where detailed information on frequent, low-salience behavior is required, diaries will provide more accurate results than memory.
8. Use words that virtually all respondents will understand. Do not use special terms unless members of the sample would be expected to know them or the term is explained in the question.
9. Increasing the length of the question by adding memory cues may improve the quality of reporting. Do not assume that the shorter, the better.
10. Recognize that, for nonthreatening behavior, respondents will generally give more accurate information about themselves than about relatives, friends, or co-workers. If cost is a factor, however, informants can provide reasonably accurate information about others-specifically, parents about children and spouses about each other.

CHECKLIST FOR ASKING THREATENING QUESTIONS

1. Open questions are better than closed questions for obtaining information on frequencies of socially undesirable behaviors.
2. Long questions are better than short questions for obtaining information on frequencies of socially undesirable behavior.
3. The use of familiar words may increase the reporting of frequencies of socially undesirable behavior.
4. To obtain reports of threatening behavior, use data obtained from informants, if possible.
5. Consider deliberate loading of the question to reduce both overstatement of socially desirable and understatement of socially undesirable behavior (e.g. *Everybody does it; Many doctors are reporting the benefits of the consumption of red wine...*) Do not depend on wording (such as, *Did you happen to..*) to improve reporting of socially undesirable behavior. Such wording may actually increase threat.
6. For socially undesirable behavior, it is better, before asking about current behavior, to ask whether the respondent has ever engaged in the behavior. For socially desirable behavior, it is better to ask about current rather than usual behavior.
7. Embed the threatening topic into a list of more and less threatening topics, to reduce the perceived importance of the topic to the respondent.
8. Consider alternatives to standard questions, such as randomized response, card sorting, and sealed envelopes.
9. Consider the use of diaries or asking questions in several waves of a panel, to improve both reliability and validity.
10. Avoid the use of reliability checks on the same questionnaire, since this will annoy respondents.
11. Ask questions at the end of the interview to determine how threatening the topics were perceived to be by the respondents.
12. Attempt to validate, even if only on an aggregate level.

Sudman & Blackburn, pgs 55-56

CHECKLIST FOR MEASURING ATTITUDES

1. Make sure the attitude objects are clearly specified.
2. Decide on the critical aspects of the attitude to be measured-affective, cognitive, and action. Do not assume that these must necessarily be consistent.
3. Measure the strength of the attitude by building a strength dimension into the question itself; by asking a separate question or questions about strength; or by asking a series of independent questions, each of which reflects the general attitude.
4. Avoid double-barreled and one-and-a-half-barreled questions that introduce multiple concepts and do not have a single answer.

Double-barreled questions: Questions in which two opinions are joined together so that respondents must answer two questions at once when their opinions about the two may diverge

One-and-a-Half-Barreled questions: A question that poses a single attitude object, and the respondent is asked to respond along a scale of degrees of favorableness and unfavorableness.

5. Consider the use of separate unipolar items if there is a possibility that a bipolar might miss independent dimensions.
6. Recognize that the presence or absence of an explicitly stated alternative can have dramatic effects on response. Specification of alternative will standardize the question for respondents.
7. Pretest new attitude questions. The use of split ballots (i.e. changing the order of the questions on test questionnaires) in pretests is highly desirable.
8. If general and specific attitude questions are related, ask the general question first.
9. When asking questions of differing degrees of popularity involving the same underlying value, ask the least popular item first.
10. In attempting to measure changes in attitude over time, ask exactly the same questions in all time periods, if at all possible.

Sudman & Bradburn pg. 121

CHECKLIST FOR MEASURING KNOWLEDGE

1. Before asking attitude questions about issues or persons, ask knowledge questions to screen out Respondents who lack sufficient information.
2. Consider whether the level of difficulty of the questions is appropriate for the purposes of the study. For new issues simple questions may be necessary.
3. Where possible, reduce the threat of knowledge questions by asking them as opinions or using phrases Such as *do you happen to know* or *can you recall, offhand*.
4. When identifying persons or organizations, avoid overestimates of knowledge by asking for additional information or including fictitious names on the lists.
5. If *yes-no* questions are appropriate, ask several on the same topic, to reduce the likelihood of successful guessing.
6. For knowledge questions requiring numerical answers, use open-ended questions to avoid either giving away the answer or misleading the respondent.
7. To increase reliability when obtaining information about a geographical area, use multiple key informants or individual respondents.
8. Consider the use of pictures and other nonverbal procedures for determining knowledge.
9. When attempting to determine level of knowledge, do not use mail or other procedures that allow the respondent to look things up or to consult others.

Sudman & Blackburn pg. 89

**EXAMPLES OF DIFFERENT
TYPES OF QUESTIONS**

NON-THREATENING QUESTIONS:

1. Closed Format:

What best describes your Counseling Services? (Please check all that apply)

- Your Services (clients under age 18)
- Domestic Violence
- Marriage and Family
- Pastoral Counseling
- Drug & Alcohol Addiction
- Other (Please specify : _____)

2. Open Format

What best describes your Counseling Services? _____

Demographic Information: Race Identification

Add: Native American: Tribal affiliation: _____
Multiracial (Please explain : _____)
Other (Please explain : _____)

3. Bounded Recall

On average, how long must clients wait to be seen?

- Within ONE Week
- Within TWO Weeks
- Within THREE Weeks
- One Month or More

4. Measuring Knowledge

Which one of the following is the MAJOR goal of the United Nations?

- To fight disease
- To maintain peace
- To spread democracy
- To fight the Taliban

THREATENING BEHAVIORAL QUESTIONS

1. **Longer threatening questions are generally better than shorter ones. It gives people some time to recover from their initial shock that the researcher would ask about such a topic, and they can begin to formulate their response in whatever carefully selected words they chose.**

We know that some kids drink alcohol and some kids don't. Some might drink alcohol a little bit, and some a lot. Some might have a few drinks now and then, whereas others drink to get drunk. Some kids drink wine or beer, and Others drink with their friends. Some kids drink wine or beer, and others drink the harder stuff like whiskey and gin. How about you? Have you drunk alcohol in your life? (Mertens, pg. 189)

Measuring Attitudes

2. **Double-Barreled Question:** A question in which two opinions are joined together so that the respondent must answer two questions at once when their opinions about the two may diverge

Are you in favor of building more nuclear power plants so that we can have enough electricity to meet the country's needs, or are you opposed to more nuclear power plants even though this would mean less electricity?

(Sudman & Blackburn, pg. 133)

3. **One-and-a-Half-Barreled Question:** A question is posed about a single attitude object, and the respondent is asked to respond along a scale of degrees of favorableness and unfavorableness. The responses start quite straightforwardly along a single dimension, but somewhere along the line a second opinion object is introduced as part of the response continuum.

(Sudman & Blackburn pgs.135-136)

*Do you think the United States was right or wrong in sending American troops to stop the **Communist** invasion of South Korea?*

_____ *Wrong*
_____ *Right*
_____ *Don't know*

(Sudman & Blackburn pg. 3)

CHECKLIST FOR INCREASING RESPONSE RATES (B)

1. Personal and phone interviews generally yield higher response rates than do mail surveys
2. Questionnaire length sometimes has an effect on response rates. Short questionnaires will be returned at a higher rate than long ones unless the respondent endorses the importance of the topic.
3. Monetary and nonmonetary incentives can increase response rates.
4. Sponsorship (through a respected organization or signature by a high-ranking administrator) can yield higher response rates.
5. Good timing is *very* important, especially with populations that organize their time around a school calendar. Avoid sending the questionnaires at the very beginning or end of a school year or around a holiday period.
6. Be sure to make it easy to return the questionnaire by providing a stamped, self-addressed envelope (and maybe a pencil). Even your choice of stamps versus metered mail can influence response rates, especially if you are able to locate a stamp that has salience for your population. For example, the U.S. Postal Service has printed stamps that depicted a deaf mother signing *I love you* to her child.
7. Consider using an attention-getting delivery mode, such as express mail, special delivery, or airmail and, possibly, printing your questionnaire on highly colored paper.
8. Handwritten envelopes seem to attract more attention now that computer-generated address labels.
9. Your follow-ups could be delivered via humorous postcards or telephone calls.

Mertens, pgs. 196-196

RESEARCH FINDINGS ON INCREASING RESPONSE RATES FOR MAIL AND WEB SURVEYS

The following is a direct quote from Mertens, pg. 196:

Mail Surveys

*Based on a meta-analysis of 115 studies of response rates for **mail surveys**, the following factors were associated with an increased return rate:*

- *Preliminary notification and follow-ups*
- *Appeals*
- *Inclusion of a return envelope*
- *Differences in postage (e.g. use of a first class stamp)*
- *Monetary incentives (Yammarion, Skinner, & Childres, 1991)*

Web Surveys

Cook, Heath, and Thompson (2000) conducted a meta-analysis of factors influencing response rates in Internet-based surveys. They found three factors that increased response rates. Follow-up contacts with nonrespondents, personalized contacts, and contacting sampled people prior to sending out the survey were the three dominant factors in higher response rates. Kittleson (1997), in a study of e-mailed based surveying, found it was possible to double the response rate with follow-up memos. As with mailed surveys, repeated follow-ups have diminishing returns and at some point risk irritating potential respondents without noticeably increasing response rates. Additionally, Dillman, Tortora, Conrad, and Bowker (2001) found that relatively plain Web surveys that load quickly resulted in higher response rates than fancier surveys that take longer to load.

ESTABLISHING VALIDITY AND RELIABILITY OF THE SURVEY

VALIDITY

Content Validity:

Table of Specification: Alignment of survey items against the questions that guide the study

- Expert Judgment:
- 1) Determining agreement among experts regarding the extent to which the study questions agree with the intent of the study; and
 - 2) Determining agreement among experts regarding the extent to which the study questions answer the specific questions that measure the specific research questions/objectives

RELIABILITY

Test-Retest with Pilot Group:

- 1) Where appropriate, administer the survey to the same group twice, within one to two weeks and measure consistency in their responses; and

Scorer/Rater Consistency:

- 1) Using an appropriate scale, ask raters to measure the extent to which the survey items measure the intended measure of the specific research questions/objectives. Measure percent of agreement

Interrater Reliability:

- 1) With the use of a numerical rating scale, correlating the agreement among raters regarding the extent to which each and/or all items measure the intent of the survey. (Also, measure of construct validity)

Intrarater Reliability:

- 1) With the use of a numerical rating scale, correlating the agreement of each individual's ratings on more than one rating of the same items.