Constructing Mail Survey Questionnaires to Maximize Rates of Return and Assure the Validity and Reliability of Responses

OBJECTIVES: Because the results generated from mail survey questionnaires can be a rich source of information for practitioners, administrators, marketers, and policy-makers, the researcher must have confidence in the validity and reliability of the estimates obtained. Following simple tenets of survey questionnaire construction can help prevent spurious results. After completing this continuing education program, participants will be able to:

- Identify methods of increasing the likelihood of a higher rate of return for mail surveys.
- Discuss the importance of a properly written cover letter.
- Describe the components of a properly written cover letter.
- Design survey questions to collect demographic, personal, and practice-related data.
- Identify some potential pitfalls when constructing attitude scales.
- Construct items comprising a summed ratings scale.

KEYWORDS: Survey, Questionnaires, Validity, Response Rate

by Shane P. Desselle, R.Ph., Ph.D.

In the first of this two-part series, the principal steps in conducting mail survey research were outlined. Each step was explained in detail, except for the methods employed when constructing the survey instrument (the “survey” or “survey questionnaire”) itself. Constructing the items and questions comprising a survey questionnaire requires knowing both psychometric techniques and common-sense pitfalls to avoid. Although inexperienced researchers often assume that constructing the instrument is the easiest part of the research project, it is often the most problematic.

The most important issues to address when constructing a mail survey instrument are the validity of the constructs being measured, the consistency of the measures, and the maximization of the survey’s rate of return. This article describes how to write survey items/questions to maximize the likelihood of achieving construct validity and reliability.

The process begins with the researcher working to maximize the survey’s rate of return; we then turn to methods to draft a cover letter, use of multiple mailing procedures, writing questions that elicit demographic data, and, finally, how to generate the items on a scale to measure the attitudes or behaviors of research subjects.

Maximizing Rate of Return

In general, the investigator seeks to maximize the number of responses obtained from the sampled population. A higher response rate affords the investigator greater confidence that the results obtained reflect those that would be obtained from the entire population; or, more correctly, it reduces the possibility that the responses obtained are biased. A sound sampling procedure and a high response rate minimize the likelihood that responses will come disproportionately from certain segments of the sampled population or from persons with predisposed attitudes or behaviors that might not reflect those of the population as a whole.

For example, if an investigator mailed out a survey to determine how satisfied plan members were with their health plans, a low response rate might raise concern that a disproportionately greater number of people extremely dissatisfied with their plan...
Constructing Mail Survey Questionnaires

had responded. A low response rate to a survey asking pharma-
cists about their propensity to provide cognitive services would
raise concern that pharmacists who already do this or have posi-
tive attitudes about such behaviors were more likely to have
returned the survey, thus overestimating pharmacists’ propensity
to provide such services. A low rate of return may also preclude
the use of certain statistical procedures, particularly if the initial
sample was itself small.

Although there is no way to guarantee high rates of return,
investigators can do several things to minimize the likelihood of
their being especially low. They must first familiarize themselves
with the pool of potential respondents to the survey. Losing sight
of who comprises the sample population could lead to mistakes
that jeopardize a good response rate. Investigators should make
certain that questions make sense to the respondents, employ con-
ventional language at an appropriate reading level, avoid using
technical jargon that may be unfamiliar to respondents, avoid
using abbreviations or acronyms, and avoid terminology or infer-
ences that respondents might find offensive.

The Cover Letter

Secondly, investigators should construct a cover letter to accom-
pany the survey instrument. The cover letter should be relatively
brief yet explain the purpose of the research project in language
that the respondent can understand. The purpose should be stat-
ed in a way that reflects the potential benefits that may accrue as a
result of the research. For example, if an investigator were con-
ducting a project to determine health literacy, the cover letter
might state that the knowledge gained from the project will be
used to tailor educational materials that will improve members’
understanding of how to maximize the benefits they gain from
their health insurance coverage.

The cover letter should also declare how the recipient was
selected, the anonymity and confidentiality of responses, and what
will be done with the data after the project is completed. The cover
letter should also inform recipients that they are in no way bound
to complete the survey, yet assure them that completion should not
cause them any undue stress (should that indeed be the case).
Finally, the cover letter should express the importance of attaining
as many responses as possible, thank the recipient genuinely for
responding, provide contact information for the principal investi-
gator, and notify the recipient of any reward associated with being
sampled or specifically for completing the survey form. Though a
considerable amount of research has been done on the effective-
ness of providing nominal monetary rewards or coupons with sur-
veys, the results have been mixed. Reward should be included if
the project budget allows for it.

Multiple Mailings

More effective than nominal monetary rewards are the use of mul-
tiple mailings. In a series of procedures now commonly referred to
as the “Dillman” approach, the investigator first issues a mailing
comprised of a cover letter, the survey instrument, a self-addressed
postage-paid return envelope, and a reward or token for the recipi-
ent, should there be one budgeted. Within one week after the ini-
tial mailing, a reminder postcard is sent to everyone in the sample,
thanking those who have already responded and resoliciting
responses from those who have not. Two weeks after the postcard
reminders are sent, the investigator mails a letter and replacement
questionnaire to nonrespondents. This mailing would again
include the questionnaire, a revised but brief cover letter, and a
self-addressed postage-paid return envelope.

Tracking who has responded and who has not can be easily
accomplished by coding the survey or the return envelopes. If an
in institutional review board has not accepted any coding procedures
or if the investigator simply chooses not to use any, then the sec-
ond distribution of the survey instrument can simply be mailed
out again to the entire sampled population with an explanation in the
cover letter that those who have already responded may disre-
gard the request and discard the survey instrument.

Four weeks after the second mailing, a final mailing similar
to the second mailing is sent, but this time by certified mail to
emphasize its importance. In addition to all of these procedures,
some investigators (such as the U.S. Census Bureau) send a brief
letter before the first mailing alerting recipients that they are
soon to receive a survey questionnaire and urging them to par-
ticipate in the research project.

An important consideration is the survey’s length. A survey that
is too long creates a burden on respondents to complete it. The
researcher should have a clear picture of what data are needed in
order to test the research hypotheses. It is often tempting to tack
additional questions onto the survey because, “Well, we’re going
through the trouble of conducting this survey. Why not ask them?
We may need that information some time in the future.” Persons
selected to receive the survey often attempt on some level to
understand the purpose of the questions comprising the survey.
Potential respondents might become disengaged or even offended
if questions appear frivolous.

A survey can become lengthy even if only the necessary ques-
tions are asked. Even some information that was deemed nec-
essary may have to be forfeited if the survey becomes too long.
Understanding the population will allow the investigator to
adequately estimate how long filling in the survey can take; typi-
ically, respondents should not be asked to spend more than
15 to 20 minutes to complete the questionnaire.

Demographic, Personal, and Practice-Related Questions

There are typically several components to any survey. Aside from
the item questions that focus on the attitude or behavior of inter-
est, there are questions soliciting personal information about the
respondents, such as age, gender, racial/ethnic background, level
of education, and level of household income. The investigator may
also be interested in obtaining other types of information, depend-
Crafting these questions might be regarded as the “easy” or unscientific part of constructing the survey, but if it is done poorly, problems may develop, such as ambiguous responses and poor rates of return. The first issue confronting the investigator is whether to employ open-ended or closed-ended question formats to solicit the information. Open-ended questions are easier to write but the answers are much harder to interpret and code for analysis. Open-ended questions are useful when: (1) respondents’ own words are essential to answering the research question, (2) respondents are willing and able to provide these answers, (3) the research is exploratory and the range of possible answers is unknown, (4) one of the investigators or someone else involved with the project has the ability to interpret, code, and analyze the responses, and (5) there is a preference for reporting the data as grouped responses rather than using statistical procedures.3

Investigators will usually use closed-end questions to solicit personal and demographic information. Two very important standards for writing these types of questions are the exhaustiveness and mutual exclusivity of the response choices. Exhaustiveness implies that the response choices adequately cover the possible range of answers. For example, an investigator surveying a population of pharmacists who is interested in their primary work setting would offer response choices that include “community-independent,” “community-chain,” “supermarket,” “hospital-distributive,” “hospital-clinical,” “long-term care,” “managed care,” and “pharmaceutical industry” at the very least. While it is appropriate and even advisable to include an “other” response category to cover the entire range of choices, if too many respondents feel they have to select “other,” the investigator loses valuable data.

Mutual exclusivity implies that the response choices are distinct; they do not overlap. Should the choices not be mutually exclusive, the respondent will not know which to choose as the answer to the question. Consider the following example:

Q. Which of the following best describes your favorite activity?
   a. Outdoor
   b. Hunting/fishing
   c. Hiking
   d. Reading
   e. Sports

Outdoor, hunting/fishing, hiking, and sports are not mutually exclusive so a respondent may have difficulty selecting the most appropriate choice.

Investigators may want to consider using open-ended questions to solicit exact age or a length of time; when a specific figure is entered, the data are richer and may allow the researcher to conduct more powerful statistical tests. For example, if age is categorized into groups such as 30–39 years, 40–49 years, etc., a respondent who is 39 years old chooses the same answer as someone who is 30 years old, and likewise for the 40– and the 49–year old. Even though the 39- and the 40-year-old are separated by only one year, they will be grouped in analyses similarly with others who are farther from their age.

It is best to take care, though, when soliciting data that may be sensitive in open-ended questions. For example, respondents may be more comfortable reporting a range for household income rather than a specific figure. In fact, respondents may not be aware of the exact figure. In general, the investigator should consider looking at how previous researchers have framed demographic questions as well as how the U.S. Census Bureau frames its census questions.

Developing Items for a Summated Ratings Scale

Most researchers use summated ratings scales rather than interval-level or multidimensional scales. Summated ratings scales are a list of item statements to which respondents indicate some level of agreement or other affective response on a numeric scale. The items comprising an attitude scale should represent the universe of existing stimuli to that referent object. In other words, researchers need to consider all aspects of a phenomenon that may constitute a person’s attitude towards it.

For example, an investigator interested in gathering customers’ attitudes about a community pharmacy should query respondents on everything about the pharmacy that evokes an affective response. This would include, but not necessarily be limited to, aspects of the pharmacy layout/design, the selection of available products, arrangement of the products, prices of product lines, the courtesy and competence of the employees, the aesthetic appearance of the pharmacy, its location, hours of operation, the professionalism of the pharmacy/prescription department staff, the speed and accuracy of prescription dispensing, counseling services offered by the pharmacy, and the ability of pharmacy staff to solve both medication- and insurance-related problems. Most of the time, because the various aspects of the referent object under study are not as obvious as the ones in the hunting/fishing example, drafting the items requires a considerable investment of time and energy.

There are at least three sources to be considered when identifying domains of the referent object under study. First and most important is the available literature, particularly primary sources. A thorough literature review will give an investigator considerable insight and inevitably bring to light certain approaches to the topic not previously considered. The review might also make it clear that the investigators’ goals are too ambitious or that the topic does
not lend itself well to investigation. On the other hand, the investigator may find a previously validated tool that measures the referent object more reliably and thus avoid the burden of creating an entirely new survey instrument.

The literature search must be comprehensive. For example, if an investigator is interested in identifying health beliefs, the literature review cannot be confined to merely plugging the terms “health” and “beliefs” into a search engine and seeing what comes up. A thorough literature review should take days or even weeks to complete. The initial queries should search for a comprehensive set of terms in a variety of databases; the investigator should read the initial abstracts and articles discovered and then reiterate the process with additional terms generated from the initial search. Novice researchers may confine their search to one database such as International Pharmaceutical Abstracts, Medline, or Micro-Medex but, depending upon the nature of the topic, other databases from the areas of psychology, business, law, education, nursing, and medicine, along with the Internet, should be utilized and may prove to be invaluable.

A second source of information is colleagues and experts in the field. The ideas that can be generated from a few discussions with colleagues (whether they are fellow practitioners, administrators, or academics) can be pleasantly surprising. Not only will their insights offer a different perspective, but the actual process of discussing the matter with peers will also be quite stimulating, helping to “grease the wheels” of the project. The literature review may have identified experts on the same or a similar subject with whom the investigator is not personally acquainted; more often than not, these experts are more than happy to share their insights and experiences. While experts are likely to have published extensively on the subject, discussions with them may reveal information not yet in print or provide the same sort of stimulation as is derived from talking with colleagues.

A third source, one that is often overlooked, is persons who meet the sampling criteria. In referring to the previous example concerning customers’ attitudes toward a community pharmacy, gathering information from actual customers can be valuable in identifying pharmacy attributes that matter to them. An investigator interested in measuring pharmacists’ attitudes and experiences with communications generated by direct-to-consumer advertising of prescription drugs should consult with practitioners when creating the survey. Gathering this type of information may simply involve interviewing a small number of persons chosen on the basis of convenience. A more formal and potentially rich data source for soliciting domains of the referent object is to use focus groups or nominal group techniques.

Once all of the relevant domains have been identified, the next step is to construct the actual item stimuli. It is usually necessary to construct several stimuli for each domain. For example, in the example on attitudes toward a pharmacy, several stimuli will be required to get at the “aesthetic appearance” domain. Consumers judge a pharmacy’s appearance by its colors, lighting, merchandising, and signage, at the very least. Generating more than one item for each domain also allows the investigator to test the validity and reliability of the stimuli comprising each domain and the domain itself (discussed further in the next section).

Cautions investigators should heed when constructing items for a scale are summarized in Table 1. An example of a vague item question would be “How are things coming along at your job?” “Coming along” is ambiguous. The respondents do not know whether this refers to a general level of satisfaction, progress on certain projects, or how well their most recent performance evaluations went.

Some common words have multiple interpretations, such as the word fair, which can mean “just,” “equitable,” “impartial,” or “not very good.” In surveys related to managed care pharmacy, words like “quality” and “access” should be avoided in items unless they are further clarified.

Basing words and phrases are those that elicit emotional responses that have little to do with the referent object. The investigator has to be careful when using terms like “abortion rights,” “socialism,” or even “capitalist” when designing survey items. The objective investigator will resist the temptation to charge up respondents, who may otherwise end up responding more to the survey itself than to the referent object. Similarly, investigators must be careful not to elicit socially desirable answers from respondents. For example, in inquiries about a pharmacist’s medication counseling behavior, “I counsel as often as other pharmacists,” or “I follow the procedures dictated by OBRA 1990” almost forces the responding pharmacist to provide an affirmative response. In cases like this, it would be better to present one or more scenarios and ask respondents how often they engage in specific behaviors.

It may be appropriate on rare occasions to lead respondents,
Constructing Mail Survey Questionnaires

particularly if there is a concern that they may provide socially desirable answers. For example, in assessing medication compliance, one might consider informing potential respondents that “many people often forget to take their medication as directed,” before asking them how often they forget to take their own medications. This technique should be used judiciously, however, as the results of leading questions may be suspect.

A very common mistake, even among experienced researchers, is to construct compound or double-barreled items. For instance, if you were asked to indicate on a scale how much you like pizza and hot dogs, you might not know how to respond. What if you love pizza, but have no appetite for hot dogs? A double-barreled question puts the respondents in a quandary, resulting in their either leaving the question blank, responding to only one of the two stimuli, or even averaging their response to the two stimuli. The remedy is simply to divide the item into two separate stimuli.

Another common error is the use of superlatives. If a respondent is asked to evaluate something like “This is the best pharmacy plan I have ever had,” it creates problems on several fronts. For one, words like “best” and “worst” are ambiguous. Secondly, the respondents may have too little experience with other pharmacy plans to make a comparison. Finally, some respondents may be relatively satisfied with the current plan but perhaps at one time had a plan that they were even more satisfied with. These respondents would have little choice but to disagree or strongly disagree with the statement despite being satisfied. Similar problems result when descriptors such as “very” and “quite” are added to items.

Investigators should consider reversing the effect of some item statements. For example, if most of the stimuli represent something negative about the referent object, inject a few items that point to something negative about it. An example of a negatively worded item on a scale designed to measure a pharmacist’s job satisfaction would read, “I do not feel as though I have an opportunity for advancement in my current job.” A respondent who is agreeing or disagreeing with all of the statements, good and bad, may not really be paying attention to the stimuli, perhaps warranting that the responses be discarded. If this is happening frequently, the investigator must question the validity of the scale, the directions provided to respondents on how to complete the survey, and perhaps the sampling procedures.

Suggestions #10 and #11 in Table 1 go hand in hand. They will help ensure the validity of the responses obtained. For example, where an investigator is concerned with attitudes and behaviors resulting from direct-to-consumer (DTC) prescription drug advertising, an item such as “Because of DTC ads, more people request specific prescription medications by name” induces speculation by the respondent on the habits of other individuals. The item should be written in a manner that solicits information about a personal consequence and reflects a potential resultant behavior on the part of the respondent. Rewritten, “Because of certain DTC ads, I am more inclined to request a specific medication by name,” the item requires the respondents to be knowledgeable only about their own intentions.

Similarly, asking someone to rank the “convenience” of a pharmacy is less effective than asking them to respond to an item stimulus such as, “Pharmacy X’s hours of operation are convenient for me.” The latter is not only more clear but can provide more specific information to discern preferences among different types of persons (by gender, occupation, age, or other variable).

Investigators may want to elicit from respondents information about a frequency of behavior or give them a timeframe for which answers to certain questions apply, for example, a general state of health or level of stress. A period that is too short, such as “yesterday,” is transient and does not necessarily indicate general health or stress if the respondent simply felt under the weather. On the other hand, asking respondents to provide information for too long a period, such as “the past year,” results in their having to engage in significant recall and make a judgment about what were probably varying degrees of health or stress. Given that sampling frame and research objectives differ from one study to the next, there is no universally accepted timeframe but periods such as “the past 30 days,” “the past four weeks,” or “the past two months” are often appropriate and may yield more reliable responses.

Conclusion

Data obtained from mail survey questionnaires can be very helpful to administrators as they make decisions. Decisions based upon a poorly constructed survey questionnaire, however, can be more deleterious than those based upon no information at all. Although survey research can be quite complicated, avoiding a few common pitfalls will enable the investigator to have more confidence in the results obtained. This continuing education article should help administrators and researchers to construct survey questionnaires that maximize rates of return and produce valid and reliable results.

REFERENCES

Upon completion of the continuing education module, pharmacists will be able to:
1. Identify situations for which surveys may be utilized to collect data.
2. Identify important research design and sampling considerations in survey research.
3. Design an appropriate cover letter to accompany a mail survey that maximizes its rate of return.
4. Select an appropriate scaling procedure.
5. Describe methods used to assess the validity and reliability of a summed ratings scale.
6. Describe procedures used to determine the potential for nonresponse bias in mail surveys.

**SELF-ASSESSMENT QUESTIONS**

1. A higher rate of return to a mail survey questionnaire is important because:
   a. It gives the investigator more confidence that the results typify the entire population.
   b. It reduces the likelihood that respondents differ in some way from nonrespondents.
   c. It may allow the investigator to use more powerful statistical procedures.
   d. All of the above.

2. Which of the following is the most effective way to improve a mail survey's rate of return?
   a. Using small font size to keep the questionnaire from becoming too lengthy.
   b. A nominal reward.
   c. Multiple mailings and reminders.
   d. Enhancing the questionnaire with graphics and illustrations.

3. Which of the following is NOT appropriate for inclusion in a cover letter accompanying a mail survey?
   a. Assure survey recipients that their responses are anonymous and will remain confidential.
   b. Warn survey recipients that the project will fail if they do not provide a response.
   c. Inform survey recipients of the benefit that will be derived from the research.
   d. Inform survey recipients of how their responses will be used and what will be done with them when the project is completed.

4. Which of the following statements is true?
   a. It is never appropriate to elicit personal or demographic information with an open-ended question format.
   b. Survey respondents typically enjoy the use of catch phrases and jargon in survey questions.
   c. Response choices to multiple-choice questions should be mutually exclusive and exhaustive.
   d. The use of open-ended questions is ideal for research that involves sophisticated statistical procedures.

5. Mutual exclusivity among response choices implies that:
   a. The response choices are distinct and do not overlap.
   b. The response choices cover the entire range of possibilities.
   c. The response choices do not offend any particular group of persons.
   d. The response choices are somewhat difficult to comprehend.

6. Which of the following is true?
   a. Items in a scale elicitating attitudes toward a referent object should be representative of the universal set of attributes possessed by that object.
   b. With enough practice, a thorough review of the literature should take only a few hours.
   c. A literature review should typically be confined to one database in order to ensure efficiency and accuracy.
   d. All of the above.

7. Which of the following are appropriate as resources for deriving item stimuli comprising a summed rating scale?
   a. The opinions of colleagues.
   b. Focus groups of persons meeting the sampling criteria.
   c. The primary literature.
   d. All of the above.

8. Which of the following would be good advice to an investigator constructing item statements to comprise a summed ratings scale?
   a. Avoid the use of double-barreled statements.
   b. Avoid the use of superlatives.
   c. Carefully intersperse items that require reverse coding.
   d. All of the above.

9. Suppose as a pharmacy manager you were designing a questionnaire to gauge the level of job satisfaction of pharmacy technicians. Which of the following item stimuli would be most appropriate?
   a. “My job is arduous.”
   b. “I have the best job of anyone I know.”
   c. “My job stimulates me to think and pays me well.”
   d. “My job provides me with an opportunity to use my own judgment.”

10. Suppose you are interested in assessing consumers’ attitudes toward switching from prescription to OTC antibiotics. Which of the following item stimuli is best? (Assume each item stimulus begins with “Allowing antibiotics to be sold over the counter . . .”)
    a. “. . . is a very bad idea.”
    b. “. . . will help me to afford to buy antibiotics for my family when needed.”
    c. “. . . will increase the susceptibility of high-risk individuals to drug-resistant microbial strains.”
    d. “. . . will increase access to antibiotics for certain people.”
11. In what type of setting do you work? (Leave blank if none of the responses below applies.)
   a. HMO
   b. PPO
   c. Indemnity insurance
   d. Pharmacy benefits management
   e. Other

12. Did this program achieve its educational objectives?
   a. Yes   b. No

13. How many minutes did it take you to complete this program, including the quiz? (Fill in on answer sheet.)

14. Did this program provide insights relevant or practical for you or your work?
   a. Yes   b. No

15. Please rate the quality of this CE article.
   a. Excellent   b. Good   c. Fair   d. Poor

### Instructions

This test affords 1 hour (0.10 CEU) of continuing pharmaceutical education in all states that recognize the American Council on Pharmaceutical Education. To receive credit, you must score at least 70% of your test answers correctly. To record an answer, darken the appropriate block below. Mail your completed answer sheet to: Academy of Managed Care Pharmacy, 100 N. Pitt Street, Suite 400, Alexandria, VA 22314. If you score 70% or more, a certificate of achievement will be mailed to you within eight weeks. If you fail to achieve 70% on your first try, you will be allowed only one retake. The ACPE Provider Number for this lesson is 233-000-02-003-H04. This offer of continuing education credit expires June 30, 2003.

### CE Exam

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>8.</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>2.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>9.</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>3.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>10.</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

11. ☐ A ☐ B ☐ C ☐ D ☐ E
12. ☐ Yes ☐ No
13. Minutes ________
14. ☐ Yes ☐ No
15. ☐ A ☐ B ☐ C ☐ D

---

**Participant Identification:** Please type or print.

**Social Security #:**
**Date:**
**Work Phone #:**

**Name:**
LAST  FIRST  MIDDLE

**Company:**

**Address:**
STREET (with Apt. No.) or PO. Box  CITY  STATE  ZIP

**State & Lic. No.:**
STATE  LICENSE NO.

**Member Type:** ☐ Active ☐ Supporting Associate ☐ Student ☐ Nonmember

**Signature:**

I verify by my signature above that I have completed this examination independently.

**Note:** If you wish you can now take this CE Exam online. Go to www.amcp.org.