Editing and Crafting Good Survey Questions and Questionnaires

Version 2.0

Timothy Triplett
Urban Institute

Acknowledgements: I have had the great fortune of working for managers that encouraged me along the path of studying and learning survey methods. This started with John Robinson, Stanley Presser, Bob Groves, and Johnny Blair while working at the University of Maryland. Continued with Kevin Wang, Olivia Golden, and Rob Santos at the Urban Institute (UI). Also, I would like to thank my close colleagues Adam Safir, Natalie Abi Habib (Vaughn), Sharon Long, Francie Ostrower, and Doug Wissoker. Special thanks to Harry Hatry for choosing me to author a chapter on using surveys in the "Handbook of Practical Program Evaluation and Brett Theodos who suggested that I should author this short book.

Contents

Introduction

Survey researchers will tell you that writing survey questions and crafting a survey questionnaire is often the most challenging and most important task in designing a survey research project. The challenge comes from the fact that it is difficult to write questions that avoid potential bias or ambiguity. The importance of questionnaire wording can easily be overlooked in studies that involve a complex survey design, yet a poor survey instrument will doom even the best designed survey research study.

A big challenge in designing questions is that researchers who are responsible for developing a survey questionnaire might have limited training in questionnaire design and are not specialists. Although, researchers are increasingly working with survey practitioners, which has led to more emphasis on improving question wording. As a survey practitioner, I often help with projects that require drafting a survey questionnaire. The discussion on questionnaire design in this document comes from my own experiences combined with what I have learned from the other researchers who have published work around the methodology of writing good survey questions.

Chapter1: Shorter surveys, shorter questions, and shorter sections

Shorter surveys

Convincing a person to participate is usually the most challenging and expensive part of a survey. Because the effort to get people to participate is so demanding, why settle for a short survey? The simple answer is that people who are asked to complete long surveys dislike them and feel that they are burdensome. The obvious downside of fielding a long survey is that they often lead to a lower survey response rate. But there are other ways in which long surveys damage the quality of survey data (Herzog & Bachman, 1981) (Galesic & Bosnjak, 2009). These include, but are not limited to, a greater number of survey breakoffs, increased item nonresponse, lack of cognitive effort put forth by the respondent and a lower likelihood that the respondent will take part in future surveys.

Shorter questions

Long questions are synonymous with writers who write long or run-on sentences. That is why they often confuse the reader and with interviewer administered surveys long questions often require the respondent to ask that the question be repeated. The goal of any well written question is clarity. The longer the question the harder it becomes to write a question that is concrete and to the point.

Respondents when filling out surveys are often not fully engaged or multi-tasking so that even beautifully written long questions can be problematic. Longer questions often lead to recall errors because respondents are not always fully focused. For example, respondents might be answering questions without considering all the information provided. Details that a respondent needs to accurately respond could get lost when the question is too long.

Look for simple edits that can shorten the question without changing the meaning of the question. Often people add needless phrases to their questions that obscure the key information rather than clarify. For example, you can always cut out the phrase “In your opinion” because a survey is usually all about a person’s behaviors or opinions.

Shorter sections

A long section of questions is another thing to avoid when constructing a questionnaire. A respondent’s focus and cognitive effort tends to trail off after they start seeing or hearing the same kinds of questions. Thus, responses to questions that get asked later in the section often have some bias since respondents are not putting as much cognitive effort in responding to them. This is one of the reasons why survey practitioners like to randomize the order that questions in a section get asked. While randomization may help spread the bias among all the questions in the section it does not fix the problem.

What you can do is break longer sections up into shorter sections. For example, if you want to ask twenty satisfaction questions consider dividing them into four or five sections, putting related items in the same section. Then you can use a transition statement "*Now I have some questions about …*" to help the respondent refocus their thoughts. Also, respondents tend to stay more focused when the questionnaire is comprised of a variety of questions. Although it might seem simpler to write questions that share the same response options, this can become mundane for the respondent leading them to put forth less cognitive effort. Changing the style of the questions and response options helps keep the respondent alert. However, changing response options can be challenging when designing interview administered surveys. Because the respondent cannot visually see the questions and must be able to recall what the options are.

Chapter2: Using appropriate words and avoiding jargon

Know your target population

The important first step in writing good survey questions is keeping in mind who your target respondents are. Characteristics of target respondents are especially pertinent to survey layout and question considerations. Think about how receptive the respondents will be to being surveyed, how much they know about the subject matter, and how sensitive they may feel about the questions asked.

The educational background of the target respondents will affect the complexity of the terms you choose to use in the survey. Writing questions directed at scientists and engineers presents different challenges from writing questions directed at recipients of food stamps. Other demographic characteristics of the target respondents are also pertinent here, such as their age and primary language spoken.

The receptivity of the target respondents to answering questions about the issues may also show whether they have the desired information, how easy it is for them to get the information requested, and how willing they are to part with the information. You need to think through any assumptions you develop about the knowledge or experience the target respondents have that is relevant to their ability to answer the questions. For example, evaluators may assume that target respondents are more familiar with service delivery procedures or with an agency's acronyms than they may be. Or evaluators may be too optimistic in thinking that respondents can provide reliable information about their experiences, such as reporting how many hours they watch television. Memories are fleeting, and respondents' abilities to recall experiences or impressions may not be enough to provide accurate data.

Same meaning for everyone

Survey questions need to convey the same meaning to all respondents. This can be challenging when the characteristics of your target respondents vary. For most surveys, there will be differences in respondents' age, education, gender, race, and ethnicity that can make it difficult writing questions that will be interpreted the same by all. Even with similar respondents it is not always easy to write questions that everyone will interpret the same. For instance, consider the question *"Was the food at last night's dinner too hot?"* - Many people are thinking about temperature, but others could be thinking that the food was too spicy.

What seemed like a standard question *"Are you driving right now?"* is often asked to make sure the respondent is not trying to complete a survey on their mobile phone while driving. When this question was asked in a survey of refugees some respondents interpreted this to mean can I drive and answered yes, thus ending

the interview, although they technically were not driving right now. A survey that was asking about people's reading habits found that some respondents who said that they read novels and short stories then said that they did not read any books because they did not consider reading a novel on an electronic device to be the same as reading a book.

To avoid writing questions that mean something different to different people the survey practitioner often cognitively tests the questions with diverse types of respondents. Cognitive interviewing is a qualitative process of studying the mental processes experienced by individuals. When employed as part of survey pre-testing, cognitive interviews can help evaluators better understand how respondents interact with an instrument and identify potential issues (Willis, 2005). Cognitive interviewing detects both overt and covert issues, including hidden subtext and questions which have more than one interpretation. Moreover, cognitive interviewing can help determine the prevalence and severity of any errors.

Avoid acronyms and verify that jargon is understood

When people author a report, they often use acronyms, but the first time the acronym is used it is usually shown in parenthesis right after what it stands for is written out. This practice does not carry over to writing a questionnaire as respondents should never be required to recall what an acronym means. So, a survey instrument should almost never include acronyms. The few exceptions would be when the acronym is so common that people will be more familiar with the acronym than what it refers to. Even then, it may be better to include both the acronym and what it stands for.

Besides acronyms, other things that will get you in trouble include using unfamiliar words, names of programs or industry jargon. Be careful when getting advice from people who are too actively involved with the subject matter. Keep in mind that your goal is to write questions that everyone will understand. For example, when writing questions for a membership or customer satisfaction survey it is common for people to assume that the target respondents will know program names or industry terms when in fact, they don't know what they mean, or they have a different understanding of them. Also, because of small sample sizes, pretesting or cognitive testing may not reveal that your question wording or industry terms are not universally understood. If in doubt, it is best to define programs or industry terms for the respondents so that everyone will have the same understanding of the concept.

Think about people new to the target population

The target population often may include new members who lack the information needed to answer the questions you would like them to answer. For example,

people new to a community probably can't evaluate services that they are receiving and someone who just joined an organization will often need more time to adequately report their satisfaction with the organization. Interviewing people new to a target population can usually be solved by assigning proper skip patterns, but sometimes it will require similar but different questions be asked of the people new to the population being surveyed. What is often done but not recommended is relying on don't know or not sure response options to handle the problem of people too new to respond appropriately. In general, it is never good ideal to ask people questions that are likely to elicit don't know or not sure responses, chapter 5 and chapter 10 provide more details on why you should usually avoid offering a don't know response option.

There are situations where it is important to interview a person new to a target population to establish a benchmark. These benchmark responses can later be compared to responses from the same person after they become a more established member of the community or organization. When writing these benchmark questions, you need to explicitly define the time frame in which you want the respondent to be thinking about when responding to your questions.

Language barriers and best practices for translating questions

If you plan to administer your survey in more than one language it is imperative that your study design has adequate resources and time for translation to be done properly. Proper translation should involve at least two separate translators for each language. The translators should write separate translated instruments and then work together on creating a single combined instrument. In many studies the respondents who are likely to fill out the translated questionnaire will come from different countries that share the same language. In this situation, it is best to choose translators that will reflect the diversity of your respondents.

Assuming there are adequate resources, it is strongly recommended that you use more than two translators on studies where there will lots of respondents using the translated instrument. Since this will lead to more than two translated versions per language you need to develop a review plan for drafting a final translated instrument. This would entail a team meeting where differences between the translated instruments can be discussed and coordinated into a single-translation instrument. Usually, one person will be assigned the role of being the adjudicator who will verify that the final translated instrument is either ready for pretesting or main data collection. If the final translated version of the instrument is used for a pretest, then revise and re-adjudicate the translation based on the pretesting results.

Lookout for terms that have changed meaning over time

One of the first steps in designing a survey instrument is to search for other studies that might have already asked some of the same questions you would like

to ask. This is a good first step because you are often able to find questions that were tested and used before. But don't assume that questions from other surveys are good questions. Even the best studies continue to ask questions that are no longer relevant, or the interpretation of the question has changed over time. A good example is from the Survey for Participation in the Arts (SPPA) that asks questions about attending live art events. The problem with using older SPPA questions is that what constitutes a live performance and what is considered art has changed over time. Questions that ask about the use of technology often become quickly outdated as people are always finding new ways of communicating, interacting, and gaining access to media. Again, think about your target population, will they all have the same understanding of the terms used in the question being asked. Also, look for these sorts of problems during the pretest or during cognitive interview testing by asking people about how they came up with their response to the question.

Avoid using positive or negative adjectives

Adjectives can enhance text so that it is more interesting, but they usually create a positive or negative slant. A positive or negative slant should be avoided when writing survey questions since they may encourage or discourage a certain response. For example, it is better to ask whether a person "enjoyed the presentation" rather than whether they "enjoyed the funny presentation" (positive slant) or "enjoyed the long presentation (negative slant). Search through your question to find and remove adjectives is a worthwhile step to include in finalizing your questionnaire.

Importance of cognitive testing

Conducting cognitive interviews can help identify potential questionnaire problems described in this book. Cognitive interviewing is a qualitative process of studying the mental processes experienced by individuals. When employed as part of survey pre-testing, cognitive interviews can help evaluators better understand how respondents interact with an instrument and identify potential issues (Willis, 2005).

Evaluators can use different cognitive interviewing techniques depending on their objective(s). Two primary cognitive interviewing techniques are the "think-aloud" method (Ericsson & Simon, 1980) and verbal probing technique (Drennan, 2002). In the think-aloud method, participants complete the survey instrument while vocalizing their thought processes. This helps evaluators better understand how respondents will perceive the question and identify any misunderstandings (Ericsson & Simon, 1980). Evaluators using verbal probing technique administer the survey instrument to test subjects without commentary and then ask probing questions after the respondent has completed the instrument (Willis, et. al, 1999).

Evaluators can use cognitive interviews as a part of survey pre-testing to accomplish two objectives. First, evaluators might use cognitive interviewing to identify problems with the survey instrument (Collins, 2003). Cognitive interviewing detects both overt and covert issues, including hidden subtext and text which might have several interpretations. Moreover, cognitive interviewing can help figure out the prevalence and severity of any errors (Blair & Conrad, 2011).

Second, cognitive interviewing can help assess and reduce the likelihood of response error by looking at self-report capacity. For example, cognitive interviewing can help identify ways of phrasing questions which will provide the most accurate information. Using cognitive modeling in conjunction with cognitive interviewing, evaluators can better understand how respondents lose information and can develop strategies to lessen the impact of any losses on the validity of survey data collected (Willis, 2005).

While cognitive interviewing can help evaluators identify problems with the survey instrument and assess the likelihood of response error, cognitive interviewing has its own set of limitations. Cognitive interviewing is both expensive and time consuming. Additionally, the sample size needs to be large enough to detect issues with the survey instrument; without an adequate sample size, serious issues may escape detection (Blair & Conrad, 2011). Moreover, even when cognitive interviewing successfully finds a problem, it does not develop a solution (Collins, 2003).

Chapter3: Appropriate recall periods

Is the time frame clearly stated?

It is surprising how often people write recall questions that do not provide the respondent with a time frame which they need to formulate their response. The problem with asking a recall question without a specified time frame is that it leads to inconsistency because respondents will be forced to decide what time frame to use. All recall questions need to include a time frame. For instance, if you are asking how often or how many times somebody did something, then you need to specify the time frame (for example – in the past week, month, year, two years). Even attitudinal questions that ask about satisfaction, importance, or likelihood usually needs a time frame since people's attitudes change over time.

Avoid changing the time frame

Keep in mind that respondents are often multi-tasking or simply not fully focused on the survey questions. Therefore, when you change time references the respondents may not pick-up on the changes. For example, switching from asking questions about the past year to a question that asks about a person's most recent experience can lead to respondent errors. However, sometimes the question that you need to ask requires a different time frame. In this situation, it is recommended that you try to ask the new time frame questions in a different section of the survey or at least before asking the questions include a transition statement that helps enlighten the respondent of the change in the time frame. For example – "*For these next questions, we would like you to report only on things that occurred in the last month that is since April 12"*.

Suppose you wanted to know whether someone did something at any time in the last year and how many times they did this in the last month. Rather than first asking about the past year, you will likely get more accurate responses if you ask first if they did something in the last month and if they say yes then ask how many times. With this approach, you can still ask those who said no to doing something in the past month a question about whether they did this in the last year.

Is the event salient enough to justify the time window?

Keep in mind that memories are usually short-lived and such your recall time window needs to be reasonable. How salient the topic or event is to the target respondent is usually an effective way of thinking about your time frame. The more salient the topic or event, the better a person will recall doing something. For example, some people will not be able to recall what they were doing several days ago but can still remember where they were and what they were doing on

September 11, 2001, when terrorists drove planes into the World Trade Center buildings.

Consider possible respondent telescoping

Telescoping is when a respondent widens the recall time frame as if they were looking through a telescope and seeing events as being closer in time than when they occurred. Less common, but sometimes respondents could shorten the recall time frame as if looking through the opposite end of the telescope. The shortening of the time frame often occurs when people are asked to recall behavior or activities that carry a stigma. While socially desirable behavior or activities often lead respondents to widening the recall period. The downside is that either way you end up with over or under counts of the proportion of people behaving a certain way.

There are no easy solutions to preventing respondents from widening or narrowing the recall period. Though there is evidence that providing specific dates may reduce the telescoping bias (Morwitz, 1997) (Janssen et al. 2006). For instance, rather than simply asking about the past year, you could add "that is since June 20, 20xx …" Asking an open-ended follow-up question about reported events might help clue you into the problem of telescoping but does not provide a solution to the problem.

There are major surveys that rely on proxy responses including the Current Population Survey (CPS) which uses one respondent to collect labor force data on all the members of a household. Proxy reporting is widely used, so the accuracy of proxy reporting has been professionally researched (Tourangeau, Rips, and Rasinski 2000). We do know that in general proxy responses have the potential to be less correct than information gotten directly from sampled respondents. However, there is evidence that proxy reporting as opposed to self-reporting may reduce telescoping bias because there is less potential of over or under reporting socially desirable/undesirable activities (Triplett, 2010). For instance, on the 2008 Survey for Public Participation in the Arts the estimate on how often women exercise regularly was higher for self -reported data than proxy reported data (by way of their spouses). Similarly, the estimate of the hours' men watch TV was lower for self -reported data than proxy reported data (by way of their spouses).

Avoid not sure or don't know responses

There are many arguments for and against providing the respondent a not sure/don't know option. Though when asking recall questions giving the respondent the not sure/don't know option is likely to be detrimental to the quality of the data being collected. For respondents, recall questions are a cognitively challenging task. So, if allowed, the respondents will often not think

enough to formulate a thoughtful response. Better to get an approximation of what they did as opposed to accepting a not sure/don't know response. Offering a not sure/don't know on a recall question also has the potential to increase the likelihood that the respondents will gratuitously choose the not sure/don't know option on other questions asked later in the survey. Keep in mind that if a person truly does not know how to respond they can always volunteer a not sure response or skip the question. See chapter 10 for more discussion on the inclusion of a not sure or don't know response.

Chapter4: Meaningful and balanced options

Are response options balanced?

Response alternatives that are presented to respondents should have the same number of positive and negative choices. Without balanced alternatives, a question becomes a leading question that favors the side that has more options. Also, people who do not have strong feelings will look to choose a middle option believing that there is a true middle, which will not be the case if the alternatives presented are not balanced. Table 1 shows a list of balanced response scales. Note that the middle ground option is listed as optional and may be left out of the scale. See Chapter 10 for information on the appropriateness of using a neutral option.

Table 1: Examples of scales that provide balanced alternatives.

Very Satisfied	Satisfied	*Neither satisfied nor dissatisfied**	Dissatisfied	Very Dissatisfied
Very important	Somewhat important		Not important	Not at all important
Very helpful	helpful		Not too helpful	Not helpful at all
Very likely	likely		Unlikely	Very unlikely
Always	Often	Sometimes	Rarely	Never
Very concerned	Concerned	*Neither concerned nor unconcerned**	Not concerned	Not at all concerned
Very Useful	Useful		Not too useful	Not at all useful
Definitely	Probably		Probably not	Definitely not
Very easy	Easy	*Neither easy or hard**	Hard	Very hard

* Optional

Are the response options realistic?

Often when designing surveys, there is an attempt to keep using the same scales to avoid confusing the respondent. As pointed out in the first chapter this can lead to a respondent becoming less focused when they start seeing or hearing the same kinds of questions. But more importantly, you should never force the use of existing scales as this may make it harder for respondents to answer and harder for the researchers to interpret the responses.

Another key thing to consider when developing response options is that they need to be realistic options. Most respondents will assume the options provided are realistic. For instance, on a split ballot experiment where a random half of respondents were asked the same question about how many hours of TV they watch on a typical day but were given different response options.

On a typical day, how many hours of TV do you watch?

Version 1:
1. Less than one hour
2. More than 1 hour, but less than 3 hours
3. More than 3 hours, but less than 5 hours
4. More than 5 hours

Version 2
1. Less than 30 minutes
2. More than 30 minutes, but less than 60 minutes
3. More than 60 minutes, but less than 120 minutes
4. More than 120 minutes

The results of this experiment were that the two different versions of the question produced the same frequency distribution. This suggests that people assume the middle categories (which respondents most often select) are reasonable ranges for someone who watches an average amount of TV even though version1 implies watching three times as much TV. If you are asking people to report the frequency at which they do something it is better to avoid range if you are not sure what the reasonable ranges would be. For instance, for the TV example, you would be better off just asking how many hours you watch TV without providing response options.

Are you using the best possible response options?

Getting the question worded correctly is critical, but equally important is providing response options that will be good enough to characterize a respondents' views.

- If you are going to provide the respondent with a scale - ask yourself, "Is this the right scale or is there a better scale?"
    - For example: Suppose you were asking a series of questions about how satisfied a person was with an event – this may include asking how satisfied were they with the length of the event – A much better question would be to ask if they thought the event was too long or too short or about right since their satisfaction would not provide information about why they were satisfied or dissatisfied (could be because it was too long or too short).
- Scales can be too short to fully exhaust the range of possible answers.
    - For example: Suppose you had a list of 8 to 10 items that you want people to prioritize in terms of importance. You might consider a larger scale so that the respondent has more flexibility in assigning different scores to things that are important (or unimportant) but not equally important.
- Scales can be too long, especially when interviewing less educated respondents by phone. Shorter scales are needed for phone surveys since the respondent cannot visually see the scale but must remember each option. In general, the scales can have more items if your target population is a highly educated group such as teachers or scientists (Younas & Porr, 2018). A good indicator that your scales may have too many points is when respondents are asking the interviewer to repeat the options or provide an answer that is not one of the scaled options.
- Scales can have the wrong labels; this is an increasing problem in the age cutting and pasting text. Reviewers usually spend their time determining whether the question is worded correctly and less time ensuring that the scale options are correct. Again, cognitively testing your survey should help discover any scales that respondents had difficulty using.

The next chapter talks about having mutually exclusive and exhaustive options. However, even with mutually exclusive and exhaustive options, your question might benefit from using better response options. For instance, suppose you were asking conference attendees to select which of the meals served at the conference was their favorite. Including all the meals as response options would mean that the question was mutually exclusive and exhaustive. However, simply providing dates and times for the meals would be inadequate compared with providing response options that provided an unbiased description of the meal to help the respondent recall the meals that were served.

Chapter5: Mutually exclusive and exhaustive options

Do any categories overlap?

You never want to put the respondent in a situation where there is more than one possible answer. This seems obvious, yet it is easy to find surveys that include questions that make this mistake. Usually, the problem occurs because you have given the respondent options that overlap. For instance, consider the question:

Where did you first hear about the festival? Was it from a:

1) Family member
2) Co-worker
3) Friend
4) Radio or TV advertisement
5) Newspaper advertisement
6) Site on the Internet
7) Other – Specify _____

For many respondents, this question will be hard to answer; friends or family members could also be co-workers; what if the advertisement was referred to you by a friend, co-worker or family member; cousins could be family members or friends; you can watch TV or listen to radio on the Internet; what if your friend told you to check a certain web site. Clearly this question has overlapping categories. In addition, this is a challenging recall question because a person may not even be sure when they first heard about the festival. My recommendation would be to first pick the categories you are most interested in learning about. Then ask the respondent about whether they got information about the festival from each of these potential sources of interest.

Try to avoid asking people to check all that apply

There are lots of surveys that contain check all that apply questions. In general, checking all that apply questions should be avoided. Often you get fewer positive responses than you would if you required the respondent to either check positively or negatively. This is especially true for response options that appear towards the end of a lengthy list. In addition, it is often incorrect to assume that a response item that is not checked means they did not do it. It could be that the option is not applicable, don't know or they did not check it because they felt that they had already checked enough options.

Research (Smyth et al. 2008) comparing questions formatted as check all that apply versus forced-choice concluded that the forced-choice questions format encourages deeper processing and therefore more optimal response behavior.

This held true across all survey modes - web, paper, and telephone surveys and mixed-mode surveys. A strong recommendation is to try and turn check all that apply response options into a format that requires the respondent to answer yes or no for each option. This will at least increase the likelihood that the respondent is reading or hearing all the response options.

Can the respondent easily find a response option that fits?

When respondents cannot find a response option, then the problem might be that there is a missing skip pattern because they should not be getting asked the question. More likely the response options are not exhaustive, so a person feels that none of the response options are appropriate. Assuming the respondent's response is unusual, then an easy solution is to include and other or something else option. However, if respondents are choosing the other response option, then there is a good chance that an important option is not being offered to the respondents.

Another reason respondents cannot find a response option is that they may not understand or be defining the options differently than you anticipated. Take this simple example: *Is your house or apartment -*

1 Owned
2 Rented
3 OTHER (SPECIFY) _____

Changing Option 1 to "Owned or being bought by someone in the household" would be a better response option for two reasons. A person with a mortgage may not consider their house as being owned yet, and secondly you probably want to count it as being owned if they are living with the person who owns the house.

Are there too many options?

Research (Krosnick and Alwin, 1987) has shown that when a respondent is given a lengthy list of response options, they tend to remember first option (primacy effect) or last option (recency effect). Most respondents cannot remember more than 4 or 5 response options. This is especially true in a phone survey. However, even on a self-administered survey, respondents often avoid fully reading all the response options when questions have more than four or five options. Also, the problem becomes worse when the response options themselves are wordy.

In some situations, you might be able to avoid too many response options by splitting the question into two or more questions. Another good strategy is to only provide the three or four most common response options and use an "Other Specify" response option to pick up the less common responses.

Should you randomize the response options?

If you think that the order of your answer options might influence the respondents' answer, then randomizing the response options makes sense. The two most common forms of response option bias are:

Primacy bias - Primacy bias is the tendency for respondents to pick one of the first options presented to them. This usually happens because it is the first choice they read or hear and agree with.

Recency bias - Recency bias is the tendency to pick an answer option presented at the end of a list because it is more memorable to the respondent.

If you suspect either of these may be occurring, then I would strongly encourage you to randomize the answer options so that each option has the same chance being at the beginning or end of the list of options. Even if you do randomize your response options, you still should be trying to avoid offering too many answer options.

Try to avoid offering a don't know or no opinion response option

In general, you should avoid offering respondents a don't know or no opinion response option. It could bias results as studies have shown that less-educated respondents are more likely to choose a no opinion or don't know response option. Also, offering this option increases the likelihood of respondents saying they don't know on other questions asked in the same survey (Tourangeau et al., 2000). Furthermore, studies have shown that including don't know/no opinion options does not improve the consistency of respondents' attitudes over time (McClendon and Alwin, 1993).

Cognitive testing has shown that for questions that require some thought, the don't know response option discourages respondents from thinking about the issue. Although when probed, respondents that chose the don't know option usually formulate opinions that lean in one direction or the other. Cognitive studies have shown that respondents who choose the no opinion or the don't know option, if encouraged, can provide substantive answers (Krosnik et al., 2002). Studies have also shown that respondents are more willing to select a not sure response option than a don't know option. Therefore, using a don't know response option is preferable to using a not sure response option.

Chapter 6: Consistency

Check that you are using consistent wording

Be. consistent and careful when writing a survey question, keeping in mind that the question wording needs to convey the same meaning to all respondents. Changing the wording may be a useful tool for authoring an interesting story, but when writing survey questions the wording needs to be consistent. For example, suppose you wanted to know about automobiles or the use of motor vehicles. People will attach different meanings to what defines an automobile versus a motor vehicle. A good approach would be to include an initial description that defines what you want, such as - "*Counting cars, vans, minivans, SUV's, pickup trucks, and motorcycles, how many vehicles are available for use by residents of your household?"* This question clearly defines what you would like the respondent to include.

Try to be consistent with recall periods

When asking respondents to recall information try to keep the recall period consistent. If the question requires a different recall period than previous questions, then make sure the respondent is easily able to pick-up on the change. For instance, include a transition statement such as "*Now, I want you to think about all the events that happened in the past 6 months that is since …"*

Consistent categories are usually helpful and sometimes detrimental

Respondents will have an easier time finishing a survey if the same response options are used. This is especially true for phone or interviewer administered surveys that require a person to remember the response options. So, consistent categories can help respondents. But the repetitive use of the same response options can lead to the respondent losing their focus. This is especially true for self-administered surveys. So sometimes it might be better to vary the response options though you need to keep all response options going from positive to negative or negative to positive as you do not want to try to trick the person.

Formatting should be simple, appealing and consistent

A variety of distinct types of questions can keep the respondent sharp, but changes to the questionnaire formatting are problematic. Formatting of the questions and questionnaire should be consistent, especially for self-administered surveys. When designing a self-administered questionnaire, it is important that you provide directions that help guide a respondent through the questionnaire. The respondent should be able to easily notice the difference between the directions and the actual questions.

Consistency is the rule - for instance be consistent with your use of capitalization, the formatting of tables, avoid changing fonts, do not switch between response options presented horizontally versus vertically, and keep open ended text boxes similar in size. The size of an open-ended text box might influence how a respondent responds and thus you should consider keeping them all the same size.

More than ever surveys are being done using various technological devices (i.e., computers, notebooks, laptops, smartphones, tablets, and so on) which require that you keep navigation through a questionnaire simple and consistent. Try to cut or eliminate the scrolling that a respondent must do by limiting scrolling to questions on the same topic. Online questionnaires should allow respondents to move backward to correct or answer an earlier question that they may have skipped or realized was answered incorrectly. Avoid forcing respondents to answer a question before moving on to the next question. A much better strategy is to remind people that they did not answer a question without forcing them to respond. The rare exception is for screening questions that you need to know to ensure they are an eligible respondent (for example if you only want female respondents - then they must respond to a question asking about their gender). Including progress indicators for on-line surveys can be useful, though they can be misleading with questionnaires that have major skip patterns.

Finally, the only way to get the formatting correct for a questionnaire that may be administered using different devices and browsers is to develop a thorough testing plan. Evaluate all potential devices and browsers with as many testers as you can find. Also, remember that not all respondents will be using their browser default settings, so try looking at your online test instrument using different browser settings. If a test reveals a problem, then you should not only fix the problem, but also schedule another test.

Chapter 7: Avoid leading respondents

Try to avoid using examples

Giving the respondent examples seems like an effective way to help respondents understand the question you are asking. The problem is that a respondent's answer will vary depending on what examples you decide to use. Thus, the examples can be seen or used as a way leading the respondent. So, try to avoid writing questions that have the phrases "such as…," "for example," or "for instance." Questions that include these phrases can usually be improved.

Consider the following question: During *the last 12 months, did you attend a live art performance? (Such as opera, ballet, or classical music performance)* For this question, the examples may help define art performances, but more likely they raise the bar of what should count as performing arts. Given these examples, it is not clear whether respondents will include theatre events, performances at art festivals, music concerts, or high school performances. Often when you feel the need to include examples, the best solution is to ask each of the examples as separate questions ending with a catch all question – "*During the last 12 months, did you attend any other live art performance – What was that?"*

Avoid using parentheses

A survey question needs to convey the same information to all respondents. So, parentheses can be detrimental since respondents may not use the information put inside the parentheses when formulating their response. This is the equivalent of allowing some of your respondents to answer a question that is worded differently. Parentheses can especially be a big problem on phones surveys where the interviewer may choose to read or not read the information put inside the parentheses. Though, parenthesis is still a problem in self-administered surveys as people have different approaches on how they deal with information put inside a parenthesis. The solution is that if you think someone will need information put inside a parenthesis then remove the parenthesis and include the information in the text of the question. For example, the following question

"*In the past 3 months, did you schedule any meetings in the human resources conference rooms (include the green, blue, or red room)?*

Should read:

"*In the past 3 months, did you schedule any meetings in the green, blue, or red human resources conference rooms?*

Using parentheses is all right for questions that repeat the same information from previous questions. For instance, if you are asking a series of questions about how

satisfied a person is with something, then using parentheses is fine after the first few questions in the series. So, after a respondent has already told you about other teachers, the following question format is fine: *(In the past 6 months, how satisfied have you been with) your math teacher (very satisfied, satisfied, dissatisfied, or very dissatisfied)*

Eliminate leading statements that might introduce bias

Researchers sometimes feel the need to add verbiage to questions to soften the question, so the respondent finds the question more palatable. For example, rather than directly asking whether someone went to church, or donated to charity they may start the question by saying we know everyone is busy or money is tight. The researcher is instinctively trying to prevent what is known as "respondent satisficing" that is the concern that people will want to report going to church or donating to charity even if they did not. However, there is no evidence that adding these leading statements improves the quality of data reported and the statements introduce potential bias. A bias that could potentially change how respondents answer subsequent questions. A better approach for dealing with respondent satisficing is to include follow-up questions about their reported behavior or activity that could be used to validate their response. For instance, asking a few questions about the last church service they attended. or donations given might result in the respondent changing their initial response or providing evidence that they did not go to church or make a charitable donation.

Also, trying to be politically correct can lead to unintentional consequences. In the 1970s there were several cross-burning events. Researchers wanted to know what percentage of the population felt favorable towards the cross-burnings but did not think people would admit to having this feeling. So, they padded the question with the statement "Some people feel the Cross-Burnings is protected by a person's first amendment rights - Do you have a favorable or unfavorable opinion of the recent cross-burnings that have taken place in Maryland?" This question included a follow-up question that asked why they felt this way. A person who reported an unfavorable opinion in a follow-up asking why said "because they thought the cross burning were not close enough to the house." So, in this example, a direct question that asked whether approved or disapproved of cross-burnings would have yielded a more accurate response.

Another leading statement that researchers often add to questions is to tell people how others have responded, so they are more comfortable responding the same way. Consider, this question – "Many people are saying that they are having trouble buying groceries, how about you are you having trouble buying groceries?" Again, this leading statement biases the question and will likely lead to overestimating the percentage of people reporting having trouble buying groceries.

Sometimes the question may not contain a leading statement, but the lead-in statement at the beginning of a section of the survey maybe biasing responses. For example, the lead-in statement “Next, I’ll be asking about some of recent programs that help low-income families” a better lead-in statement would be “Next, I’ll be asking about some of recent programs for low-income families.” In this example the term “help” implies that the programs are do something good and that could bias how people respond about the programs. Always try to keep lead-in statements neutral, including lead-in statements that are read at the beginning of a section of the survey.

Chapter 8: Avoid ANDs and ORs

Any question that has an "and" is likely to be a double-barreled question. That is a question which is asking a person about two or more separate issues or topics but only allowing for one response. Consider the following question – *How satisfied were you with the drink <u>and</u> food offered at the pool party?* A person who enjoyed the food but did not like the drink options will have tough time answering this question. A simple but incorrect fix would be to change the question to read - How *satisfied were you with the drink <u>or</u> food offered at the pool party?* Using an "or" rather than "and" conjunction usually helps the respondent, but the response to a question that uses "or" as a conjunction is usually hard to analyze. In this example, a person will now probably answer the question based on their disappointment with either the food or drink, but no information is learned about which item (food or drink) was the problem.

Fixing questions that include an "and" or "or" conjunction can easily be done by breaking the question into more than one question. The right approach for our example would be to ask - *How satisfied were you with the drink offered at the pool party?* - followed by the question - *How satisfied were you with the food offered at the pool party? A* reason that researchers end up asking questions that include ANDs or ORs is that they are often under pressure to cut questions asked on the survey. However, you should always think about whether it shortens the time it takes to complete the survey. One could easily argue that it is easier and therefore possibly faster to answer two separate questions about food and drink compared to formulating a response to single question that combines food and drink.

Chapter 9: Questions can be too burdensome, too easy, or too factual

Questions should not be too burdensome or too easy

How much effort you can expect depends on who are your target respondents. Often researchers want to know what is better, a 3-point, 5-point, 7-point or 10-point scale. There is no simple answer. For less engaged or less educated respondents, larger scales may be too burdensome. While shorter scales for more engaged or higher educated respondents may be too easy and not fully capture the variability of the responses. So, scales need to consider how engaged or educated your respondents will be.

Another cognitively challenging task that researchers ask of their respondents is to rank options. As a rule, I discourage asking respondents to rank a list of response options. The effort put forth to rank options will vary among respondents reducing the usefulness of the responses. Questions that ask people to come up with percentage estimates that in the end should sum to 100% can also be a burdensome task. For responses that should sum to 100%, consider not forcing the responses to sum to 100% and simply normalize their responses to sum to 100% when cleaning the data file.

Other burdensome questions include questions that are too long or questions that respondents will have trouble accurately responding to. For example, consider the following question:

"*In the past month, how many times were you late for work?"*

For a person that is seldom late this is an easy question to answer, but for someone who is often late, this can be a burdensome question. In this situation, it would be easier for the respondent if you were to create categories such as more than once a week, about once a week, once every two weeks, once a month, never.

Sometimes questions require respondents to go and find their answer and that can be problematic. For example, consider the following question:

"How much did you pay for auto insurance last year?"

Few people might know this without having to do some checking, but most people will either guess or leave the question unanswered. For surveys where the respondent will need to look up information, when possible, the respondent should be given advance notice about the information that they will need to complete the survey.

The Bottom-line when deciding whether a question or set of questions is too burdensome, keep in mind your target respondents. While you usually want to

avoid asking difficult or challenging questions, when questions become too easy it could lead to people taking short cuts. Again, conducting cognitive interviews to test your questionnaire can be an effective way of determining whether a question or series of questions are too burdensome.

Challenges of asking factual question

Often researchers would like to ask people factual or knowledge questions to measure how well people understand things. Questions about whether a person knows the name of a specific program or whether they can name their State Senators, or do they know how many Supreme Court Justices there are. These are interesting research questions but asking them can be problematic.

For some people factual or knowledge questions may seem intimidating and reactions when asked factual or knowledge questions could invoke anger. Respondents may feel that they are being tested and their lack of knowledge potentially could lead to stress or sadness. Another problem is that people may feel the need to cheat by asking someone to help them answer or looking up the answer. With online surveys, looking up answers is easy and often tempting.

Using a lead-in statement explaining the purpose of asking factual questions can potentially help lessen the respondents' concerns about providing a response that maybe wrong. With an on-line or self-administered survey, you might consider emphasizing that they should not look up the answers. However, telling people not to look up the answers could backfire and increase the chance of the respondent looking up the answer. Try not to begin your survey with these types of questions since they may lead to people quitting and try not to end with these items since these items might leave people with a negative perception of your survey.

Chapter 10: Appropriate use of a neutral option

One question that often arises about using scales to measure respondents' perceptions is whether to include a middle value. Some experts advocate forcing respondents to choose a positive or negative response rather than giving them a midpoint (neutral response), such as 3 on a 1 to 5 scale. The view against using a middle or neutral value is based on the fear that too many respondents will prefer it. My own experience is that offering midpoints does not lead respondents to rely too much on them. I advocate using an odd-numbered scale, such as 1 to 5, 1 to 7 or 0 to 10.

Although I like including a neutral option for questions that use scales, I usually argue against offering respondent a don't know or not sure response option for the following reasons:

- Foreign-Born and Less-educated respondents are more likely to choose this category.
- Offering this option increases the likelihood of not responding to other items.
- Studies have shown that including these items does not improve the consistency of respondents' attitudes over time.
- For questions that require thought, these items discourage respondents from thinking about the issue. (makes people lazy)
- Respondents who do not have clearly formulated opinions usually lean in one direction or the other.
- Recent cognitive studies have shown that respondents who choose these options could, if encouraged, provide substantive reason.
- Removing don't know or not sure categories usually does not impact results.

If in doubt, consider doing cognitive testing asking people why they choose a neutral or not sure response. Cognitive testing can often determine whether a don't know or not sure option is needed, especially on recall questions or questions that ask someone to report on someone else's behaviors.

Chapter 11: Question order effects

Evaluators need to anticipate how receptive target respondents will be to the questions asked of them. The order of the questions should reflect how respondents are likely to perceive the intrusiveness or sensitivity of the questions. Surveys often start with non-controversial questions that are extremely easy to answer and inoffensive. The questions then can move toward more sensitive areas, such as requesting respondents to evaluate services, and end with requests for demographic information that will help to disaggregate responses.

In general, ask the broader questions after the more specific questions. For example, consider asking respondents to rate the quality of their recycling services and the garbage collection. Respondents' rating of garbage collection is likely to be influenced by the recycling services being provided. Asking first; "*How would you rate recycling services in your community*" before asking "*How would you rate garbage collection in your community*" clues the respondent that you want them to consider garbage collection separate from the recycling services.

Consider a lead-in statement for items in which there is likely to be an order effect in either direction. For example, consider a respondent trying to evaluate how well the police, the courts, and local leaders have been doing in preventing crime in the community. No matter what order you choose, there is likely to be some order effect, so consider using a lead in statement like the following: "*Now, I'd like you to tell me how effective the police, the courts, and local leaders have been in preventing crime in your community: First, How about ….*" By telling what the options are in advance, the respondent can take them all into consideration before providing their responses to each individual option.

Both avoiding long sections and proper formatting can help reduce order effects. Breaking long sections into smaller sections helps by encouraging respondents to refocus their thoughts. For self-administered surveys, visual formatting can help respondents know when a question is meant to be dependent on previously asked questions.

Also effective is building in real time edit checks that alert the respondent or interviewer when their response to a question is unusual given an earlier response. For instance, if you were to first ask a person "about how many miles a person drove their car to get to work?" and then later asked "How many minutes did it take to get to work?"- You could have the programmed survey instrument check to see if the responses are reasonable and if not verity the responses. You certainly would not believe that someone could get to work in 15 minutes when they had to drive thirty miles, unless you believe they were driving 120 miles per hour.

Sometimes order effects are caused by questions that require the respondent to consider similar but different scenarios. For example, asking questions about other people living in a respondent's household can be different from asking questions about a respondent's family. If you asked a person that lives in a multi-family household "*How many people live in your household?*" and "*How many people in your family got a flu shot last year?*" Their responses to these two questions could easily differ depending on the order the questions get asked. In this situation and others like it, consider specifically telling the respondent who they should include or exclude. For example, "*Including yourself and non-family members, in total how many people live in your household?*" and "*Including you, your spouse or partner and your children, how many people in your immediate family got a flu shot last year?*"

When to randomize question order

Randomizing the order questions are asked can help mitigate bias to any one question. Randomizing the order that a series of questions is asked is always a good ideal when there is potential bias based on the order that questions are asked. However, this is not a perfect solution since randomizing is not likely to reduce order effect bias but spreads the bias across questions. A better solution would be to try to rewrite the questions or move the questions that most likely influence each other into different sections of the survey. Long lists of questions that use rating scales are known to have primacy effects (Schuman & Presser, 1996). That is asking a question earlier on a list of questions increases the chances of the questions yielding a favorable response. Again, potential order effect bias is another good reason to break large sections of questions into smaller sections.

Testing for question order effects

A pre-test or pilot test will usually have too small a sample size to test whether responses differ depending on the order questions get asked. If the questions were randomized and asked on an earlier survey, then testing for order effects would be useful. Depending on your findings, you could try to improve or move around those questions that are most impacted by the order they get asked. Also, use cognitive testing which provides insight into how respondents formulate their responses and thus is a valuable tool for identifying potential question order bias and possible solutions.

Chapter 12: Mode considerations

The work involved in preparing the survey instrument depends on the mode of data collection. Therefore, a decision on which data collection mode to use should be decided early on, so that there is enough time to convert your questionnaire into a survey instrument. The four most common modes are traditional mail surveys, telephone surveys, face-to-face interviews, and online Web surveys.

Mail surveys are now less common because of the increasing movement towards on-line Web surveys, but traditional mail surveys are still a popular form of data collection. The three distinct advantages of mail surveys are that they are usually less expensive, a complete list of addresses is usually obtainable, and they yield less response bias when the questions are sensitive. Some disadvantages of mail surveys are comparatively lower response rates, a response bias toward more educated respondents, higher nonresponse rates for individual questions, and the questionnaire needs to be short, with little or no skip patterns (that is, instructions that ask respondents to ignore some questions).

Telephone surveys have been extremely popular because they often yield high response rates and less item nonresponse, provide more control of the question ordering, allow the use of longer questions, and skip patterns, and allow callers to ask respondents to recall information during the interview. Disadvantages of telephone surveys are that they are more expensive, it may be more time-consuming to write and test questions for them, and there is more bias when asking for sensitive or personal information. Higher cooperation rates and the ability to reach people by telephone have been two major advantages of telephone surveys, but these advantages are now on the decline. When the phone rang forty years ago, no one assumed the person on the other end of the line was going to try to sell him or her something. Thirty years ago, the only way to know who was calling was to say, “Hello.” Today, people screen their calls with caller ID, an answering machine, or a privacy manager.

Face-to-Face Surveys is the oldest method and still yields the highest response rates and is the best method for asking open-ended questions (questions that do not limit responses to predefined response options) or questions requiring visual aids. However, these surveys are usually expensive, require longer testing and data collection periods, and are inappropriate for surveys that include sensitive questions. Also, sampling usually involves interviewers conducting interviews in a small geographical area, which can create a clustering effect that will decrease the precision of estimates. These surveys are appropriate for captive audiences, such as institutionalized clients.

Online web surveys are still a relatively new method of survey data collection. However, there has been a recent proliferation in their use, which has paralleled the dramatic worldwide growth of Internet access. The lower cost of Web surveys compared to telephone or face-to-face interviews is another important reason

behind the increased use of Web surveys. Web surveys can take advantage of the established HTML and Java script standards that survey designers can use to create complex questionnaires that can manage skip patterns and recall information from earlier questions. Also, Web surveys can provide enticing graphics or visual aids to help guide respondents. However, when designing a web base questionnaire, you usually need to keep things simple because respondents will choose to fill out a web survey using a small handheld device like a smartphone of small tablet. Also, there is concern with providing information over the Internet, despite the data being more secure than information provided by e-mail. Though, respondents are starting to trust giving information on the Internet, as evidenced by the increasing willingness of people to fill out forms and purchase things with their credit cards on the Internet. Invitations to participate in a web survey are usually done by sending an email invite but increasingly we are seeing invites sent as a short message service (SMS) test message.

Interactive Voice Response (IVR) surveys are pre-recorded automated phone surveys in which the people respond vocally to questions or type numerical responses using their keypad. IVR enables you to create pre-recorded automatic telephone surveys during which the people respond to a brief series of questions.

An inbound IVR survey is a method where an invitation to participate in the survey is sent by email, letter, SMS, or printed marketing materials and the respondent calls in to take the survey (usually calling a toll-free number). With an outbound IVR survey, the respondent is called to participate in the survey by way of an automated system, from telephone information that was collected. In this method, it is important to consider the volume of calls going out to avoid overwhelming the IVR system.

Surveys sometimes use a 0-10 point, or 1-10-point scale which captures a certain degree of nuances from an analytic standpoint. With an IVR survey, these scales can present difficulties when they require the respondent to enter a two-digit rating from telephone keypad. For instance, the system may fail to capture both digits when given a rating of “10” and record it as a “1” instead. To avoid this problem, use a 5-point scale, or even, a 7-point scale. If a 10- or 11-point scale is chosen, the IVR technology should allow for ample response time and/or confirmation of the scores entered.

Since an IVR survey is automated, there is a greater chance of a drop-off than with an interviewer-administered survey. So, try to keep the survey as brief as possible, with key questions at the beginning of the survey, to maximize the number of responses. Another IVR best practice is to limit the use of open-ended questions to avoid additional drop-off on the survey.

Mixed-Mode Surveys are becoming more common. Given the difficulties in persuading people to respond to surveys, survey practitioners have increasingly been offering people more than one way of responding. There are surveys for which

using a combination of data collection modes will increase participation. For instance, when conducting a job satisfaction survey, you may be able to collect most responses using the Web, but for employees who do not use a computer, you may need to call them or provide them with a paper survey. The downside of mixed-mode surveys is that they cost more to design, and you need to be careful that the mode of data collection does not influence the results. In general, survey results do not vary when self-administration modes of data collection, such as mail or Web surveys, are combined, but results for interviewer-administered modes of data collection (phone or face-to-face) often differ from the results collected through self-administered modes of data collection (Dillman et al., 2009).

There are other ways to conduct surveys. You may use pencil-and-paper questionnaires especially with a captive audience—for instance, asking people to fill-out a survey at the end of a meeting, surveying students in the classroom, or having clients complete a survey while they fill out required forms. It is often possible to bolster the response rate by taking advantage of a captive audience. Until recently, e-mail was a viable alternative to mail for conducting surveys with populations for whom you could get e-mail addresses. You can collect the data faster and potentially more efficiently through e-mail. However, concerns about who has access to e-mail messages, filters and firewalls that prevent unsolicited e-mail, and increases in computer viruses sent by e-mail are all contributing to a declining interest in e-mail surveys.

The emergence of hand-held and mobile devices that often have integrated features such as voice, photography, video, text, email, GPS, apps, and other features has given survey practitioners new measurement tools to study public opinion, attitudes, and behaviors. For instance, travel surveys can now include a GPS application that can keep track of a person's time spent at distinct locations. The GPS application may be a suitable alternative to requiring respondents to keep a survey travel diary or least aid the person in accurately filling out their diary. Keep in mind that new measurement tools are likely to introduce new forms of measurement error and privacy concerns. So, plan to do rigorous testing before you start using new devices to collect data.

Another key point is that, if you plan to compare results with a previous survey, try to use the same mode of data collection; using a different mode of data collection could introduce unintentional bias.

Chapter 13: Testing Survey Questions

When people think about testing survey questions, they often think of conducting a pretest. Although pretesting is important it is not the only method for testing survey questions and probably the least effective method. Expert panel reviews and cognitive testing often find problems that would be undetected during a traditional pretest. Depending on timing and the project budget a pilot study and random experiments are other ways to test survey questions. A combination of testing methods is recommended since each of the methods described in this chapter tend to find distinct types of problems.

Expert panel review

Getting a group of experts that are familiar with the topics covered by your survey questions is often the most useful method of testing survey questions. Topical experts will identify survey questions that are incorrectly worded. A subject matter expert also may be able to identify better versions of the questions you want to ask. It is important to let the expert know the objectives of your research, who is the target population, and the mode of data collection. You should encourage your expert reviewers to think about whether the questions you are asking will lead to the information you are trying to obtain and whether other questions need to be asked.

It is also useful to recruit a survey design expert to review your survey questions, although they may not have knowledge about the topic, they will be able to identify more generic questionnaire problems. Also, the survey design expert will be looking for things that may not be clear to all respondents and the lack of topical expertise should be helpful at identifying instances where wording might be too specific.

If you are writing new survey questions, the expert panel should be the first and probably will be the most important method in testing your survey questions.

Cognitive testing

The importance and how to do cognitive testing was already discussed in the last section of Chapter2. Another strength of cognitive testing is that it complements other testing methods. With expert panels, you often get half-baked questions and questions that have never been asked before. Using cognitive testing after getting feedback from your expert panel is a terrific way finding problems with questions that the experts suggested you include in your survey. After the expert panel has done their job, you should consider recruiting some of these experts to serve as cognitive test respondents. Similarly, consider recruiting respondents that participated in a pilot or pretest to serve as cognitive test respondents. The advantage here is that you have their original answers and can now probe more deeply into how they formulated those responses.

Traditional Pretest

No matter how much question testing you do, pretesting a survey instrument with a representative sample of the population of target respondents is essential. The questions, mode of administration, and procedures should be the same in the pretest as planned for the survey. Even if questions are borrowed from previous studies or other agencies or authorities, the questions still need to be asked of a small sample of the target respondents to ensure clarity and understandability. Often more than one pretest is needed in general, the final pretest should look as much like the final survey as possible.

A good-sized sample for a pretest is twenty to twenty-five completed interviews or written surveys, with more needed for questionnaires that have skip patterns. With questionnaires administered by interviewers, try to record the pretest interviews; this will allow you and others to evaluate the respondents' understanding of the questions. Also, if the total population is small, you may need to include the findings from the pretest sample in the analysis, noting questions that changed because of the pretest. However, if your sample size is large enough, I recommend not including the findings from the pretest sample in the analysis results.

Pilot test

A pilot test is often described as a large-scale pretest. However, it is better to think of a pilot test as a feasibility test that answers the question of whether the survey will work. A pilot test is small scale trial of what you expect your main data collection to look like. So, a pilot is not usually the best method for testing survey questions but can identify questionnaire problems that were not caught by the other testing methods.

Random Assignment

A frequent problem survey designers face is how to revise a poor or inadequate question when one of the research objectives is to look at change over time. Random assignment is a method for assigning survey respondents to answer either the original question or a revised version of the question. Random assignment gives each respondent an equal chance of being asked the original or the revised question. Like flipping a coin to see which question you get asked. Although, unless conducting a paper and pencil survey the random assignment procedure is computerized.

Random assignment is not guaranteed to control all extraneous variables. It is always possible that just by chance, the respondents getting the question might turn out to be different from the respondents getting the original question. So, even with

random assignment you may need to use survey weights when comparing the newer question responses to the response to the older question. However, the larger your sample size the less likely you will see difference among your extraneous variables.

The results of the random assignment will tell you whether the new wording of your question is getting the same responses. If not, you will at least know what differences to expect.

Validity Testing

What does it mean when you hear that survey questions have been validity tested and therefore should not be changed. Validity is defined as the degree of agreement between the claimed measurement and the real world. There are four categories of validity testing; 1) face validity 2) content validity; 3) construct validity; 4) and reliability.

- Face Validity – Face validity asks whether the survey or question measure what it intended to measure? Use targeted respondents (not experts) to answer your questions to see if they can answer them as you would expect them to. Check for face validity before any other validity tests.

- Content Validity - Content validity seeks to answer the question of whether the current test covers all relevant items needed to answer the research question. Use an expert panel to answer the question: Is the question essential to the intended measurement? Recruit a panel of subject matter experts and then ask them whether your intended questions or survey is relevant to your intended research issue. Create the following content validity ratio (CVR): CVR = [(ne - N)-N/2] / 2

    - ne= number of experts in the panel answering yes
    - N= total number of experts in the panel

- Construct Validity - Construct validity is the degree to which the test measures what the theory claims. There are two kinds of construct validity: (a) convergent validity and (b) discriminant validity. A convergent construct validity exists when what is expected to be correlated indeed turns out to be correlated, thus H0: r = 0 and HA: r not equal to 0. The result shows that H0 is incorrect and, thus, is rejected. Whereas, in discriminant validity, r = 0; H0 cannot be rejected. Use correlation coefficients as the unit of analysis.

- Reliability - Reliability is the ability of a questionnaire to consistently produce the same results. You can perform test -- re-test for characteristics that should not change. Another reliability test would be comparing your estimates with other reputable studies or with administrative data. Also, you could field your survey in multiple waves comparing estimates in each wave.

There are many reasons to be cautious when using validated questions. If the answer is yes to any of the following questions, then the validity testing that was done may not apply to your survey. For instance, were the survey questions tested:

- On the same target population that you intend to interview?
- Using a different survey mode of data collection?
- Using a different language?
- Using concepts that have changed over time?
- Using standards that may have changed over time?
- Also, it is possible that validity testing has found questions that are reliable and consistent but still stink because the measurement error associated with the questions is reliable and consistent.

Bibliography

Abi-Habib, Natalie, Adam Safir, Timothy Triplett, and Pat Cunningham. *NSAF 2002 Methodology Series Report No. 12: NSAF Questionnaire*. Urban Institute., 2003. http://webarchive.urban.org/publications/900699.html.

Alwin, Duane F, Erin M Baumgartner, and Brett A Beattie. "Number of Response Categories and Reliability in Attitude Measurement†." *Journal of Survey Statistics and Methodology* 6, no. 2 (2018): 212–39. https://doi.org/10.1093/jssam/smx025.

Antoun, Christopher, Mick P. Couper, and Frederick G. Conrad. "Effects of Mobile versus PC Web on Survey Response Quality A Crossover Experiment in a Probability Web Panel." *Public Opinion Quarterly* 81, no. S1 (2017): 280–306. https://doi.org/10.1093/poq/nfw088.

Antoun, Christopher, and Stanley Presser. "Open Questions Self-Administered on the Web versus Interviewer-Administered in Person: The 2016 American National Election Study Mode Comparison." *Public Opinion Quarterly* 88, no. 2 (2024): 249–67. https://doi.org/10.1093/poq/nfae012.

Basson, Danna. "The Effects of Digital Recording of Telephone Interviews on Survey Data Quality." 2005.

Beatty, Paul, Carol Cosenza, and Floyd Fowler. *Experiments on the Design and Evaluation of Complex Survey Questions*. 2019. https://doi.org/10.1002/9781119083771.ch6.

Behr, Dorothée, and Michael Braun. "How Does Back Translation Fare Against Team Translation? An Experimental Case Study in the Language Combination English–German." *Journal of Survey Statistics and Methodology* 11, no. 2 (2023): 285–315. https://doi.org/10.1093/jssam/smac005.

Belli, Robert F., Lynette M. Smith, Patricia M. Andreski, and Sangeeta Agrawal. "Methodological Comparisons Between CATI Event History Calendar and Standardized Conventional Questionnaire Instruments." *Public Opinion Quarterly* 71, no. 4 (2007): 603–22. https://doi.org/10.1093/poq/nfm045.

Bethlehem, Jelke, Fannie Cobben, and Barry Schouten. *Handbook of Nonresponse in Household Surveys (Wiley Handbooks in Survey Methodology)*. 1st ed. Wiley, 2011. http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0470542799.

Bickart, Barbara A, Joan M Phillips, and Johnny Blair. "The Effects of Discussion and Question Wording on Self and Proxy Reports of Behavioral Frequencies." *Marketing Letters* 17, no. 3 (2006): 167–80.

Blair, Johnny, Ronald F Czaja, and Edward A Blair. *Designing Surveys: A Guide to Decisions and Procedures*. Sage Publications, 2013.

Blair, Johnny, Geeta Menon, and Barbara Bickart. "Measurement Effects in Self vs. Proxy Response to Survey Questions: An Information-Processing Perspective." *Measurement Errors in Surveys*, 1991, 145–66.

Bound, John, Charles Brown, and Nancy Mathiowetz. "Chapter 59 - Measurement Error in Survey Data." In *Handbook of Econometrics*, edited by James J. Heckman and Edward Leamer, vol. 5. Elsevier, 2001. https://doi.org/10.1016/S1573-4412(01)05012-7.

Bradburn, Norman M, and Carrie Miles. "Vague Quantifiers." *Public Opinion Quarterly* 43, no. 1 (1979): 92–101.

Callegaro, M., K.L. Manfreda, and V. Vehovar. *Web Survey Methodology*. Research Methods for Social Scientists. SAGE Publications, 2015. https://books.google.com/books?id=A_0aCAAAQBAJ.

Chang, Linchiat, and Jon A. Krosnick. "Comparing Oral Interviewing with Self-Administered Computerized Questionnaires An Experiment." *Public Opinion Quarterly* 74, no. 1 (2010): 154–67. https://doi.org/10.1093/poq/nfp090.

Christian, Leah Melani, and Don A. Dillman. "The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions." *Public Opinion Quarterly* 68, no. 1 (2004): 57–80. https://doi.org/10.1093/poq/nfh004.

Christian, Leah Melani, Don A. Dillman, and Jolene D. Smyth. "Helping Respondents Get It Right the First Time: The Influence of Words, Symbols, and Graphics in Web Surveys." *Public Opinion Quarterly* 71, no. 1 (2007): 113–25. https://doi.org/10.1093/poq/nfl039.

Conrad, Frederick G., and Johnny Blair. "Sources of Error in Cognitive Interviews." *Public Opinion Quarterly* 73, no. 1 (2009): 32–55. https://doi.org/10.1093/poq/nfp013.

Conrad, Frederick G, Mick P Couper, Roger Tourangeau, and Andrey Peytchev. "Use and Non-Use of Clarification Features in Web Surveys." *Journal of Official Statistics* 22, no. 2 (2006): 245.

Converse, Jean M, and Stanley Presser. *Survey Questions: Handcrafting the Standardized Questionnaire*. Sage, 1986.

Couper, Mick P., Frederick G. Conrad, and Roger Tourangeau. "Visual Context Effects in Web Surveys." *Public Opinion Quarterly* 71, no. 4 (2007): 623–34. https://doi.org/10.1093/poq/nfm044.

Couper, Mick P, Eleanor Singer, and Roger Tourangeau. "Understanding the Effects of Audio-CASI on Self-Reports of Sensitive Behavior." *The Public Opinion Quarterly* 67, no. 3 (2003): 385–95.

Couper, Mick P, Roger Tourangeau, and Kristin Kenyon. "Picture This! Exploring Visual Effects in Web Surveys." *Public Opinion Quarterly* 68, no. 2 (2004): 255–66.

Cremeens, Joanne, Christine Eiser, and Mark Blades. "Factors Influencing Agreement between Child Self-Report and Parent Proxy-Reports on the Pediatric Quality of Life Inventory™ 4.0 (PedsQL™) Generic Core Scales." *Health and Quality of Life Outcomes* 4, no. 1 (2006): 1.

Dahlhamer, James M., Adena M. Galinsky, and Sarah S. Joestl. "Asking about Sexual Identity on the National Health Interview Survey: Does Mode Matter?" *Journal of Official Statistics* 35, no. 4 (2019): 807–33. https://doi.org/doi:10.2478/jos-2019-0034.

Dashen, Monica. "The Effects of Retention Intervals on Self-and Proxy Reports of Purchases." *Memory* 8, no. 3 (2000): 129–43.

DeBell, Matthew. "Experimental Effects of Advance Postcards, Survey Title, Questionnaire Length, and Questionnaire Content on Response Rates and Incentive Costs in a Mail Non-Response Follow-Up Survey." *Survey Practice*, ahead of print, March 24, 2022. https://doi.org/10.29115/SP-2022-0002.

Debell, Matthew, Catherine Wilson, Simon Jackman, and Lucila Figueroa. "Optimal Response Formats for Online Surveys: Branch, Grid, or Single Item?" *Journal of Survey Statistics and Methodology* 9, no. 1 (2019): 1–24. https://doi.org/10.1093/jssam/smz039.

Dillman, Don A. *Mail and Internet Surveys : The Tailored Design Method*. J. Wiley, 2000. N/A.

Dillman, Don A. *Mail and Internet Surveys : The Tailored Design Method*. Wiley, 2007.

Dillman, Don A. *Mail and Telephone Surveys*. Vol. 3. Wiley Interscience, 1978.
Dillman, Don A, Glenn Phelps, Robert Tortora, et al. "Response Rate and Measurement Differences in Mixed-Mode Surveys Using Mail, Telephone, Interactive Voice Response (IVR) and the Internet." *Social Science Research* 38, no. 1 (2009): 1–18.
Dillman, Don A, Jolene D Smyth, and Leah Melani Christian. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 2014.
Dykema, Jennifer, Nora Schaeffer, Dana Garbaski, Erik Nordheim, Mark Banghart, and Kristen Cyffka. "The Impact of Parenthetical Phrases on Interviewers' and Respondents' Processing of Survey Questions." *Survey Practice* 9, no. 2 (2016). http://www.surveypractice.org/index.php/SurveyPractice/article/view/348.
Elangovan, N., and E. Sundaravel. "Method of Preparing a Document for Survey Instrument Validation by Experts." *MethodsX* 8 (January 2021): 101326. https://doi.org/10.1016/j.mex.2021.101326.
Fowler, F.J. *Survey Research Methods*. Applied Social Research Methods. SAGE Publications, 2013. https://books.google.com/books?id=CR-MAQAAQBAJ.
Fowler, Floyd J. *Improving Survey Questions: Design and Evaluation*. Vol. 38. Sage, 1995.
Galesic, Mirta, and Michael Bosnjak. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly* 73, no. 2 (2009): 349–60. https://doi.org/10.1093/poq/nfp031.
Galesic, Mirta, Roger Tourangeau, Mick P. Couper, and Frederick G. Conrad. "Eye-Tracking Data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding." *Public Opinion Quarterly* 72, no. 5 (2008): 892–913. https://doi.org/10.1093/poq/nfn059.
Garbarski, Dana, Nora Cate Schaeffer, and Jennifer Dykema. "Interviewing Practices, Conversational Practices, and Rapport Responsiveness and Engagement in the Standardized Survey Interview." *Sociological Methodology*, 2016, 0081175016637890.
Gnambs, Timo, and Kai Kaspar. "Socially Desirable Responding in Web-Based Questionnaires: A Meta-Analytic Review of the Candor Hypothesis." *Assessment* 24, no. 6 (2017): 746–62.
Graham, Matthew H. "'We Don't Know' Means 'They're Not Sure.'" *Public Opinion Quarterly* 85, no. 2 (2021): 571–93. https://doi.org/10.1093/poq/nfab028.
Groves, Robert M. "Actors and Questions in Telephone and Personal Interview Surveys." *Public Opinion Quarterly* 43, no. 2 (1979): 190–205.
Groves, Robert M. *Survey Errors, and Survey Costs*. Vol. 536. John Wiley & Sons, 2004.
Groves, Robert M, Floyd J Fowler Jr, Mick P Couper, James M Lepkowski, Eleanor Singer, and Roger Tourangeau. *Survey Methodology*. Vol. 561. John Wiley & Sons, 2009.
Gwartney, P.A. *The Telephone Interviewer's Handbook: How to Conduct Standardized Conversations*. Research Methods for the Social Sciences. Wiley, 2007. https://books.google.com/books?id=HU32fO1chNUC.
Hatry, Harry P. *Customer Surveys for Agency Managers: What Managers Need to Know*. The Urban Institute, 1998.
Herzog, A Regula, and Jerald G Bachman. "Effects of Questionnaire Length on Response Quality." *Public Opinion Quarterly* 45, no. 4 (1981): 549–59.
Hess, Jennifer, Jeffrey Moore, Joanne Pascale, Jennifer Rothgeb, and Catherine Keeley. "The Effects of Person-Level versus Household-Level Questionnaire Design on Survey Estimates and Data Quality." *The Public Opinion Quarterly* 65, no. 4 (2001): 574–84.

Hilton, Charlotte Emma. "The Importance of Pretesting Questionnaires: A Field Research Example of Cognitive Pretesting the Exercise Referral Quality of Life Scale (ER-QLS)." *International Journal of Social Research Methodology* 20, no. 1 (2017): 21–34.

Holbrook, Allyson, Young IK Cho, and Timothy Johnson. "The Impact of Question and Respondent Characteristics on Comprehension and Mapping Difficulties." *Public Opinion Quarterly* 70, no. 4 (2006): 565–95. https://doi.org/10.1093/poq/nfl027.

Holbrook, Allyson L., Jon A. Krosnick, David Moore, and Roger Tourangeau. "Response Order Effects in Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes." *Public Opinion Quarterly* 71, no. 3 (2007): 325–48. https://doi.org/10.1093/poq/nfm024.

Jann, Ben, Julia Jerke, and Ivar Krumpal. "Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism." *Public Opinion Quarterly* 76, no. 1 (2012): 32–49. https://doi.org/10.1093/poq/nfr036.

Johnson, Timothy, Diane O'Rourke, Noel Chavez, et al. "Social Cognition and Responses to Survey Questions Among Culturally Diverse Populations." In *Survey Measurement and Process Quality*. John Wiley & Sons, Inc., 1997. http://dx.doi.org/10.1002/9781118490013.ch4.

Kaminska, Olena, Allan L. McCutcheon, and Jaak Billiet. "Satisficing Among Reluctant Respondents in a Cross-National Context." *Public Opinion Quarterly* 74, no. 5 (2010): 956–84. https://doi.org/10.1093/poq/nfq062.

Kaplan, Robin L, Tywanquila Walker, and Rebecca L Morrison. "Beyond Functional Limitations: A Multimethodological Approach to Questionnaire Pretesting of New Disability Measures." *Public Opinion Quarterly* 89, no. SI (2025): 624–50. https://doi.org/10.1093/poq/nfaf029.

Katz, Jonathan, Jasmine Luck, and Jenna Fulton. "Lessons Learned from Conducting Paired Cognitive Interview Studies to Examine the Feasibility of Proxy Reporting." *Survey Practice* 15, no. 1 (2022). https://doi.org/10.29115/SP-2022-0001.

Kojetin, BA, and Paul Mullin. "The Quality of Proxy Reports on the Current Population Survey (CPS)." *50th Annual Conference of the American Association for Public Opinion Research*, 1995.

Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity." *Public Opinion Quarterly* 72, no. 5 (2008): 847–65. https://doi.org/10.1093/poq/nfn063.

Kreuter, Frauke, Joseph W. Sakshaug, and Roger Tourangeau. "The Framing of the Record Linkage Consent Question." *International Journal of Public Opinion Research* 28, no. 1 (2016): 142–52. https://doi.org/10.1093/ijpor/edv006.

Krosnick, Jon A. "Questionnaire Design." In *The Palgrave Handbook of Survey Research*. Springer, 2018.

Krosnick, Jon A, and Duane F Alwin. "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement." *Public Opinion Quarterly* 51, no. 2 (1987): 201–19.

Krosnick, Jon A, and Leandre R Fabrigar. "Designing Rating Scales for Effective Measurement in Surveys." *Survey Measurement and Process Quality*, 1997, 141–64.

Krosnick, Jon A, Allyson L Holbrook, Matthew K Berent, et al. "The Impact of" No Opinion" Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice?" *Public Opinion Quarterly* 66, no. 3 (2002): 371–403.

Lax, Jeffrey R., Justin H. Phillips, and Alissa F. Stollwerk. "Are Survey Respondents Lying about Their Support for Same-Sex Marriage? Lessons from a List Experiment." *Public Opinion Quarterly* 80, no. 2 (2016): 510–33. https://doi.org/10.1093/poq/nfv056.
Leeuw, Edith de. "To Mix or Not to Mix Data Collection Modes in Surveys." *J Off Stat* 21 (November 2004).
Li, Shan. "Measuring Cognitive Engagement: An Overview of Measurement Instruments and Techniques." *International Journal of Psychology and Educational Studies* 8 (July 2021): 63–76. https://doi.org/10.52380/ijpes.2021.8.3.239.
Liu, Mingnan, Sunghee Lee, and Frederick G. Conrad. "Comparing Extreme Response Styles between Agree-Disagree and Item-Specific Scales." *Public Opinion Quarterly* 79, no. 4 (2015): 952–75. https://doi.org/10.1093/poq/nfv034.
Lundmark, Sebastian, Mikael Gilljam, and Stefan Dahlberg. "Measuring Generalized Trust: An Examination of Question Wording and the Number of Scale Points." *Public Opinion Quarterly* 80, no. 1 (2016): 26–43. https://doi.org/10.1093/poq/nfv042.
Lynn, Peter, and Olena Kaminska. "The Impact of Mobile Phones on Survey Measurement Error." *Public Opinion Quarterly* 77, no. 2 (2013): 586–605. https://doi.org/10.1093/poq/nfs046.
Malhotra, Neil. "Completion Time and Response Order Effects in Web Surveys." *Public Opinion Quarterly* 72, no. 5 (2008): 914–34. https://doi.org/10.1093/poq/nfn050.
Malhotra, Neil. "Order Effects in Complex and Simple Tasks." *Public Opinion Quarterly* 73, no. 1 (2009): 180–98. https://doi.org/10.1093/poq/nfp008.
Malhotra, Neil, Jon A. Krosnick, and Randall K. Thomas. "Optimal Design of Branching Questions to Measure Bipolar Constructs." *Public Opinion Quarterly* 73, no. 2 (2009): 304–24. https://doi.org/10.1093/poq/nfp023.
McClendon, McKee J, and Duane F Alwin. "No-Opinion Filters and Attitude Measurement Reliability." *Sociological Methods & Research* 21, no. 4 (1993): 438–64.
McConnell, John. *How to Design, Implement, and Interpret an Employee Survey*. Amacom, 2003.
McFarland, Sam G. "Effects of Question Order on Survey Responses." *The Public Opinion Quarterly* 45, no. 2 (1981): 208–15. JSTOR.
Meadows, Keith. "Cognitive Interviewing Methodologies." *Clinical Nursing Research* 30, no. 4 (2021): 375–79. https://doi.org/10.1177/10547738211014099.
Menon, Geeta, Barbara Bickart, Seymour Sudman, and Johnny Blair. "How Well Do You Know Your Partner? Strategies for Formulating Proxy-Reports and Their Effects on Convergence to Self-Reports." *Journal of Marketing Research*, 1995, 75–84.
Mneimneh, Zeina N., Steven G. Heeringa, Roger Tourangeau, and Michael R. Elliott. "Bridging Psychometrics and Survey Methodology: Can Mixed Rasch Models Identify Socially Desirable Reporting Behavior?" *Journal of Survey Statistics and Methodology* 2, no. 3 (2014): 257–82. https://doi.org/10.1093/jssam/smu008.
Morwitz, Vicki G. "It Seems Like Only Yesterday: The Nature and Consequences of Telescoping Errors in Marketing Research." *Journal of Consumer Psychology* 6, no. 1 (1997): 1–29. https://doi.org/10.1207/s15327663jcp0601_01.
Neuert, Cornelia E. "Do Forced-Choice (FC) Questions Trigger Deeper Cognition than Check-All-That-Apply (CATA) Questions?" *Journal of Survey Statistics and Methodology* 8, no. 4 (2019): 617–35. https://doi.org/10.1093/jssam/smz015.

Newcomer, Kathryn E., and Timothy Triplett. "Using Surveys." In *Handbook of Practical Program Evaluation*. John Wiley & Sons, Inc., 2015. http://dx.doi.org/10.1002/9781119171386.ch14.

Olson, Kristen, and Jolene D Smyth. "The Effects of Mismatches between Survey Question Stems and Response Options on Data Quality and Responses." *Journal of Survey Statistics and Methodology* 7, no. 1 (2018): 34–65. https://doi.org/10.1093/jssam/smy005.

Peytchev, Andy, Mick P. Couper, Sean Esteban McCabe, and Scott D. Crawford. "Web Survey Design: Paging versus Scrolling." *Public Opinion Quarterly* 70, no. 4 (2006): 596–607. https://doi.org/10.1093/poq/nfl028.

Presser, Stanley, Mick P. Couper, Judith T. Lessler, et al. "Methods for Testing and Evaluating Survey Questions." *Public Opinion Quarterly* 68, no. 1 (2004): 109–30. https://doi.org/10.1093/poq/nfh008.

Redline, Cleo, and Don A Dillman. "The Influence of Alternative Visual Designs on Respondents' Performance with Branching Instructions in Self-Administered Questionnaires." *Survey Nonresponse*, 2002, 179–93.

Roberts, Caroline, Emily Gilbert, Nick Allum, and Léïla Eisner. "Research Synthesis: Satisficing in Surveys: A Systematic Review of the Literature." *Public Opinion Quarterly* 83, no. 3 (2019): 598–626. https://doi.org/10.1093/poq/nfz035.

Schaeffer, Nora Cate. "Hardly Ever or Constantly? Group Comparisons Using Vague Quantifiers." *Public Opinion Quarterly* 55, no. 3 (1991): 395–423.

Schiff, Kaylyn Jackson, B Pablo Montagnes, and Zachary Peskowitz. "Priming Self-Reported Partisanship: Implications for Survey Design and Analysis." *Public Opinion Quarterly* 86, no. 3 (2022): 643–67. https://doi.org/10.1093/poq/nfac032.

Schmidt, Stefanie R, Kevin H Wang, and Freya L Sonenstein. "Designing a Household Survey to Address Seasonality in Child Care Arrangements." *Evaluation Review* 32, no. 2 (2008): 216–33.

Schuman, Howard, and Stanley Presser. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Sage, 1996.

Sha, Mandy. "The Use of Vignettes in Evaluating Asian Language Questionnaire Items." *Survey Practice* 9, no. 3 (2016). http://www.surveypractice.org/index.php/SurveyPractice/article/view/336.

Shaeffer, Eric M., Jon A. Krosnick, Gary E. Langer, and Daniel M. Merkle. "Comparing the Quality of Data Obtained by Minimally Balanced and Fully Balanced Attitude Questions." *Public Opinion Quarterly* 69, no. 3 (2005): 417–28. https://doi.org/10.1093/poq/nfi028.

Sigelman, Lee, Steven A. Tuch, and Jack K. Martin. "What's in a Name?: Preference for 'Black' versus 'African-American' among Americans of African Descent." *Public Opinion Quarterly* 69, no. 3 (2005): 429–38. https://doi.org/10.1093/poq/nfi026.

Silber, Henning, Endre Tvinnereim, Tobias H Stark, et al. "Lack of Replication or Generalization? Cultural Values Explain a Question Wording Effect." *Journal of Survey Statistics and Methodology* 10, no. 5 (2022): 1121–47. https://doi.org/10.1093/jssam/smab007.

Sirken, Monroe G. *Cognition and Survey Research*. Vol. 322. Wiley-Interscience, 1999.

Smyth, Jolene D., Leah Melani Christian, and Don A. Dillman. "Does 'Yes or No' on the Telephone Mean the Same as 'Check-All-That-Apply' on the Web?" *Public Opinion Quarterly* 72, no. 1 (2008): 103–13. https://doi.org/10.1093/poq/nfn005.

Smyth, Jolene D., Don A. Dillman, Leah Melani Christian, and Mallory McBride. "Open-Ended Questions in Web Surveys: Can Increasing the Size of Answer Boxes and Providing Extra Verbal Instructions Improve Response Quality?" *Public Opinion Quarterly* 73, no. 2 (2009): 325–37. https://doi.org/10.1093/poq/nfp029.

Smyth, Jolene D, Kristen Olson, and Jerry Timbrook. "Why Do Cell Phone Interviews Last Longer? A Behavior Coding Perspective." *Public Opinion Quarterly* 82, no. 3 (2018): 553–82. https://doi.org/10.1093/poq/nfy022.

Strack, Fritz. "'Order Effects' in Survey Research: Activation and Information Functions of Preceding Questions." In *Context Effects in Social and Psychological Research*, edited by Norbert Schwarz and Seymour Sudman. Springer New York, 1992. https://doi.org/10.1007/978-1-4612-2848-6_3.

Sturgis, Patrick, and Ian Brunton-Smith. "Personality and Survey Satisficing." *Public Opinion Quarterly*, September 17, 2023, nfad036. https://doi.org/10.1093/poq/nfad036.

Sudman, Seymour, and Norman M Bradburn. *Asking Questions: A Practical Guide to Questionnaire Design.* 1982.

SÜRÜCÜ, Lütfi, and Ahmet MASLAKÇI. "VALIDITY AND RELIABILITY IN QUANTITATIVE RESEARCH." Articles. *Business & Management Studies: An International Journal* 8, no. 3 (2020): 2694–726. https://doi.org/10.15295/bmij.v8i3.1540.

Sweitzer, Matthew D, and Hillary C Shulman. "The Effects of Metacognition in Survey Research: Experimental, Cross-Sectional, and Content-Analytic Evidence." *Public Opinion Quarterly* 82, no. 4 (2018): 745–68. https://doi.org/10.1093/poq/nfy034.

Tedesco, Heather, Andrew L Zukerberg, and Elizabeth Nichols. "Designing Surveys for the next Millennium: Web-Based Questionnaire Design Issues." *ASC International Conference*, 1999, 103–12.

Thomas, Randall K, and Jonathan D Klein. "Merely Incidental?: Effects of Response Format on Self-Reported Behavior." *Journal of Official Statistics* 22, no. 2 (2006): 221.

Todorov, Alexander. "Cognitive Procedures for Correcting Proxy-Response Biases in Surveys." *Applied Cognitive Psychology* 17, no. 2 (2003): 215–24.

Toepoel, Vera, and Mick P. Couper. "Can Verbal Instructions Counteract Visual Context Effects in Web Surveys?" *Public Opinion Quarterly* 75, no. 1 (2011): 1–18. https://doi.org/10.1093/poq/nfq044.

Tourangeau, Roger, Mick P. Couper, and Frederick Conrad. "Color, Labels, and Interpretive Heuristics for Response Scales." *Public Opinion Quarterly* 71, no. 1 (2007): 91–112. https://doi.org/10.1093/poq/nfl046.

Tourangeau, Roger, Mick P. Couper, and Frederick Conrad. "Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions." *Public Opinion Quarterly* 68, no. 3 (2004): 368–93. https://doi.org/10.1093/poq/nfh035.

Tourangeau, Roger, Robert M. Groves, and Cleo D. Redline. "Sensitive Topics and Reluctant Respondents: Demonstrating a Link between Nonresponse Bias and Measurement Error." *Public Opinion Quarterly* 74, no. 3 (2010): 413–32. https://doi.org/10.1093/poq/nfq004.

Tourangeau, Roger, Aaron Maitland, and H. Yanna Yan. "Assessing the Scientific Knowledge of the General Public: The Effects of Question Format and Encouraging or Discouraging Don't Know Responses." *Public Opinion Quarterly* 80, no. 3 (2016): 741–60. https://doi.org/10.1093/poq/nfw009.

Tourangeau, Roger, Lance J Rips, and Kenneth Rasinski. *The Psychology of Survey Response*. Cambridge University Press, 2000.
Tourangeau, Roger, and Ting Yan. "Sensitive Questions in Surveys." *Psychological Bulletin* 133, no. 5 (2007): 859.
Tourangeau, Roger, and Cong Ye. "The Framing of the Survey Request and Panel Attrition." *Public Opinion Quarterly* 73, no. 2 (2009): 338–48. https://doi.org/10.1093/poq/nfp021.
Triplett, Tim. "Can Your Spouse Accurately Report Your Activities? An Examination of Proxy Reporting." *Survey Practice* 3, no. 1 (2010). http://www.surveypractice.org/index.php/SurveyPractice/article/view/116.
Triplett, Tim, Brian Tefft, and Rob Santos. "Who Has Trouble Reporting Prior Day Events?" *Survey Practice* 10, no. 1 (2017). http://www.surveypractice.org/index.php/SurveyPractice/article/view/376.
Truebner, Miriam. "The Dynamics OF 'Neither Agree nor Disagree' Answers in Attitudinal Questions." *Journal of Survey Statistics and Methodology*, no. smz029 (October 2019). https://doi.org/10.1093/jssam/smz029.
Weisberg, Herbert F. "72Measurement Error Due to Respondents, I: Question Wording Problems." In *The Total Survey Error Approach: A Guide to the New Science of Survey Research*, edited by Herbert F. Weisberg. University of Chicago Press, 2005. https://doi.org/10.7208/chicago/9780226891293.003.0005.
Wenz, Alexander, Tarek Al Baghal, and Alessandra Gaia. "Language Proficiency Among Respondents: Implications for Data Quality in a Longitudinal Face-To-Face Survey." *Journal of Survey Statistics and Methodology* 9, no. 1 (2021): 73–93. https://doi.org/10.1093/jssam/smz045.
Westland, Heleen, Sigrid Vervoort, Marijke Kars, and Tiny Jaarsma. "Interviewing People on Sensitive Topics: Challenges and Strategies." *European Journal of Cardiovascular Nursing* 24, no. 3 (2025): 488–93. https://doi.org/10.1093/eurjcn/zvae128.
Willis, Gordon B. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Sage Publications, 2004.
Wright, Daniel B, George D Gaskell, and Colm A O'Muircheartaigh. "How Much Is 'Quite a Bit'? Mapping between Numerical Values and Vague Quantifiers." *Applied Cognitive Psychology* 8, no. 5 (1994): 479–96.
Yan, Ting, Hanyu Sun, and Anil Battalahalli. "Using Machine Learning to Evaluate Questions in a Multilingual Survey." *Survey Practice* 19 Special Issue (March 2025). https://doi.org/10.29115/SP-2024-0021.
Yeager, David Scott, and Jon A. Krosnick. "Does Mentioning 'Some People' and 'Other People' in an Opinion Question Improve Measurement Quality?" *Public Opinion Quarterly* 76, no. 1 (2012): 131–41. https://doi.org/10.1093/poq/nfr066.