

qualtrics EXPERIENCE MGMT™

# The Qualtrics handbook of question design

**DAVID L. VANNETTE, PH.D.** Principal Research Scientist at Qualtrics

3 **INTRODUCTION**

4 **Q&A**

6 **DATA QUALITY**
7 Goals for Evaluating Survey Quesitons

8 **UNDERSTANDING THE DATA GENERATION PROCESS**

10 **THE THREAT OF SATISFICING**
10 Causes of Satisficing
13 Forms of Satisficing
15 Combating Satisficing

19 **DESIGNING QUESTIONS TO MAXIMIZE DATA QUALITY**
19 Choosing the Best Response Format
23 Rating Scale Point Decisions

27 **QUESTION WORDING**

31 **USING AN EXISTING QUESTION VS. WRITING YOUR OWN**

32 **CONCLUSION**

# Introduction

Survey data are only as good as the questions that generate them. This short handbook introduces some key best practices for question design along with the theoretical and empirical rationale behind those best practices. By following these best practices and principles, you will be able to design questions that will enable your respondents to give you their best data. The recommendations made in this guide are based on a large academic literature about question design in survey methodology along with evidence for best practices that we have discovered at Qualtrics. A few key references are provided throughout, along with a list for further reading, with each tip based on theory and/or empirical evidence that has been discovered through research on survey methodology.

# Q & A

Asking and answering questions seems intuitive—after all, you probably ask and answer dozens of questions each day, maybe hundreds if you have a small child at home. Given our vast experience with asking and answering questions in an interpersonal context, it is not terribly surprising that many researchers seem to assume that question design in the survey context is intuitive, too. However, if you were to review any randomly selected survey, you would find that intuition-based question design does not typically result in a questionnaire that enables respondents to provide their best data.

On the other end of the spectrum, many researchers adopt questions designed by other people without critically evaluating their quality or fitness for the research purpose at hand. There is a common misperception that just because a question has been used before, perhaps by many other people or for a long time, that it must be good. In many cases, these questions were designed for a particular application and may be misapplied outside of the specific context that they were developed for. These questions are also often not vetted by trained survey methodologists, but developed by subject matter experts—people who generally understand very well what they want to measure but not necessarily the best way to measure it.

In both cases, an important issue is how the researcher implicitly treats both the question and the answer that produce a survey data point. An operational definition of a "question" might be "a request for specific information that the respondent is expected to be able to provide." On the surface, this seems like a perfectly plausible definition for a question, and it's one that many researchers implicitly adopt. However, this definition could perhaps more realistically be rephrased as, "a request for specific information that the respondent thinks the researcher wants them to provide." This small change in phrasing highlights a key aspect of the process between asking a question and answering a question—the thought processing that occurs when a respondent interprets a question. In some cases, the interpretation may differ from the literal meaning of the words; in other cases the respondent may feel that there is a socially desirable response or may feel compelled to give an inaccurate response for some other reason. Additionally, the context of the question may influence how the respondent interprets its meaning. Whether you like it or not, responses to question 17 in your survey can easily be affected by the preceding 16 questions because they can influence how your respondent interprets what they believe you are asking about.

Misinterpreting a question can lead to your respondents answering a different question than the one you intended to ask. It's important to ask yourself: Would you know if this happened for one of your respondents? Or all of your respondents? In most cases the answer to both questions is no, which should give any good researcher pause. Understanding how respondents think about your survey questions and the implications for your design decisions is critically important for collecting valid data.

Similarly, researchers have an implicit understanding of what an answer is when analyzing survey data. In most cases this understanding can be summarized as "a response to a request for information." But an answer to a survey question can also be understood to be "a response to a request for information that is interpreted (or misinterpreted), considered, edited, and mapped onto a response option." This alternative, and perhaps more realistic, definition of an answer once again highlights the role that respondent cognition plays in the provision of data and demonstrates how standard analysis processes are not capable of accounting for the complex thought process that generates the data you collect.

Given the underappreciated complexity of asking and answering questions, careful and intentional design of survey questions is of paramount importance to minimizing the opportunity for error.

# Data quality

The ability to draw correct conclusions or insights from survey data depends on the quality of the data. Throughout this handbook, you will find many references to "data quality" and, given the many potential definitions of data quality, it is important to set a framework for how we intend it to be understood in this context. In the question design context, we will focus on two dimensions of data quality: reliability and validity. Understanding these two constructs is critical to the work of question design.

Reliability refers to the extent to which a measurement process provides internally consistent or repeatable results. Internal consistency in this context typically means that items that theoretically should be correlated actually are correlated when examined. For example, it is well-established that human height and weight are very strongly correlated. Consequently, a researcher may expect that survey measures of height and weight should be correlated. If the researcher does not find this expected association upon examining a dataset, it should be a cause for concern about the quality of the data. A similar example would be the correlation between political ideology and political party among Americans; if these things are not correlated then there may be a data quality issue. Researchers typically test this dimension of reliability by computing the correlations between questions they expect to be associated.

Repeatability, or test-retest reliability, is another dimension of reliability that may be more familiar to most researchers. Using the example from before, if a researcher gathers self-report data on respondent height and weight on Monday, and then again from the same respondents on Tuesday, it would be very concerning in terms of data quality if the responses were substantially different. To assess this dimension of reliability, researchers will commonly ask the same question or questions multiple times in a survey with the expectation that different responses to the same question may be an indicator of low-quality data.

Validity is the dimension of data quality that researchers are often most concerned about, and it generally refers to the extent to which a measurement process is actually measuring the thing that it is intended to measure. There are a handful of ways that validity can be operationalized:

- **Construct validity** — how closely does the measure "behave" like it should based on established measures or the theory of the underlying construct?
- **Content validity** — how well does the sample of questions asked reflect the domain of possible questions?
- **Predictive validity** — what is the strength of the empirical relationship between a question and the gold standard?
- **Face validity** — what does the question look like it's measuring?

Unfortunately, the last definition is the one that we see used most commonly by researchers to evaluate validity. The apparent or face validity of a survey question is a poor criterion, but this doesn't prevent researchers from using it to evaluate the likely quality of the data collected with a question. The other approaches to assessing validity listed above are much more robust, despite being more difficult to implement. Ideally, some combination of construct, content, and predictive validity would be applied when assessing the validity of a survey question.

Taken together, reliability and validity are the basis for what we broadly refer to as "data quality" in the context of survey question design. Ensuring that survey questions produce high-quality data is incredibly important for drawing correct conclusions. All recommendations presented in this handbook aim to use knowledge derived from empirical research to help you understand best practices for designing questions that will produce high-quality data.

## GOALS FOR EVALUATING SURVEY QUESTIONS

There are a few key goals that need to inform all question design decisions. The first is that all elements of a question should reduce the opportunity for respondents to make mistakes. Questions should be clear so that the risk of misinterpretation is low. It needs to be easy for respondents to provide valid, reliable, and accurate answers to every question. The second goal is to minimize the difficulty of administering the survey. Researchers should always aim to use questions that can be asked and thoughtfully answered by respondents as easily as possible. While we don't want our respondents to "speed" through the survey, we want to make it as easy as possible for them to provide high-quality answers. Lastly, all else equal, we would like respondents to find the process of answering questions enjoyable and not frustrating. This means designing questions that minimize opportunities for confusion, are as simple as possible to process and answer, and that are relevant to the respondent.

# Understanding the data generation process

Whenever researchers analyze data, assumptions are made about the mental processes that produced those data. With surveys, an implicit assumption made by many researchers is that each question was answered using what survey methodologists often call the "optimal cognitive response process."

There are four components of this process that respondents must mentally engage with (adapted from Tourangeau, Rips, and Rasinski 2000):

**1. COMPREHENSION**

Respondents must understand the question and any instructions. They must determine what is meant by the question as it may differ from the literal interpretation of the words; identify the information being sought by the question; and link key terms from the question text to relevant concepts.

**2. RETRIEVAL**

Respondents must generate an information retrieval cue and then search their memory for the relevant generic and specific information.

**3. JUDGMENT**

Respondents must assess the completeness and relevance of the information retrieved from memory, draw conclusions and inferences based on the accessibility of the information, and integrate the information retrieved. This involves synthesizing the information from memory and making determinations about the concepts in the question.

**4. RESPONSE**

Respondents then take the summarized information and map their judgment onto a response category, editing the judgment for the format of the requested response as necessary.

When respondents perform the four steps above to answer a question they are said to have "optimized" the response process. Most research that analyzes and interprets survey data makes an implicit assumption that this is the process each respondent used to generate their answer (data) for each question. How realistic this assumption is varies dramatically by project, so in many cases it may not be a valid assumption to make.

When presented with this model of the response process, many researchers react with disbelief—many of them have taken surveys before and will admit that, even as research professionals, they do not carefully engage in each of these steps for every question on a survey. This acknowledgement is important because the quality (reliability and validity) of the data that are collected by surveys typically depend on the degree to which respondents complete these steps for each question.

As researchers, we cannot directly control the care that respondents direct toward this process. However, question and survey design decisions can influence respondents heavily—for better or worse. With this in mind, survey designers should keep two goals in mind: make completing the response process as easy as possible for all respondents and avoid making it easy for respondents to shortcut this process when they are answering questions.

In the next section, we highlight an influential theory about how respondents may shortcut the response process, along with number of survey design mistakes you may be making that allow respondents to shortcut the optimal response process and provide low-quality data.

# The threat of survey satisficing

The word “satisficing” is a combination of  “satisfactory” and “sufficient.”  In survey methodology literature, satisficing refers to a theory that describes the common practice of a respondent taking cognitive shortcuts while answering survey questions. Developed by Stanford Professor Jon Krosnick (Krosnick 1991; 1999) this theory suggests that survey respondents engage in the previously mentioned cognitive response process to varying degrees when responding to questions.

When respondents engage in satisficing behavior, they provide responses that often have lower reliability and validity (i.e., lower data quality.) As stated in the section above, when a respondent completes the entire optimal response process, they are said to be “optimizing” their response. However, in many cases, a respondent may only partially search their memory or incompletely integrate the information retrieved from memory—this is known as “weak satisficing.” By contrast, “strong satisficing” refers to the case where a respondent skips the memory search or the information integration steps altogether and simply uses cues from the question, response options, or context to identify an acceptable response.

## Causes of satisficing

Previous studies have identified a number of factors that seem to cause respondents to engage in satisficing. By understanding these causes, researchers can make informed decisions about how to design questions and questionnaires in ways that will minimize the opportunities for respondents to satisfice. While the researcher does not have control over every potential cause, very often the data quality can be improved by applying good question design principles. In the next section, we highlight the known causes of satisficing along with factors that tend to influence these causes.

## TASK DIFFICULTY

Respondents may find the task of completing a survey difficult for a number of reasons. Factors such as the number of words in a question, the familiarity of words, and words with multiple definitions may all make the interpretation phase more difficult.

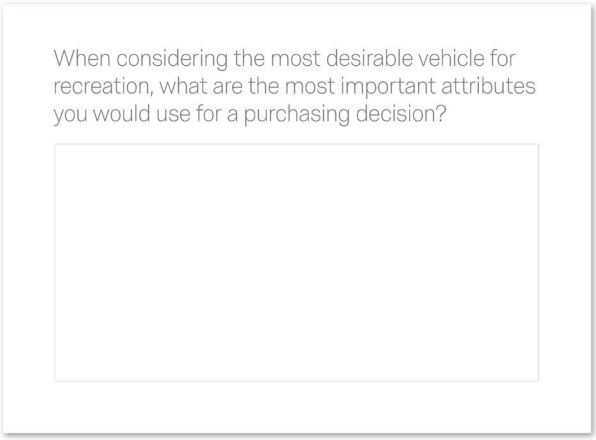



**FIGURE 1A**
Bad example

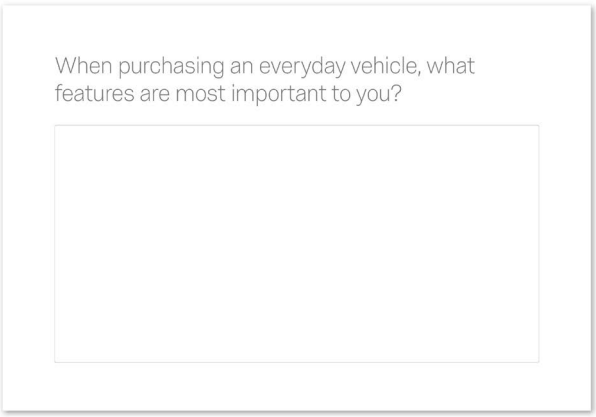



**FIGURE 1B**
Good example

To limit confusion, aim to use the words that are needed to communicate the necessary information to the respondent and no more. Choose words that are commonly used and attempt to avoid words with multiple definitions. During the retrieval phase, answering questions about past events is more difficult for respondents than answering questions about the present. Similarly, answering about multiple subjects is more difficult than answering about a single subject. In general, avoid asking about the past any more than necessary and focus the question on a single subject, using multiple questions to address additional elements as necessary.

In terms of judgment, respondents find it more difficult to make comparative judgments than absolute judgments, and judgments that are decomposable are easier to make than those that are not (e.g., if a person dislikes all carbonated beverages then it will be easier to rate a new flavor of soda than if they like some carbonated beverages and not others.) In the response phase, respondents find it more difficult to understand numeric scale labels compared to verbal labels. In addition, using unfamiliar words in response categories makes the process more difficult. As a general rule, always provide verbal labels for all points on a response scale.

## RESPONDENT ABILITY

Mental ability of the respondent is another factor found to cause satisficing, with low-ability respondents being most likely to engage in satisficing behavior. The cognitive skill of respondents is one of the dimensions of respondent ability that is implicated, and researchers often use education as a rough proxy for cognitive skill because it is difficult to include a full assessment of cognitive ability in most surveys. Respondents with limited experience thinking about the question topic prior to taking the survey are more likely to engage in satisficing. Relatedly, respondents that have pre-existing judgments about the survey topic tend to satisfice less. While researchers generally have no control over the cognitive abilities of the respondents in their samples, it is important to recognize this as a cause of satisficing, particularly for studies that may survey low-ability populations.

## RESPONDENT MOTIVATION

Respondent motivation is a limited and precious resource for researchers. When motivation is low, satisficing behavior tends to increase. Respondent motivation is affected by many factors, the first being the need for cognition, which is a psychological term that describes the extent to which a respondent is inclined to expend significant mental energy on a particular task. Respondents that have low need for cognition are more likely to satisfice than those with high need for cognition. Another motivating factor for respondents is a feeling of accountability; respondents that do not feel accountable for their responses are more likely to satisfice. Relatedly, respondents that do not personally believe that the topic of the survey is important or that the survey itself is important are more likely to satisfice. Lastly, fatigue saps respondent motivation, and a good predictor of fatigue is the number of prior questions in the survey. The more questions that have been asked already, the more likely a respondent is to satisfice.

# Forms of satisficing

Now that we've highlighted some of the causes of satisficing, we can turn to the question of how this behavior harms data quality. Satisficing behavior takes on a variety of forms when respondents are completing a survey. In most cases, there are design decisions that researchers can use to make satisficing easier or more difficult for the respondent to engage in.

## ACQUIESCENCE

The first form of satisficing that we will highlight is called "acquiescence response bias," or a respondent's tendency to agree with suggestions. This is most commonly seen in questions that use agree-disagree response scales; in these question types respondents have a bias toward agreeing, regardless of the content of the statement they are evaluating.

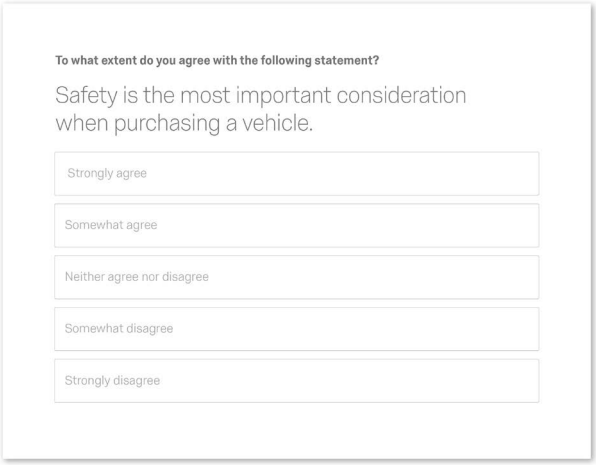



**FIGURE 2**

This also commonly happens with true/false questions, where respondents are more likely to report "true" than "false," and yes/no questions, where respondents demonstrate a bias toward "yes." Using any form of these response scales makes it easier for respondents to engage in satisficing behavior rather than going through the optimal response process. In general, avoid using generic response scales and instead use response scales that are specific to the subject that your question is asking about. For example, if you were asking about the degree of satisfaction or dissatisfaction that your respondent felt about an experience, you could formulate it as an agree-disagree statement: "I was satisfied with my experience," and provide response options ranging from "strongly agree" to "strongly disagree." Or, you could use the best-practice approach of using a construct-specific response scale: "How satisfied or dissatisfied were you with your experience?" with response options ranging from "extremely satisfied" to "extremely dissatisfied."

## STRAIGHTLINING

Straightlined, or non-differentiated responses to rating questions, are another form that satisficing can take. Chances are, if you've ever used a matrix or grid question type in a survey, you've found that at least some of your respondents have provided the same answer for each question in the grid. This is straightlining.

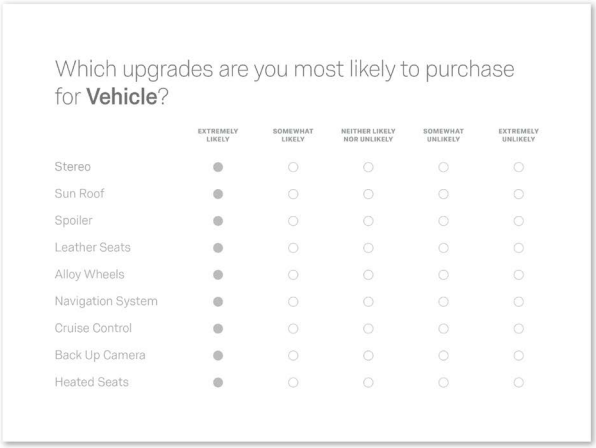



**FIGURE 3**

In the worst-case scenario, respondents that straightline are not even reading the individual questions or statements but are simply clicking answer choices in a straight line to get through the survey as quickly as possible. It is important to note that not all straightlined responses are necessarily invalid, and as a precautionary measure, researchers will sometimes include a reverse-coded version of the same question in order to catch respondents that report, for example, both liking and disliking the same thing. But it can be extremely difficult to separate valid straightlined responses from those produced by satisficing behavior.

Best practices for avoiding straightlining in surveys include:

- Avoid using grid or matrix question types
- Use construct-specific response scales
- Ask one question per page

### PRIMACY

Ideally, we want respondents to carefully read the question and response options before providing an answer. However, when a respondent is satisficing, they are more likely to simply identify the first reasonable response option provided and select it without reading any further. In web surveys this is typically the first reasonable response at the top of a vertically oriented scale or on the left of horizontally oriented scales. This tendency results in an effect known as "primacy" and can introduce a bias into your data.

For questions that offer categorical responses, the best practice is to randomize the order of all of the response alternatives, which usually reduces potential bias at the expense of increased variance. For rating scales, randomizing which end of the response scale is on the top or left—depending on whether the orientation of the scale is vertical or horizontal—may also help reduce bias (Malhotra 2009).

## Combating satisficing

There are two primary tools that researchers can use to combat survey satisficing: task difficulty and respondent motivation. By designing questions to reduce the difficulty of the cognitive response process and maximize respondent motivation, researchers can reduce the chances that respondents will engage in these negative response behaviors. Fortunately, taking steps to reduce satisficing is also likely to increase the validity and reliability of responses.

### TASK DIFFICUTY

To reduce the difficulty of responding to questions, we recommend that researchers take three steps:

1. Make questions easy to understand
2. Minimize distractions
3. Keep the survey short

It is important to note that when we refer to making the questions easy to understand, the goal is to help respondents optimize the cognitive response process and provide accurate, valid, and reliable responses—NOT simply fast responses. For example, respondents are able to click through matrix or grid questions very quickly, perhaps indicating that this is an easy question type for respondents. But research indicates that this question type actually may require greater cognitive effort to optimize responses to than if the same questions are asked individually.[1]

To minimize distractions, researchers should consider asking a single question per page and asking related questions together.

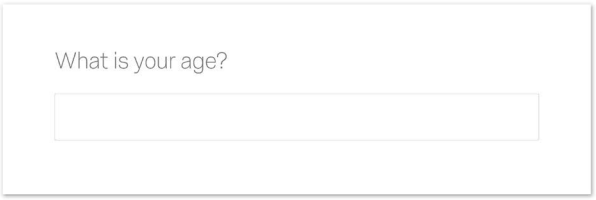



**FIGURE 4A**

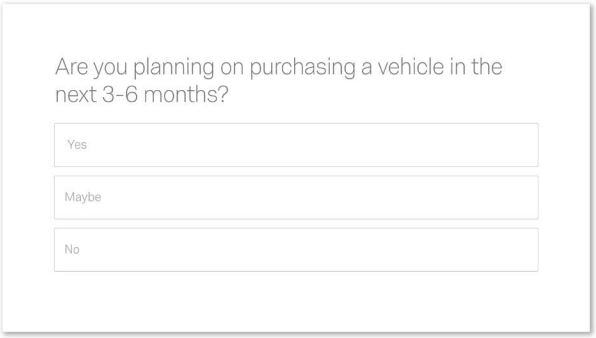



**FIGURE 4B**

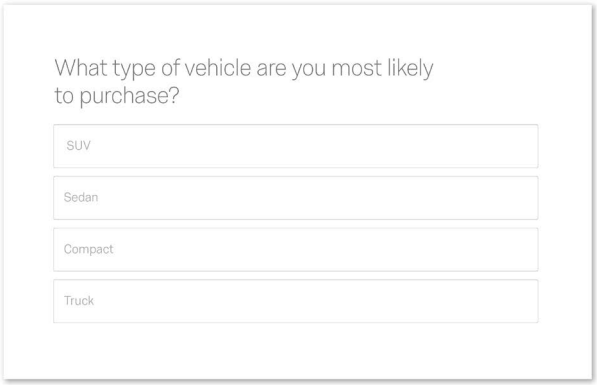



**FIGURE 4C**

Keeping the survey short typically means asking the questions that are necessary for your research goals and no more—there is no room for "pet" or "nice to know" questions if you want high-quality data. In general, we find that web surveys that take longer than 10 minutes are much more likely to suffer from low-quality data, as respondents fatigue and begin satisficing. Typically, respondents can provide about 30 responses (to average questions) in 10 minutes, but with any new or revised survey it is important to pre-test it yourself on a few people to see how long it takes.

## RESPONDENT MOTIVATION

In recent years, response rates to surveys have typically been extremely low. This suggests that those individuals that do participate and become respondents must have some amount of motivation to take the survey. Capitalizing on this motivation by taking steps to avoid diminishing it and also attempting to increase it whenever possible may not only help keep respondents in the survey but will also help them to provide higher-quality responses. There are five approaches that we recommend for getting the most out of your respondents' motivation:

1. Ask them to commit to provide their best responses
2. Leverage the importance of the survey
3. Leverage the importance of their responses
4. Use incentives and gratitude to increase engagement
5. Keep the duration of the survey short (10 minutes or less)

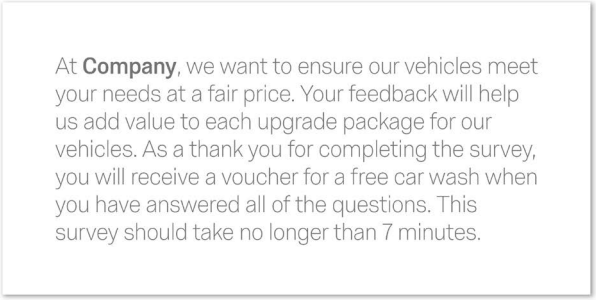



**FIGURE 5**

Early work in survey methodology indicated that asking a respondent to commit to providing their best data actually had positive effects on the quality of responses they gave (Cannell, Miller, and Oksenberg 1981). At Qualtrics, we have recently replicated this finding in the web survey context across 14 countries. To implement this, you can simply ask your respondents at the beginning of your survey if they will commit to providing their best data. We believe that, because people feel a desire to be internally consistent with statements and commitments that they have made (Cialdini, Cacioppo, and Bassett 1978), they are more likely to provide high-quality survey responses after they have committed to do so. In our 14-country study, we found that over 98% of respondents were willing to commit to providing their best responses when asked.

In terms of leveraging the importance of the survey, the best practice is to reaffirm the decision made by your respondents to participate by providing an indication that the topic of the survey is important. Similarly, you can remind your respondents that their responses actually matter and are important. Incentives can be an effective method of increasing respondent motivation as well. Reminding the respondent that they will be paid or will be entered into a lottery for a prize can not only keep respondents from leaving the survey, but can reduce satisficing and other negative response behaviors. Lastly, the duration of the survey affects task difficulty and respondent motivation simultaneously, making it critically important to keep the survey as short as possible. Generally, we recommend that web surveys not take the average respondent more than 15 minutes. If you're anticipating that many respondents may arrive at the survey using a mobile device, then the duration should be even shorter (probably not longer than 7 minute), to avoid large numbers of respondents losing motivation and breaking off from the survey.

# Designing questions to maximize data quality

## Choosing the best response format

In this section we discuss the most commonly used survey question types and highlight when each may be most appropriate. These question types include open-ended questions, ranking questions, and rating questions. Each type has a distinct set of benefits and disadvantages, and knowing when to use each can make a huge difference in the quality of your data. It is important to weigh these benefits and disadvantages carefully when designing a survey question.

### OPEN-ENDED QUESTIONS

Open-ended responses to questions can be some of the most reliable, valid, and rich data collected in a survey. Unfortunately, in the web survey context they are most often used in ways that don't allow the researcher to realize their full potential. Most commonly, they are used as 'Other (specify): _____' response alternatives to categorical questions for which the researcher either doesn't know the entire universe of possible responses or feels the list would be too long to present. The other very common usage is in the format of general "feedback" or "comments" boxes where the respondent is simply expected to type comments about anything related to the topic of the survey or the survey itself.

The reason that open-ended questions are not used more often is that respondents generally do not like them very much—this is because they are more cognitively demanding and time-consuming to provide high-quality answers to. But if you use open-ended questions judiciously, it is possible to avoid both excessive cognitive demand and long completion times. Asking open-ended questions that are very specific and easy to answer will allow you to realize the benefits of this powerful question type and avoid annoying or fatiguing your respondents. For example, in many cases, when a number is being requested, it is best to use an open-text question. Asking for a person's age in years and letting them type the number is easier for the respondent to do than selecting from a drop-down list. It's also more precise than selecting an age group.

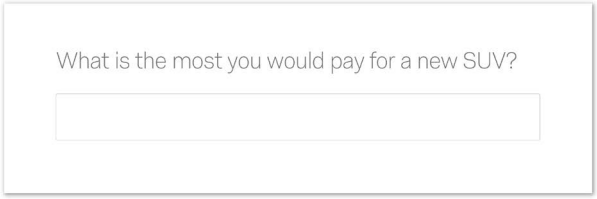



**FIGURE 6**

When using the validation options available in Qualtrics you can also ensure that only plausible responses are given (e.g., you can set the text-box to only accept numbers between 18 and 100). In general, it is best to not use a lot of open-ended questions because they can lead to respondent fatigue faster than other types of questions and are more likely to induce your respondents to leave the survey before completing it.

As mentioned above, the 'Other (specify): _____' response alternative is commonly used for categorical questions. Though it can potentially capture some additional responses that were not included in the categorical options provided, this approach also comes with a cost. Research indicates that respondents tend to select options that are provided rather than typing in their own response; this can lead to underestimates of the options that are written in by the respondents. Furthermore, the magnitude of the underestimate can be difficult to accurately assess. In general, the best practice is to use an open-ended response format when the full range of possible responses cannot be provided in a list for the respondents to select from or if the list would be so long that respondents might not carefully read each alternative. Using the 'Other (specify): _____' option should be avoided because the resulting data may be misleading.

## RATING

Rating questions are the most commonly used question type in web surveys. These questions obtain assessments of one object at a time, typically along a single dimension (e.g., satisfaction, importance, likelihood, etc.)

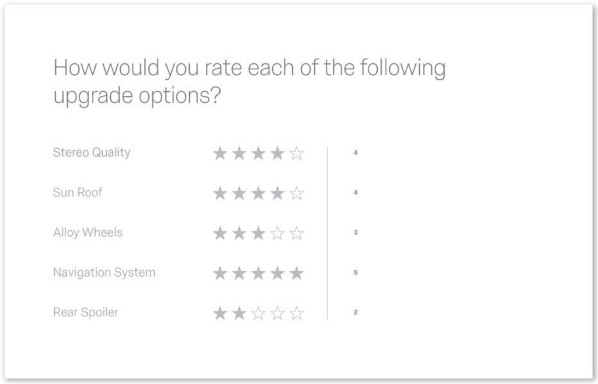



**FIGURE 7**

These questions are popular for a number of good reasons. First, they are comparatively easy for respondents to answer, both in terms of the cognitive burden of the question and the provision of a response. Unsurprisingly, respondents prefer the rating question type over ranking questions. Rating questions also generally have shorter completion times than ranking questions. Finally, the data from rating questions is typically more straightforward and easier to analyze than the data from ranking questions.

However, rating questions do pose some tradeoffs when compared with alternatives. The first is that lower effort on the part of respondents may produce lower data quality—this is the tradeoff of using questions that do not require as much cognitive effort on the part of respondents to produce a reasonable answer. The second downside is that responses tend to be a bit less reliable and change more over time. Lastly, rating questions are susceptible to response styles, which describes the tendency of some respondents to consistently avoid the ends of rating scales (or always give answers at the ends), give acquiescent responses, or give straightlined responses.

## RANKING

Ranking is a powerful and under-utilized question type that is becoming increasingly popular as researchers outside of the field of market research embrace conjoint designs for their projects. But even apart from conjoint, ranking questions have a huge amount of value for many kinds of research questions.

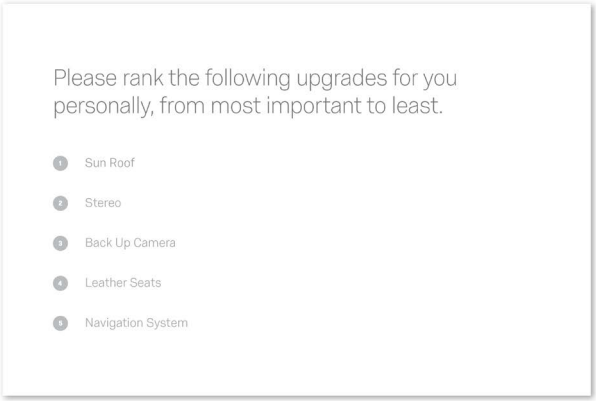



**FIGURE 8**

Ranking questions have a couple of key advantages. The first is that they provide comparisons between multiple things at one time. When a consumer enters a convenience store to purchase a soda, they will typically only purchase one of the many options available. They may report on rating questions that both Coca-Cola and Pepsi are equally preferred, but when faced with the decision they are likely to only choose one. A ranking question type is able to force the differentiation between these items in a way that a rating question type cannot. Consequently, ranking questions are useful when the desired outcome is a comparison or choice. When the goal is to evaluate relative performance, importance, preference, satisfaction, and many other measures, the ranking question format is worth considering.

The second advantage afforded by ranking questions is that they are often more reliable than rating questions, particularly for items at the ends of the ranking scale where respondents typically have the strongest differentiation. This means that respondents are able to repeatedly provide the same rankings of items more consistently than when they are asked to provide ratings.

## RATING VS. RANKING

When choosing between rating and ranking, it is important to evaluate which is most appropriate for the research being conducted. In some cases, ranking will be the best tool; in other cases, rating will be better. As a general rule of thumb, when life forces a choice between alternatives, ranking may be the better option. In these cases, you want the question type that more closely reflects the decision process that the respondent will engage in when making a choice. In most other cases, rating will be the best choice.

# Rating scale point decisions

## HOW MANY SCALE POINTS SHOULD BE USED?

The primary goal for choosing the number of points to include in rating scales is to differentiate between respondents as much as validly possible while still maintaining high reliability or consistency in responses (e.g., if the same question is asked twice in the survey, we would typically hope for the same response on the scale.) Determining the number of scale points is a balancing act, which creates a tension when trying to maximize data quality. Including more scale points might differentiate responses more, whereas fewer scale points might produce more reliability.

Fortunately, survey methodology research on this subject provides some guidelines for best practices that enable optimal validity and reliability. The results of this research suggest that the optimal number of scale points ranges from 5 to 9—with fewer points, you lose the ability to differentiate as much as you could between respondents, and with more scale points, the reliability of responses tends to drop off.

For most survey response scales, the optimal number of scale points will be either 5 or 7. If the construct being measured can range from zero to some positive value (this is called a unipolar construct), then a 5-point unipolar response scale is best.

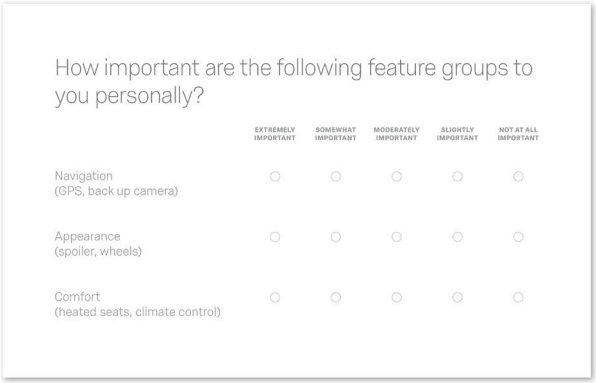



**FIGURE 9**

For example, in assessing how probable a customer is to make a return visit to a store or website, it does not make sense to try to measure a negative probability because, mathematically, probability ranges from 0 (definitely won't visit) to 1 (definitely will visit). So in this example the appropriate scale is a 5-point unipolar response scale. If the construct being measured can range from some negative value to some positive value (this is called a bipolar construct), then a 7-point bipolar response scale will be best. For example, in assessing satisfaction with a product or service, the construct (satisfaction) can range from extremely dissatisfied, a negative value of satisfaction, to extremely satisfied, a positive value of satisfaction. As a result, a 7-point bipolar response scale is the best choice for this question.

There is one important caveat to consider about the advice provided above. Web surveys are taken on more types of devices now than ever before, and an important consideration for web survey designers is the screen size on which their questionnaires are likely to be viewed. If many respondents are expected to come to the survey on a mobile device with a small screen, such as a smartphone, then it becomes important to ensure that response scales will render on the screen without any need for horizontal scrolling. In this case, 5-point response scales may be preferable to 7-point scales, even for bipolar constructs.

You may have noted that the recommendations above are both for scales with odd numbers of scale points, meaning that there will be a midpoint. Many researchers like to force respondents to choose one side of the scale or another, thinking that a midpoint response is equivalent to a "don't know" or " no opinion" response. But the empirical research actually indicates that this is not the case for respondents. In fact, reliability seems to be highest when a midpoint is provided, meaning that forcing respondents to take one side or another may introduce inaccuracy into the data.

## LABELING RESPONSE SCALE POINTS

There are a few different ways that response scales are labeled. Some use verbal labels and others use numeric. Some only have the endpoints labeled, others label the ends and the midpoint, and yet others label each point. So which approach is best?

It is important to keep the goals of scale point labels in mind when making this decision.
These goals are:

1. The meanings of each scale point should be easy for respondents to interpret
2. The meaning of each scale point should be clear (unambiguous)
3. All respondents should interpret the meanings of each scale point identically
4. The labels should differentiate respondents from each other as much as is validly possible
5. The resulting scale should include points that correspond with all points on the underlying construct's continuum

From these goals, the resulting best practices include labeling all scale points. The meaning of unlabeled scale points is ambiguous. As a result, in an answer scale with both labeled and unlabeled points, respondents may be attracted to the points that have labels. This tendency to select labeled scale points will produce a bias toward those points, resulting in clusters of respondents at labeled points. In short, partially labeled response scales may produce biased data due to respondents being attracted to scale points that are easier to interpret.

Another best practice is to verbally label rating scale points rather than numerically labeling them since numbers alone are often ambiguous in their meaning and difficult for respondents to interpret. In general, it is best to omit numeric labels for rating scales altogether, because they may be interpreted differently by different respondents.

Not only should the scale points be labeled verbally, but the labels should also match the construct being asked about in the question text. Researchers often attempt to use the same response scale for many questions, most commonly the agree-disagree scale. This practice may lead to less thoughtful answers on the part of respondents and may also make interpretation of the questions and response scales more complex. For example, if a researcher wishes to assess satisfaction, they may make the statement, "I was satisfied with _____," and provide a response scale ranging from "strongly agree" to "strongly disagree." In this case, the respondent must make an assessment about their degree of satisfaction and then determine how that assessment maps onto agreement with a statement, rather than simply answering a question about satisfaction.

A better approach is to determine the construct of interest (satisfaction, in this case) and then ask a question with response options that are specific to that construct. The resulting question using this approach would be, "How satisfied or dissatisfied were you with _____?" and the response scale mapped onto this construct would range from "extremely satisfied" to "extremely dissatisfied."

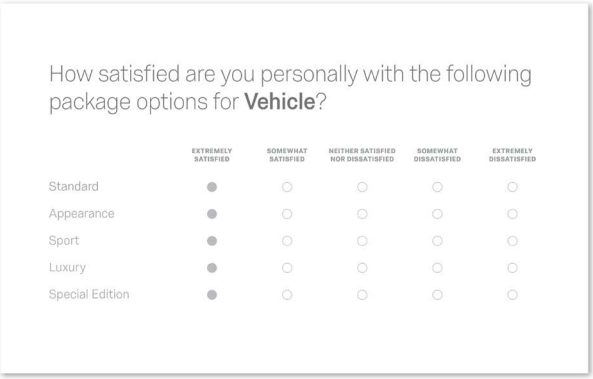



**FIGURE 10**

In this formulation of the question, using a construct-specific response scale, the respondent makes their assessment about their degree of satisfaction and then can provide an answer without needing to make the additional step of determining how their assessment of the construct maps onto agreement with a statement.

Another reason to use construct-specific response scales is that they may reduce straightlining. When a respondent is asked to give answers on the same response scale many times, there is an increased chance of fatigue setting in, and it becomes more likely that they will begin to satisfice by straightlining. This is most prevalent with matrix or grid question types, but can also happen when questions are presented individually.

Lastly, respondents presume equal spacing of scale points and the underlying continuum. This means that respondents expect that the response scale covers the entire range of possible answers along the target construct and that the scale points provided are evenly divided across that range. As a result, scales should be balanced across the continuum of possible responses and the labels chosen for the scale points should reinforce this. If a bipolar construct (negative to positive values) is measured using a unipolar response scale (zero to positive values), the lack of correspondence between the response scale and the construct will result in biased data.

# Question wording

Researchers have known for decades that question wording matters, and that even subtle changes in the wording of a question can produce substantial differences in the resulting data. With this in mind, it is important to use questions that have been validated or tested to ensure that the ways that respondents interpret the question match exactly what the researcher believes they are measuring. Without this testing, the question may inadvertently measure something different than intended without the researcher ever finding out.

## GOALS OF QUESTION WORDING

When writing a question, there are a few guiding goals that are worth keeping in mind. The first is to ensure clarity by only mentioning a single subject and construct – failure to do this can result in "double-barreled" questions. For example, a question that asks, "How satisfied or dissatisfied were you with our product selection and quality?" is problematic because it mentions two subjects: (1) the selection, and (2) the quality of products.

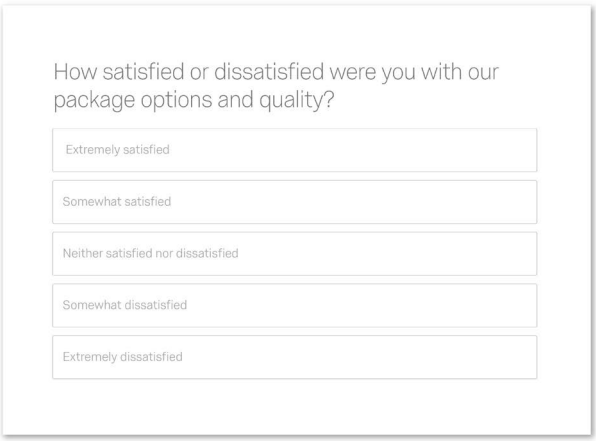



**FIGURE 11A**
Bad example

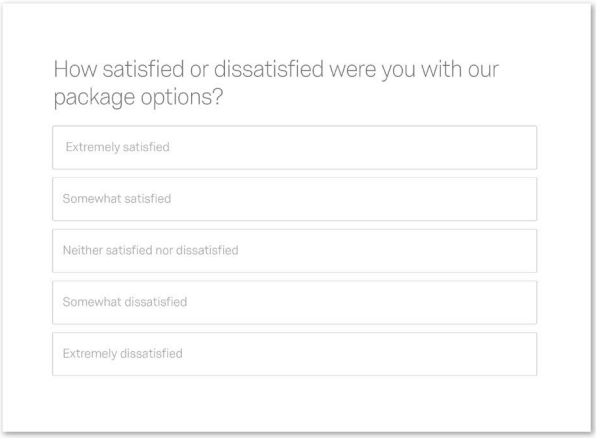



**FIGURE 11B**
Good example

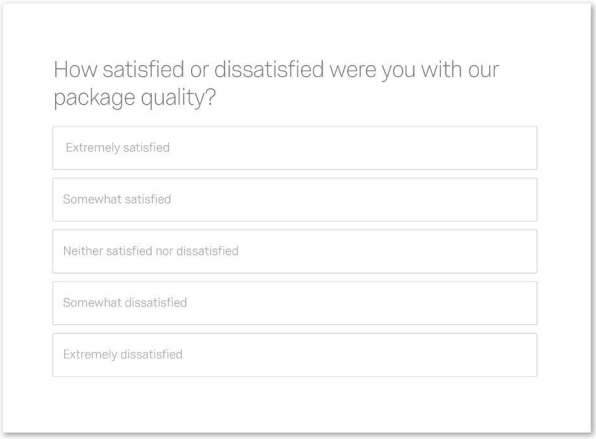



**FIGURE 11C**
Good example

A respondent may feel extremely satisfied with the quality of the products while being extremely dissatisfied with the selection. For that respondent, it is unclear how to respond to the question as it is worded. The better approach would be to create separate questions to address each construct.

The second goal is to have the question mean the same thing to all respondents. Put another way, each respondent should interpret the meaning of the question identically. If respondents have different interpretations of the meaning of the question, the resulting data will be invalid because it will be impossible to determine what question each respondent thought they were answering.

The third goal to bear in mind is to use words economically. It is important to use as many words as are needed to convey the idea of the question clearly to all respondents, but additional words present more opportunities to introduce elements that could confuse respondents. Keeping questions short also means that respondents can read them more quickly, which should help keep the survey duration short.

## WORD CHOICE GUIDELINES

There are a number of guidelines or best practices for word choice when writing survey questions. The first is that words used in survey questions should have only one meaning— this is easy to verify using a dictionary. Words and sentences should be simple, to maximize the ease of reading and comprehension. A useful rule of thumb is that words with fewer syllables and sentences with fewer words are typically simpler. Readability scores calculated using online tools are often useful for assessing the complexity of the words and sentences that form a question.

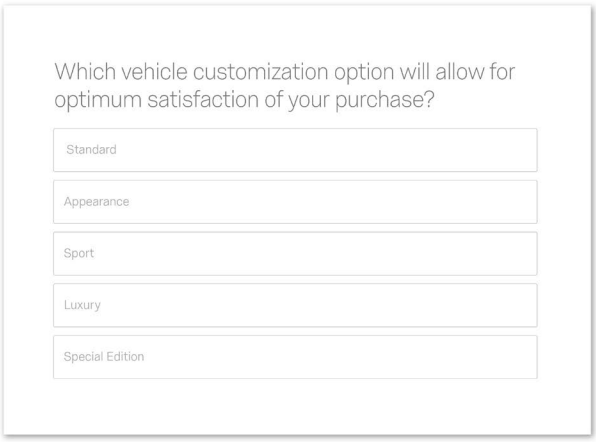



**FIGURE 12A**
Bad example

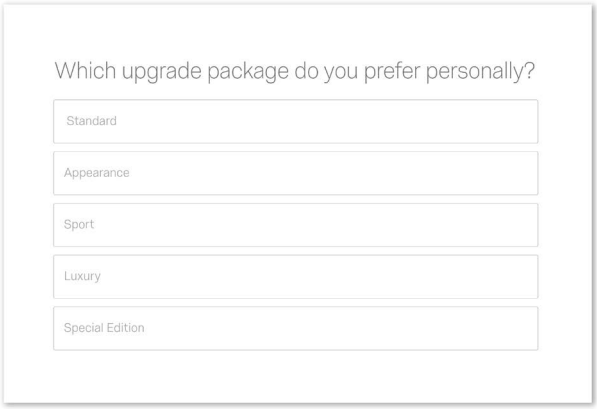



**FIGURE 12B**
Good example

The conventional wisdom—which has been supported by most empirical research on the topic over the years—suggests that, in general, questions should be worded to

- Be simple, direct, comprehensible
- Not use jargon
- Be specific and concrete (rather than general and abstract)
- Avoid ambiguous words
- Avoid double-barreled questions
- Avoid negations
- Avoid leading questions
- Include filter questions
- Read smoothly out loud
- Avoid emotionally charged words
- Allow for all possible responses

If you follow these recommendations, it is much less likely that your survey questions will confuse or frustrate your respondents, and your data are more likely to be valid and reliable. Surprisingly, many researchers fail to test their own survey questions by reading them out loud or asking others to state in their own words what they think a question is asking. These simple steps would solve many problems that we see with survey questions.

# Using an existing question vs. writing your own

In many cases, when deciding how to measure something, there may be existing questions, such as the Net Promoter Score questions or more domain-specific questions. Many researchers want to know whether or not they should try to find an existing question for their survey or if they should design their own. There are three general types of cases to consider when making this decision:

1. If there is an existing question that perfectly matches your research needs, use it.
2. If there is no existing question that perfectly matches your research needs, write your own.
3. If there is a question that seems to be a somewhat close match for your research needs, then consider using it and pre-testing it against one that you have written. If space allows, you may actually want to field both questions, but hopefully the pre-test is decisive.

It is certainly a best practice to use validated survey questions that have known measurement properties, but fielding a question that doesn't collect exactly the data needed to address your research question can be a waste of both yours and your respondents' time. It is important to remember that whatever question you ask should always collect the best possible data for your research question.

# Conclusion

The goal of this handbook has been to highlight some key best practices for question design and the rationale behind them. Additional specific recommendations for survey design can be found in the Survey Methodology 101 section of the Qualtrics Resource Library.

Writing good surveys questions is both science and art, and it's not intuitive for most people. Applying the best practices from the academic survey methodology literature outlined in this handbook (and from many other resources) will help you gather the most valid, reliable, and accurate data—and, as a result, the best insights.

In particular, being aware of what is going on in the heads of your respondents and how the design decisions made when creating a survey can either enable positive or negative response behaviors can play a powerful role in shaping the quality of your data. Question wording, response option wording, and response format can each have substantial effects on the quality of your data, particularly if they enable respondents to apply negative response strategies, such as satisficing.

Every question should be designed to collect the best possible data for the research question that it is intended to address. At a minimum, this may mean avoiding response formats and scales that enable satisficing response strategies. In other cases, this may mean designing new questions when existing ones do not perfectly match the needs of your research.

Throughout this handbook we have emphasized the importance of pre-testing any new or edited survey. This pre-testing process is the single most important tool that a researcher can use to catch errors, confusing or inadvertently challenging questions. Asking colleagues or friends to take a survey and provide feedback is a simple step that many researchers fail to take when writing a survey.

There are many potential sources of error that can affect the quality of survey data. Some of these are outside of the control of the researcher; others can be influenced directly or indirectly by decisions made in the design, fielding, and analysis phases of the research process. Question design is one of the opportunities that researchers have to directly influence the quality of the data that a survey is used to gather. With all of the other ways that data quality can be degraded, there is no reason to miss the opportunity that question design offers to maximize data quality. We hope that this handbook provides some actionable recommendations that will help you design better survey questions and generate higher-quality data and results for your research.

# References and additional reading


Cannell, Charles F, Peter V Miller, and Lois Oksenberg. 1981. "Research on Interviewing Techniques." *Sociological Methodology* 12: 389–437.

Cialdini, Robert B, John T Cacioppo, and R Bassett. 1978. "Low-Ball Procedure for Producing Compliance: Commitment Then Cost." *Journal of Personality.*

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5(3): 213–36.

Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50(1): 537–67.

Malhotra, Neil. 2009. "Completion Time and Response Order Effects in Web Surveys." *Public Opinion Quarterly* 72(5): 914–34.

Tourangeau, Roger, Lance J Rips, and Kenneth A Rasinski. 2000. *The Psychology of Survey Response.* Cambridge University Press.


## ADDITIONAL READING


Krosnick, Jon A. and Stanley Presser. 2010. "*Question and Questionnaire Design*" in the Handbook of Survey Research

Schaeffer, Nora Cate, and Stanley Presser. 2003 "*The Science of Asking Questions.*" Annual Review of Sociology

Sudman, Semour. and Norman Bradburn. 1996. "*Thinking About Answers*"

Vannette, David L, and Jon A. Krosnick. 2014. "*Answering Questions: A Comparison of Survey Satisficing and Mindlessness*" in The Wiley Blackwell Handbook of Mindfulness

Vannette, David L, and Jon A. Krosnick. (forthcoming). "*The Palgrave Handbook of Survey Methodology*"


[1] https://www.sesrc.wsu.edu/Dillman/papers/1997/A%20Theory%20of%20Self-Administered%20Questionnaire%20Design.pdf

ERE EMPLOYEE PRODUCT BRAND CUSTOMER EMPLOYEE PRODUCT BRAND CUSTOMER EMPLOYEE PRODUCT B

qualtrics EXPERIENCE MGMT™