Question Wording and Scale Response Guide

Statewide SURVEY Colloquium

INTRODUCTION

Constructing effective questions is central to the survey design process. This guide offers best practices for the questionnaire design process. It also offers guidelines on question wording and tips for well-constructed survey scales.

Writing good questions is as much an art as it is a science. Survey design is always about "writing questions for a *particular* questionnaire which means constructing them for a *particular* population, for a *particular* purpose, and placement next to another *particular* question" (Dillman 2010).

In other words, there are trade-offs between following guidelines and letting the research context, goals, and history guide the questions. To that end, while guidelines on question scales help, some situations necessitate retaining an existing (but less desirable) scale (e.g., when seeking year-over-year comparisons).

QUESTION WORDING CONSIDERATIONS

POSSIBLE REVISIONS

1 **Make sure the question applies to the respondent** *(i.e., use branching)*

NOT APPLICABLE TO ALL

If you made a trip on transit yesterday after 5:00 p.m., how many minutes did the entire trip take (door-to-door)?

[ ] minutes

1. **Did you make a trip on transit yesterday after 5:00 p.m.?**
   O Yes ➔ **go to Question 2**
   O No ➔ **skip to Question 3**
2. **How many minutes did the entire trip take? (door-to-door)**
   [ ] minutes
3. ...

2 **Ask one question at a time** *(i.e., avoid double-barrel questions)*

DOUBLE-BARREL QUESTION

To what extent do you agree or disagree:
"Transit should be faster and safer."

O Strongly disagree
O Disagree
O Slightly disagree
O Neither agree nor disagree
O Slightly agree
O Agree
O Strongly agree

1. **To what extent do you agree or disagree: "Transit should be faster."**
   O Strongly disagree
   O Disagree
   O Slightly disagree
   O Neither agree nor disagree
   O Slightly agree
   O Agree
   O Strongly agree
2. **To what extent do you agree or disagree: "Transit should be safer."**
   O Strongly disagree
   O Disagree
   O Slightly disagree
   O Neither agree nor disagree
   O Slightly agree
   O Agree
   O Strongly agree

3 **State positive and negative sides in the question stem**

ONE-SIDED QUESTION

Do you favor using tolls for corridor improvements?
O Favor
O Oppose

Do you favor or oppose using tolls for corridor improvements?
O Favor
O Oppose

4 **Make sure "yes" means yes and "no" means no** *(i.e., no double negatives)*

DOUBLE NEGATIVES

Do you favor or oppose not allowing toll increases without the prior approval of the Florida State Legislature?
O Favor
O Oppose

Do you favor or oppose requiring the approval of the Florida State Legislature for toll increases?
O Favor
O Oppose

5 **Use complete sentences with simle sentence structures**

INCOMPLETE QUESTION

Number of years lived in Florida

[ ] Years

POSSIBLE REVISIONS

How many years have you lived in Florida?

[ ] Years

6 **Ask for the specific unit desired in the question and provide answer spaces that are sized appropriately for the response task**

VAGUE UNIT

In an average week, how often do you take transit?

[   ]

In an average week, on how many days do you take transit?

[ ] Days per week (0-7)

SCALE AND RESPONSE CONSIDERATIONS

Survey scales allow respondents to express sentiments in a way that can be turned into hardcoded data. Good scales should be exhaustive, meaningful, and not leave room for interpretation by respondents. This guide discusses Likert scales which are rating scales that use psychometric categories to capture the intensity of an opinion or feeling across a symmetrically arranged series of statements. There are alternatives to Likert such as semantic differential, briefly explored at the end of this document.

Bipolar vs. Unipolar Scales

General rule: Bipolar constructs should be measured with bipolar scales, and unipolar constructs with unipolar scales.

One of the first things to decide when choosing an appropriate scale for closed-ended questions is what type of construct is being assessed.

- A **unipolar construct** includes responses that range from zero to positive (or when a single attribute can be absent or present). For instance, the likelihood of an event occurring or the effectiveness of a training program are unipolar constructs and are best measured with a unipolar scale.

- A **bipolar construct** can have responses that fall on one side or the other of a midpoint, where the midpoint is a neutral category. For instance, evaluations of "good/bad," "liking/disliking," or how "helpful/unhelpful" are best measured using bipolar scales.

> **> TIP: Using Midpoints with Bipolar Scales**
>
> Some researchers avoid bipolar scales with a midpoint. This avoidance stems from the assumption that a midpoint response is a "don't know" response in disguise or that respondents avoid taking a position because they lack motivation to respond thoughtfully. However, many researchers counter that offering a midpoint improves data quality and makes the responses more (not less) reliable and valid.

Number of Scale Points for Closed Questions

Research suggests that unipolar scales are most accurate when five-point scales are used. Conversely, bipolar scales are most accurate when a seven-point scale is used (with three scale points on each side of a truly neutral midpoint). Below are unipolar and bipolar scales with good measurement properties.

OPTIMAL UNIPOLAR RATING SCALES

- O Extremely
- O Very
- O Moderately
- O Slightly
- O Not at all
- ⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯
- O A great deal
- O A lot
- O A moderate amount
- O A little
- O None at all
- ⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯
- O Always
- O Most of the time
- O About half the time
- O Sometimes
- O Never

OPTIMAL BIPOLAR RATING SCALES

- O Extremely good
- O Moderately good
- O Slightly good
- O Neither good nor bad
- O Slightly bad
- O Moderately bad
- O Extremely bad
- ⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯
- O Like a great deal
- O Like a moderate amount
- O Like a little
- O Neither like nor dislike
- O Dislike a little
- O Dislike a moderate amount
- O Dislike a great deal

Statewide
SURVEY
Colloquium

Use Exhaustive and Mutually Exclusive Answer Choices

POSSIBLE REVISIONS

NEITHER EXHAUSTIVE NOR MUTUALLY EXCLUSIVE

From which one of these sources did you first learn about the closure of the Veterans Expressway between Exit 3 and Exit 13?

O Radio
O Television
O Someone at work
O While at home
O While traveling to work

Questions that are not exhaustive leave out answer options that some respondents can be reasonably expected to select. For instance, in the above example on the left, radio, television, and from someone at work are clearly not the only information sources that are possible. Mutually exclusive answer options posit that only one response option can be true for a single respondent. In the above example on the left, the answer options are not mutually exclusive since it is possible, for instance, to have heard about the closures on the radio while traveling to work. The question is therefore better split into two questions where all reasonable answers are covered, but only one answer per question can be true for each respondent.

From which one of these sources did you first learn about the closure of the Veterans Expressway between Exit 3 and Exit 13?

O Radio
O Television
O Internet
O Newspaper
O Another person
O Other, *please specify:* _____

Where were you when you first heard about the closure?

O At work
O At home
O Traveling to work
O Other, *please specify:* _____

Use construct-specific scales whenever possible

GENERIC SCALE

How often have you used TNCs (e.g., Uber, Lyft) during the past year?

O Never
O Rarely
O Occasionally
O Regularly

The construct of interest is frequency of using TNCs <u>during the past year</u>. While both scales tap into frequency of usage, the scale to the right consists of more concrete, specific time periods over the course of a year.

How often have you used TNCs (e.g., Uber, Lyft) during the past year?

O Not at all
O A few times during the year
O About once a month
O Two to three times a month
O About once a week
O More than once a week

Randomize Response Options if Concern Exists About Order Effects

For scale responses that are nominal in nature and do not have an implied order (e.g., modes of transportation, brands of cars), it is a good practice to always randomize response options. For scale responses with an implied order (age or income ranges, Likert scales), retain the order and do not randomize.

Labels for Scale Points of Closed Questions

For online or paper surveys, scale points should have clear labels; this requires labeling each point and not only the end points of the scale. Scales that label each point can help avoid ambiguity and are higher in reliability and validity and less susceptible to context effects (meaning that they are less susceptible to be influenced by aspects other than the question content at hand, such as the formatting of the question or prior, unrelated questions) . Numbered scale labels (e.g., 1 = strongly disagree, 2 = disagree) can be omitted if verbal labels are presented (e.g., strongly disagree, disagree).

For surveys conducted over the phone, to decrease respondent burden, labeling each scale point is not recommended. Instead, only label the most extreme points when reading the question (e.g., "On a scale from '1' to '5' where '1' means not at all satisfied and '5' means extremely satisfied").

Statewide
SURVEY
Colloquium

Using "Don't Know"

For factual type questions, offer a "do not know" option when it is reasonable to assume that not all respondents know the answer to the question (e.g., "How much did you pay on your most recent trip on the Beachline Expressway?"). However, for attitudinal or opinion questions, leave out this option when possible as research has shown that people will select "do not know" for attitudinal or opinion questions to reduce personal burden.

Using "Other, Please Specify"

Unless you can be certain that you are offering exhaustive answer options to a closed question, include an "other, please specify" option. However, respondents will still be more likely to select one of the offered answer choices than to provide their own response under "other, please specify."

Alternative Survey Scales

A common alternative to a Likert Scale is a Semantic Differential Scale that uses bipolar adjectives to describe opposite sides of a sentiment, perception, or likelihood and asks the respondent to indicate where their position lies on a numerical spectrum between the polarized adjectives.

Indicate how effective or not effective you feel the new overhead signs describe toll policies on the Veterans Expressway?

Not Effective						Effective
O	O	O	O	O	O	O
1	2	3	4	5	6	7

Like Likert Scales, five- or seven-point scales should be used to limit the response options to a manageable number of categories, and option labels should be numbered to simplify the respondent's task of assessing the spectrum. Although these scales are less dependent on the respondent's interpretation of words (e.g., in a Likert scale phrases like 'Moderately', 'Slightly' or 'Somewhat' can mean different things to different people), the scale intervals are still subjective and do not necessarily provide additional accuracy. Research has found that Semantic Differential scales are more demanding for respondents to complete than Likert Scales leading many survey practitioners to prefer Likert type scales that include well-designed and standardized response options.

REFERENCES AND ADDITIONAL RESOURCES

Dillman, D. A., Smyth, J. D., & Christian, L.M. (2009). Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method. John Wiley Co. Hoboken, NJ.

Krosnick, J. A., & Presser, S. (2010). Questionnaire design. In J. D. Wright & P. V. Marsden (Eds.), Handbook of Survey Research (Second Edition). West Yorkshire, England: Emerald Group.

Pasek, J., & Krosnick, J. A. (2010). Optimizing survey questionnaire design in political science: Insights from psychology. In J. Leighley (Ed.), Oxford Handbook of American Elections and Political Behavior. Oxford, UK: Oxford University Press.

Friborg, O., Martinussen, & M., Rosenvinge, J.H. (2006) Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. Personality and Individual Differences, 40, 873-884.

Statewide
SURVEY
Colloquium