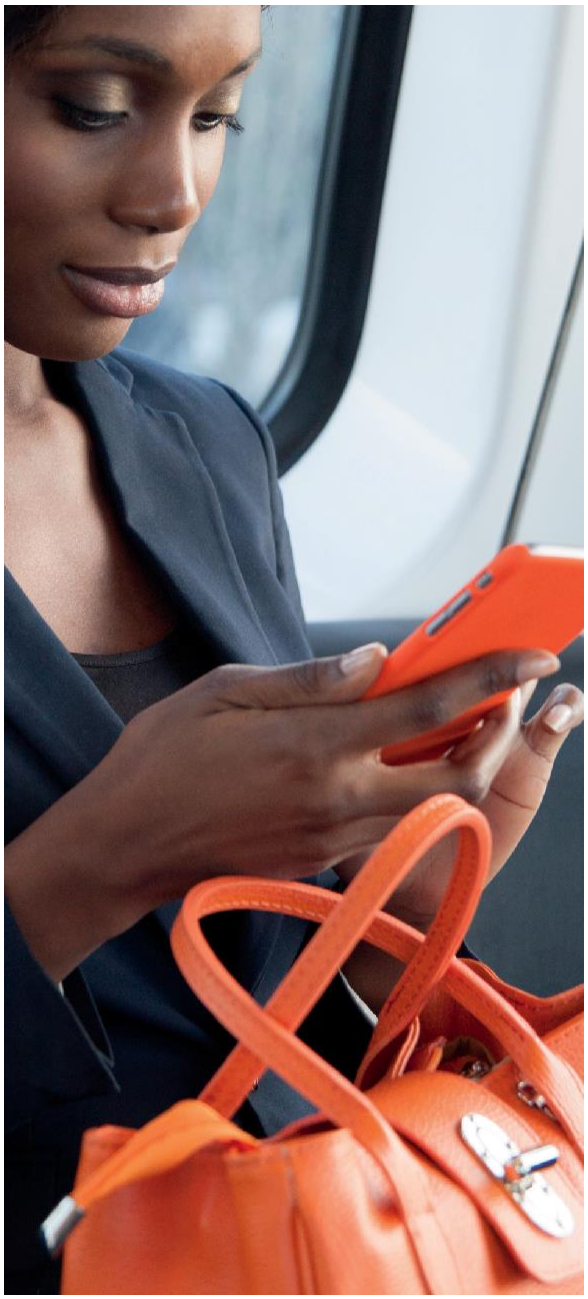
BEST PRACTICES IN MOBILE-FIRST DESIGN

RANDALL K. THOMAS,
CHIEF SURVEY METHODOLOGIST
& SENIOR VICE PRESIDENT

FRANCES M. BARLAS, PH.D.,
SENIOR RESEARCH SCIENTIST
& VICE PRESIDENT

January 2019

More people than ever are taking surveys online – and as smartphones have become go-to devices for accessing the internet, the incidence of taking surveys on mobile phones has also risen dramatically.



Currently, Ipsos finds that over 40% of general population samples are starting our surveys on a smartphone in the US. These respondents tend to be younger, have less education, and are more likely to be African American or Hispanic. Similar trends are occurring across other countries, as well.

Because researchers are accustomed to designing online surveys for desktop and laptop computers, taking a survey with a smartphone can be a frustrating experience for respondents. Questionnaires are often long, and screens may not render properly on small devices, leading to sub-optimal responses (straight-lining, item nonresponse, etc.) and higher rates of break-off.

Putting mobile devices – and respondents – first

As people's reliance on mobile devices continues to grow, it is essential that researchers put respondents and their smartphones at the center of questionnaire design.

When researchers incorporate a "mobile-first" philosophy into everything they develop, they will see:

- Improved respondent survey experience across all completion devices
  - Higher satisfaction and future participation
  - Lower levels of break-offs and less costly samples
  - Shorter surveys and higher satisfaction
- Higher quality and increased validity of the data
- More representative samples

GUIDING PRINCIPLES FOR MOBILE-FIRST ONLINE QUESTIONNAIRE DESIGN

The foundational principles of mobile-first surveys are easy to understand — and, in some cases, quite familiar. Many are new incarnations of challenges researchers have faced for decades — boosting motivation and minimizing frustration, detachment, and break-off among our most precious resource, our respondents.

Smart approaches to online mobile-first survey design fall generally into three categories.

1. Limit cognitive effort

Cognitive effort during survey taking is directly related to

- The total number of tasks that we ask respondents to complete and the length of time that it takes to complete them.
- The complexity of the tasks that we present to respondents.
- The extent to which respondents are asked to remember events accurately.
- The level of discrimination between responses that we require from respondents — more response categories require greater discrimination.
- The repetitiveness and redundancy of tasks (using the same rating scale repeatedly or asking multiple questions about the same concept); greater repetition requires additional cognitive effort to maintain attention.

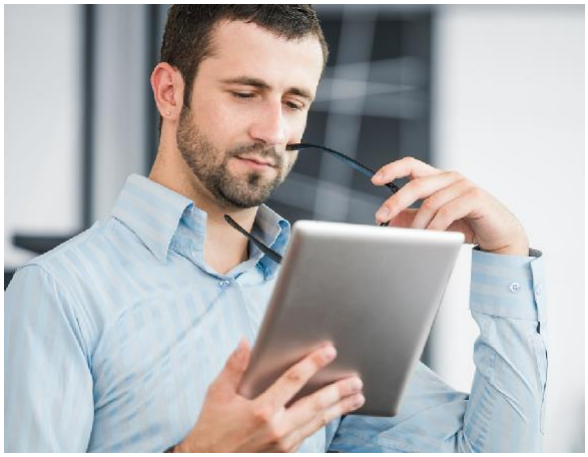
2. Reduce manual effort

Survey designers create manual effort for respondents when they fail to minimize the number and difficulty of manual tasks required to complete the entire survey (e.g., dragging and dropping elements, clicking into an open-ended text box and then typing or entering numbers into numeric boxes).

3. Make the survey motivating

Interest and incentives directly contribute to a respondent's motivation to complete a survey. Keeping the survey interesting and engaging and offering sufficient incentives will improve respondents' willingness to complete the survey.

Decreasing cognitive and manual effort will reduce respondent burden and the amount of time it takes to complete a survey. Making the survey motivating will improve respondents' attention to the questions and the accuracy of their answers.



SPECIFIC ONLINE QUESTIONNAIRE DESIGN RECOMMENDATIONS

For the last several years, we have been conducting research-on-research to determine best practices in designing online surveys, with a focus on improving the survey experience for respondents on mobile devices. The recommendations that follow are the distillation of our extensive learnings after having conducted over 100 experiments and assessing the feedback from thousands of respondents.

Limiting cognitive effort

Organize topics.
There should be a natural flow of content, so that topics and questions are not disjointed (which can often be disorienting to respondents).

Introduction and short introductory question. The introduction to a survey is a critical point; if you have a lot of text or begin with a difficult task, respondents will break-off quickly. We strongly recommend as short an introduction as possible, combined with a very brief dichotomous response item on the very first screen (see Example 1).

Chop unnecessary questions.
Focus on must-have questions; delete "nice-to- have" items.

Use transitions.
When changing topics, use very brief transition sentences or phrases. Present such transitions on the same screens as the next topic, not by themselves on separate screens. Placing transitions on their own screens increases manual effort and time to download, which adds to respondent burden.

Use easily understandable language.
Write the questionnaire for someone with lower cognitive abilities (think of writing for an average 10 year old). Use fewer words and words with fewer syllables.

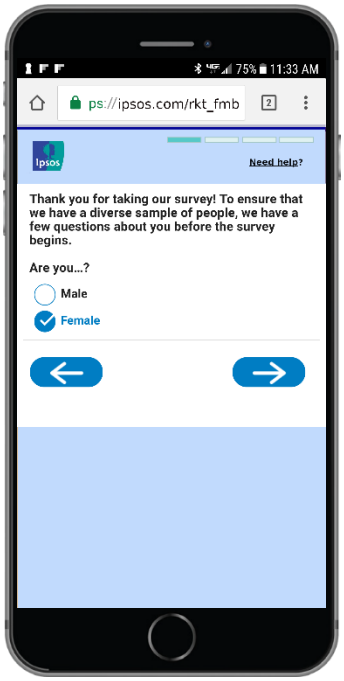
Keep all text short.
To minimize break-off, keep introductory, question and response, transition, and instruction texts as short as possible.

Use short sentences.
Do not express too many ideas in a single sentence. Break up long or complex sentences into shorter ones. Review each question, and delete sentences if they are not necessary. Finally, shorten and simplify the remaining words for each and every sentence.

Use shorter responses.
Eliminate any unnecessary words in your response options and shorten or simplify words.

Use the smallest number of responses possible.
Present fewer responses for the respondent to select from.

Example 1
Introduction combined with question



Use a 4-category unipolar, fully anchored scale. For attitude measurement, we have not found that scales with more numerous scale points, end-anchored scales, or bipolar scales are better or more valid than a simple 4-category, unipolar, fully anchored scale.

Reducing manual effort

Do not require respondents to scroll horizontally.
Redesign your questions and responses to minimize or eliminate horizontal scrolling on devices. Use the alternatives to traditional grids that we suggest below.

Use an optimal number of screens – in most cases, use fewer screens.
For example, present introductory text *with* a question, as mentioned above. Each additional screen in a survey is associated with higher rates of break-off. However, we also do not want to present too many questions on a single screen; scrolling down vertically to see more questions - particularly on a desktop or laptop - can also be frustrating for respondents and can lead to break-offs. Think about the optimal number of screens for your survey — questions should not be alone on their own screens; bank questions together, but do so sensibly.

Use clickable cells/radio buttons for response entry.
We have not found any more efficient method of response entry than the use of clickable buttons. Sliders, drag and drop, numeric entry, and drop- down boxes all make entering a response more time consuming than a clickable cell, and therefore introduce a higher level of respondent fatigue.

Allow respondents to complete a survey on their preferred devices.
Do not block respondents on mobile devices from completing particular items or from the entire survey. Also, do not ask respondents who are on a smartphone or tablet to take the survey on a desktop or laptop device.

Our research has found that most respondents will not switch to the device researchers request; and, once you ask them to take the survey on a device other than their preferred one, they are much less likely to start the survey in the first place.

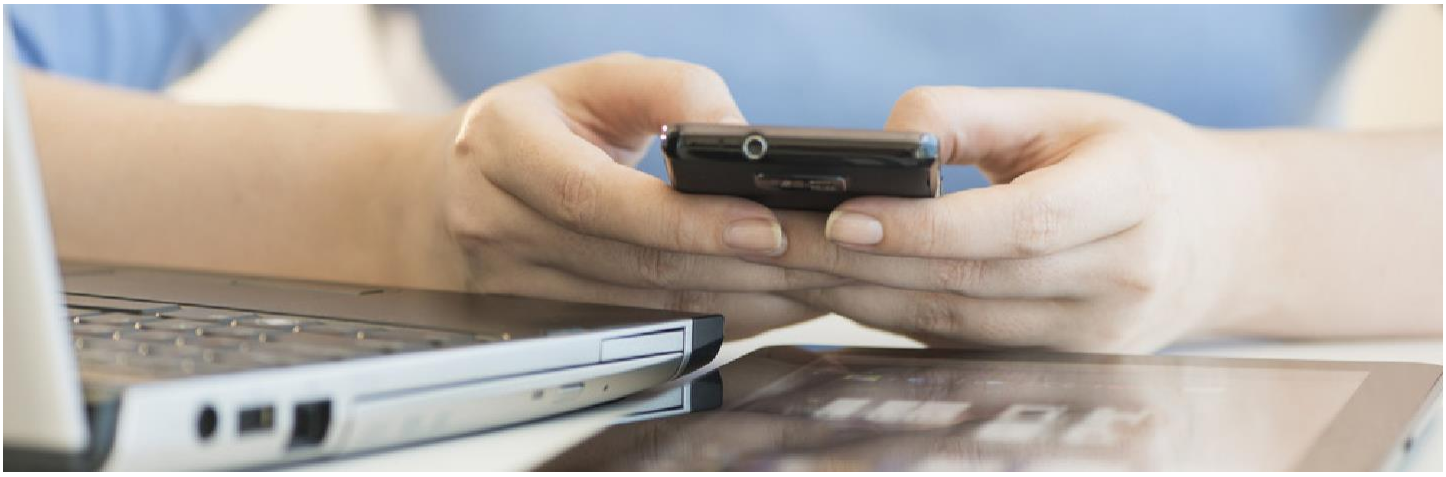
In addition, if respondents start a survey on a device that has been blocked, they are unlikely to switch devices and complete the survey. Due to the differential higher rates of non-completion on one device, trying to have respondents take the survey on a different device will reduce the representativeness of the sample, which can increase bias and lower the incidence of qualification, raising sample costs.

90%

of those taking a survey on a smartphone will start the survey in portrait mode.

Few will take the time to switch to landscape during the survey.

Allow respondents to use their devices with the screen orientation they prefer.
Do not ask respondents to change how they are holding their smartphones (such as asking them to switch from portrait to landscape mode). Our research has indicated that most people will not do as asked. About 90% of those taking a survey on a smartphone will start the survey in portrait mode; few will take the time to switch to landscape during the survey.



Making the survey motivating

Avoid redundancy.
Asking redundant items can affect respondents' levels of interest and increase break-off rates. Using redundant items is not necessary for either reliability or validity when conducting surveys with sufficient sample size. It is the number of people completing an item that produces reliable and valid results, not the number of items measuring the same idea. Redundancy can be eliminated through factor analyses or regression of items on a criterion to determine the best items to use.

Mix things up.
Variety is the spice of life. Respondents like tasks that change, and sometimes questions about different topics. Make sure you ask about more than what people do (their behaviors); also include questions about their attitudes and beliefs. Respondents like surveys that make them think more about their opinions.

Optimal length.
Less is more. Generally, lower levels of sub- optimal responding occur with shorter and well- designed surveys, interesting topics, and sufficient incentives. For minimal incentives, a well-designed and interesting survey can be between 10 and 15 minutes without showing any detrimental effects. Even with an increase in incentives, we start to see a rise in sub-optimal responses with surveys that are between 15 and 20 minutes, with issues becoming significant at over 20 minutes.

General principles

Unified design and presentation.
Present questions and responses consistently across all devices. Questions and responses should not be displayed differently on larger screens than on smaller screens. When we present response formats differently across devices (such as vertically on a smartphone and horizontally on a desktop), we may obtain different answers based solely on the variations in presentation - which is a device effect. Similar to mode effects, we need to avoid introducing device effects into our studies.

Present questions and responses consistently across all devices - on large or small screens.

Test your online survey on a smartphone.
Always test your survey on a smartphone in portrait mode, since this is the orientation that most people use when taking surveys on a smartphone. Ideally, we would test the survey with older, smaller versions of smartphones; if there will be a problem with presentation, it will occur on such devices. You should also test on both Androids and iPhones (besides normally testing with different browsers).

RECOMMENDATIONS FOR SPECIFIC QUESTION TYPES

Single-response ordinal/interval scales

A single-response scale, whether presented individually or in a grid, is one of the most commonly used forms of measurement in surveys. With a single-response scale, respondents select an answer among a set of ordered options that best describes their feelings, attitudes, behaviors, or intentions. They provide respondents a single dimension of judgment (e.g., importance, liking, satisfaction, likelihood, frequency). For self- reported measures of attitudes and intentions, no other form of measurement has been demonstrated to have higher validity or reliability than a *well-designed* single-response scale.

- We recommend using no more than five response categories, and three or four may be ideal for a smartphone design. **Our research has shown that in most cases using a scale with more than four categories does not improve the validity of the data.**

- **Unipolar, fully labeled scales with no neutral point** (e.g., 'Not at all important' to 'Extremely important' or 'Do not like' to 'Very strongly like') have been demonstrated to be generally more valid than bipolar scales with a neutral point (e.g., 'Very unimportant' to 'Very important' and 'Strongly dislike' to 'Strongly like'). They have also appeared to be superior to end-labeled scales (which give meaning only to end categories). **Unipolar, fully labeled scales have been associated with the highest validity and greater discriminability for survey measurement.** (See Appendix for recommended scale types.)

- If you use a 'Not sure', 'Not familiar,' or 'Not applicable' non-substantive response, make sure there is a visual separation between the graded response scale and the non-substantive response.

- If you have a number of single-response items, consider grouping them with at least a few per screen, rather than only one per screen. (See the guideline "Optimal number of screens" on page 5, paragraph 3.) Reducing the number of screens required for a survey reduces break-offs.

Unipolar, fully labeled scales have been associated with the highest validity for survey measurement.

- Single-response scales often can have response order effects; respondents' choices can be affected by the order of the presented options. Generally, we recommend controlling for response order effects by:

  - Randomizing the order of the responses. With a single-response ordinal scale, randomize the order of response presentation so that half of respondents receive the most-to-least order and half receive the least-to-most order (you just need to randomize once by respondent, not for every question); **or**

  - Presenting only a least-to-most order. This reduces the top-box bias and allows for better differentiation between rated targets, preventing bunching of respondents at the top end of the scale.

Single- or multiple- response nominal / categorical items

A categorical single-response or multiple- response item has responses that are not ordinally related. This might be a task that asks respondents to select one automobile feature (for single response) or all features (for multiple response) from a set that are most likely to affect their car purchases. Alternatively, the question could ask them to select, from a listing of national parks, the one or all that they are most likely to visit in the next year (see Example 2 for single response and Example 3 for multiple response).

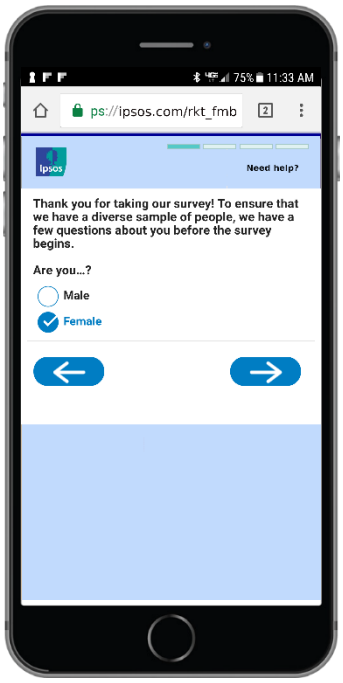
- If there is a natural way of organizing the responses – one that is normally how the respondents would quickly understand the responses, such as alphabetical order – use that method. For example, states or provinces are generally presented alphabetically in most forms and surveys.

- If there is no natural order for organizing the responses, then we strongly recommend randomizing the order of presentation, such that each respondent sees a different order, to control for order effects. If respondents only need to choose one response from a visual list, they are more likely to read and select from the first few options presented rather than considering each item in a long list.

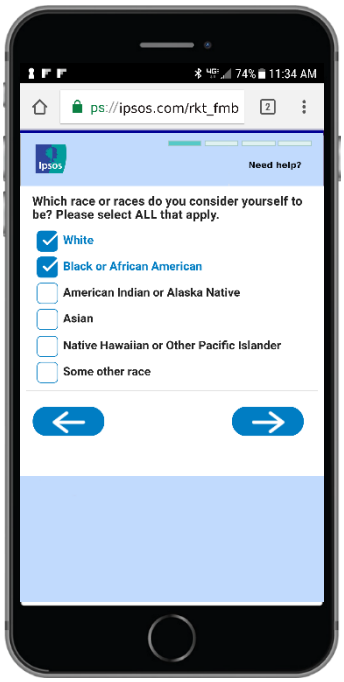
- Unless you can ensure that at least one response will apply to a respondent, a 'None' response should be presented after the responses. A 'None' response should be separated and visually distinguishable from the others. This category may not be needed if your instructions include something like 'Please select the one response from the responses listed below, even if the best response for you is not listed.'

Grids can be deadly boring for respondents, and are not very mobile friendly.

Example 2
Single-response format



Example 3
Multiple-response format



Grid questions (single response per item)

A grid question presents a number of elements that respondents are to evaluate, typically organized in rows. The responses are the same for each element and are most often shown in columns. A grid question is one of the most efficient measurement tools we have in online survey design. **However, grids can be deadly boring for respondents if carried to the extreme, and are not very mobile friendly.** (See Example 4 on page 9.)

- In designing the response scale for a grid, follow all recommendations concerning single-response ordinal/interval scales above.

- Keep question text, element text (the items presented in rows), and response text short.

- Because grids take up a lot of horizontal space (with elements to the left and responses to the right), they are often very difficult for smartphone respondents. Based on our research, the preferred default alternative display is an accordion format (see Examples 5 and 6 on page 9).

This allows for longer item and response text and more response categories, if necessary. The accordion format also takes up less vertical space than other options.

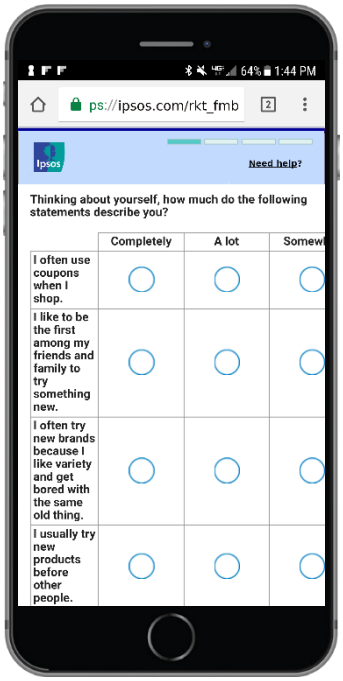
- The second default option is to present the items as banked items (with responses horizontally displayed under each item – see Example 7). This banked approach can present up to 4 or 5 *short* responses but takes up more vertical space than the accordion format. Banked is not a viable response format if scales have more than 5 categories or utilize long words; it should only be used with very short scales.

- The third, and least preferred, option is to use the traditional grid with floating headers. Floating the headers as you scroll up and down improves the respondent experience to a small degree. As we mentioned, the traditional grid generally takes up too much horizontal space and often requires horizontal scrolling to accurately respond. Rarely can more than 3 short responses be shown on the screen of a smartphone, even with short elements to rate.

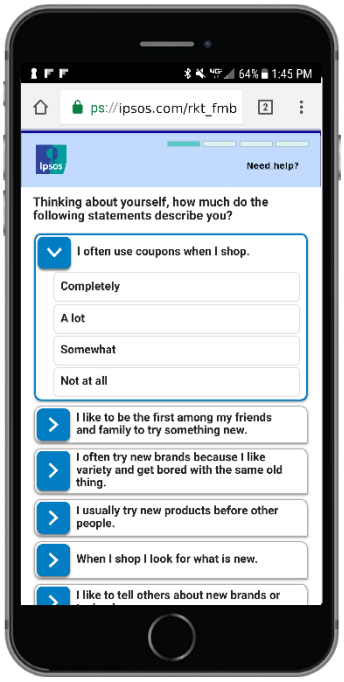
The simpler you can make the task presented on the screen, the higher the quality of your data and the better the respondent experience.

While some researchers may be tempted to pack as much as possible onto a single screen, complex grids should be broken apart into simpler, separate questions. A complex grid might have a person provide multiple ratings for each element along a row (sometimes using drop-down boxes for each column of evaluation). The simpler you can make the task presented on the screen, the higher the quality of your data and the better the respondent experience.

Example 4
Traditional grid

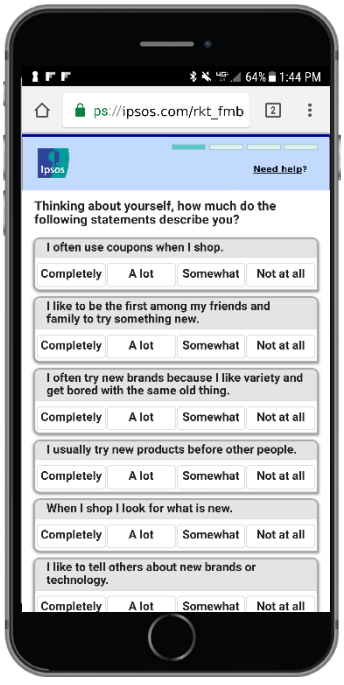


Example 5
Accordion format



Example 6
Accordion format –
earlier elements have been answered

Example 7
Banked item format



Grid questions (multiple response per item)

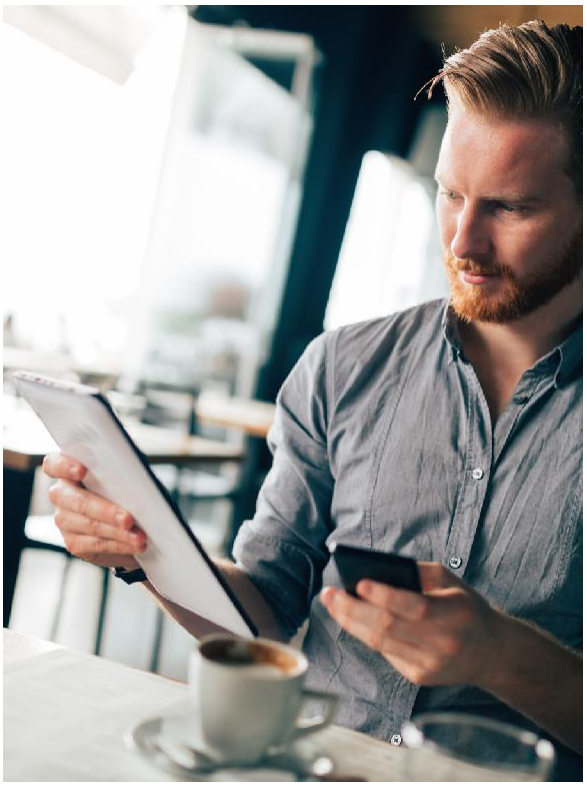
- First, review all recommendations concerning multiple response items above.
- The preferred default is an accordion format (see Examples 5 and 6 on page 9) with responses displayed vertically under each element.
- We do *not* recommend a traditional grid for multiple responses; respondents find multiple response in a grid format very confusing.
- Respondents are generally accustomed to giving only a single response per row due to the pervasive use of single-response grids. As a consequence, we see many respondent errors with a multiple-response grid.

Numeric box

- Numeric boxes add to respondent burden because it takes longer for respondents to answer these types of questions (mostly due to having to manually enter the numbers) when compared to radio buttons/clickable cells that have responses with numeric ranges (see Example 8).
- If you do use numeric boxes, we recommend doing so sparingly.

Allowing wider ranges and permitting out-of-the-ordinary responses also lets you evaluate data quality.

- Numeric ranges should allow for all possible responses, not just the values you want. In other words, set a wide or liberal range.
- Allowing wider ranges and permitting out-of- the-ordinary responses also lets you evaluate data quality. Outliers can be capped at the top or bottom of the expected range for analyses.



Example 8
Numeric format

Text box

- Having respondents type in responses is time consuming and more effortful. We recommend only including them if you are going to take the time to code the responses.
- Note that there can be a default maximum number of characters for text boxes – often 1024 characters. If you think that responses might be longer, you will need to specify an increase in the character limit.
- If you want respondents to list a number of elements (such as companies, brands, or reasons why), you should use a series of one- or two-line boxes, rather than one big box. You will get more responses with separate, individual boxes than one large box.

Comparative dependent tasks: Constant sum and ranking

Two of the most common comparative dependent tasks used in survey research are constant sum and ranking tasks. '**Comparative**' means that evaluations of a specific element take into account a consideration of the other elements. '**Dependent**' means that the number (evaluation) assigned to one element affects the numbers assigned to the other elements.

A **constant sum** methodology provides respondents with a fixed number of points that the researcher asks the respondent to allocate among a set of elements. As an example, respondents might be asked to distribute 100 points among a set of features of automobiles that would affect their purchase of a car. Respondents assign points to each feature based on its perceived importance to them but cannot assign more than 100 points (the constant sum) across all features. The numbers chosen are dependent – the higher the number of points assigned to one feature, the fewer the points that can be assigned to any other.

For a **ranking task**, respondents assign an order to features/elements based on a dimension of judgment. Similar to the constant sum example,

respondents can rank order a set of features from most important to least important in their influence on purchasing a vehicle. The numbers assigned are again dependent – when a feature is given the highest rank, this means that the other features must be assigned lower ranks (each value is used only once).

- We generally do not recommend using comparative dependent tasks if you can avoid them for the following reasons:

  - Tasks like constant sum or ranking are more difficult for respondents than rating tasks; they take respondents twice as long to answer as a comparable set of rating scales – especially as the number of elements to rank increases. A large part of this is the amount of cognitive effort it takes to compare elements and the increased manual effort to enter more responses.

  - There has been no substantial empirical evidence that constant sum or ranking tasks yield better, more accurate, or more valid information than an ordinal scale with independent ratings, when the scales are well designed (e.g., using unipolar fully labeled scales).

  - You lose statistical power with these kinds of comparative tasks; because of the dependent nature of the numbers assigned, less powerful statistical analyses must be used than for scales with independent ratings. Therefore, to detect differences among the elements being evaluated, you will need a larger sample size than with independent ratings.

- If you use a constant sum or ranking task, we recommend numeric boxes for response entry.
- If you use a constant sum or ranking task, we do not recommend using more than 5 elements to rank or to allocate points – respondents have more difficulty with more than 5 elements.
- Because they take less cognitive and manual effort, a couple of reasonable alternatives to constant sum or ranking tasks are:

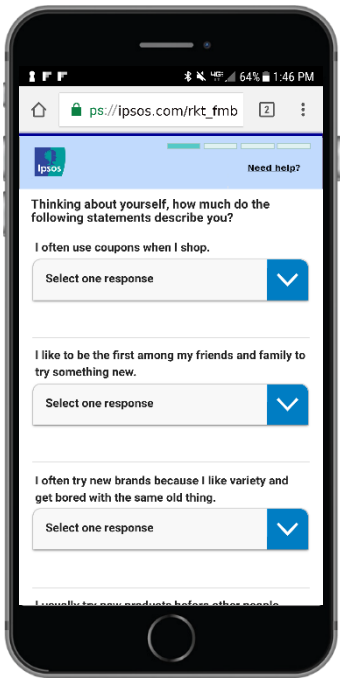
A max-diff task – respondents select the elements with the most and least presence of the dimension (e.g., most or least important). This is easier for respondents and takes less time to click responses.

A top-two task – respondents select the top and next ranks among a set of elements. Like max-diff, this is easier for respondents and additionally provides better differentiation at the top end of evaluations.

Drop-down boxes

- We generally do not recommend the use of drop-down boxes – empirical evidence shows much higher error rates in response selection. They can also be difficult for some respondents to scroll through on smartphones – and, as a result, are associated with higher rates of nonresponse on smartphones (see Example 9).
- We recommend using clickable lists in which all responses are displayed – even for long lists, such as state of residence.
- If you must use drop-down boxes, save them for lists that are naturally ordered (such as alphabetized states or order of months in year).

Example 9
Drop-down forma



Interactive scales and other tools

Generally, we do not recommend using interactive measures (such as sliders, drag and drop).

- Interactive measures take longer to complete and do not yield more valid results than alternative, simpler measures.
- Some devices may not render interactive measures properly, which could interfere with respondents' ability to complete the task.
- If you do use interactive tools, we recommend very sparing application; be sure to provide an alternative for those who are unable to use the tool.

Next steps

As people's use of smartphones – and mobile devices of future generations – becomes more pervasive, researchers will need to revisit these guidelines, with an eye always to making the respondent experience as simple and rewarding as possible. We expect that technology may prove both an ally and, at times, an obstacle in achieving these all-important goals; it is up to researchers and their clients to keep a close watch on that balance.

Randall K. Thomas
is Chief Survey Methodologist and Senior Vice President in the U.S. Public Affairs team at Ipsos. Randall K Thomas is an internationally recognized survey methodologist and is responsible for overseeing the quality of research processes and projects for Ipsos. He serves as a special consultant on projects requiring challenging multi-mode studies and custom panel designs. Trained in social psychology at Ohio State University, Randall has been one of the most prolific researchers on measurement using web-based surveys for the past 15 years and has more than 25 years of experience in experimental survey methodology across modes of interviewing and countries. His research interests focus on measurement of attitudes, intentions, and behaviors across devices and modalities.

Frances M. Barlas, Ph.D.
is Senior Research Scientist and Vice President in the U.S. Public Affairs team at Ipsos.
With over 15 years of experience in survey research, Frances M. Barlas is a leading survey methodologist in online surveys. She is responsible for conducting research-on-research to improve all aspects of survey methodology at Ipsos, focusing on survey sampling, weighting, and measurement. Frances also leads research and development initiatives to advance the operational efficiency and statistical integrity of Ipsos' probability-based KnowledgePanel® and other Ipsos research assets. She received her Ph.D. in sociology from Temple University.

Questions? Contact us!

Randall K. Thomas randall.thomas@ipsos.com
Frances M. Barlas, Ph.D. frances.barlas@ipsos.com

APPENDIX

Recommended scales for most commonly used measures

Intensity Measures	
Importance	How important is X? • Not important • Somewhat important • Very important • Extremely important
Usefulness	How useful is X? • Not useful • Somewhat useful • Very useful • Extremely useful
Describes	How much does X describe you? • Does not describe • Describes somewhat • Describes very well • Describes extremely well
Applies	How much does X apply to Y? • Does not apply • Somewhat applies • Very much applies • Completely applies

Evaluative Measures	
Quality	How good is X? • Not good • Somewhat good • Very good • Excellent
Liking	How much do you like X? • Do not like • Somewhat like • Strongly like • Very strongly like
Satisfaction	How satisfied were you with X? • Not satisfied • Somewhat satisfied • Very satisfied • Completely satisfied
Agreement	How much do you agree with X? • Do not agree • Somewhat agree • Strongly agree • Very strongly agree
Support	How much do you support X? • Do not support • Somewhat support • Strongly support • Very strongly support

APPENDIX (continued)

Recommended scales for most commonly used measures

Behavioral Measures	
Likelihood	How likely will you be to do X? • Not likely • Somewhat likely • Very likely • Extremely likely
Frequency	How often do you do X? • Never • Sometimes • Often • Very often
Extent	How much do you do X? • Not at all • A little • Some • A lot