DISM
DUKE INITIATIVE ON
Survey Methodology

TIPSHEET – IMPROVING RESPONSE SCALES

As important as it is to have good question wording in a survey, it is equally important to have well constructed and reliable question response options. Researchers should first consider whether open- or closed-ended questions are more appropriate to gather the desired data. Surveys typically rely more on closed-ended questions, so researchers need to be thoughtful in how they construct closed response options. This document discusses points researchers should consider when writing responses and offers tips on how to improve those scales.

Open-Ended or Closed-Ended?

Open-ended questions pose a question to respondents, but allow them to answer in their own words. Closed-ended questions provide respondents with pre-set answer options, and the respondent must select the most appropriate response(s) from among those listed.

Open-ended questions allow respondents to provide more nuanced answers that include a great deal of detail and rationale. Researchers often find open-ended questions useful for exploratory research in which the responses serve as a guide in writing closed-ended questions for future surveys, developing hypotheses, and shedding deeper light on why certain patterns are found in closed-ended results. The disadvantage of open-ended responses is that researchers must develop a reliable method for content analyzing those responses and devote resources to the actual content analysis. This can be a time consuming process and is one of the reasons why closed-ended questions are often preferable.

Closed-ended questions offer their own advantages. Analysis of these questions is often easier than for open-ended items because no content analysis is necessary. Closed-ended questions also help standardize the survey process by presenting systematic cues to respondents that are lacking in open-ended items. The main disadvantage of closed responses is that they offer the researcher little ability to get below the surface of the superficial response and uncover the thought processes that led to it. Also, elites or experts tend to find closed-ended questions frustrating. People possessing a lot of knowledge about a specialized topic often want to share more of their

knowledge and opinions than a closed-ended question allows, which is one reason why researchers might prefer open-ended interviews for these populations.

Most surveys contain a mix of open- and closed-ended items. A good rule of thumb is to make closed questions the default. Use them as much as possible, but include a few open questions if deeper insight into respondent thought is essential or useful on a small subset of questions.

Don't Know/Unsure

Researchers should think carefully about giving respondents the option to answer "don't know" (DK) or "unsure." Survey analysts typically treat these responses as missing data, resulting in fewer cases for analyses. As respondents tend to choose DK/unsure for systematic reasons, the exclusion of these cases tends to affect analyses significantly. Here are some points to consider when thinking about whether to include DK options:

- *DK/Unsure may be a valid response, especially for knowledge and recall items.*
  Some respondents may genuinely not know their response, especially on knowledge or recall items. Forcing these respondents to choose a substantive response may result in haphazard guessing and, therefore, lower quality data. However, the researcher may prefer the respondent to make an educated guess in these instances. Think carefully about why respondents would choose DK on any question before offering it as an option.

- *DK/Unsure is often a way to satisfice on attitude questions.*
  Respondents do not always have strong or consistent opinions about every topic. This is especially true when respondents are ambivalent, meaning they have a mixture of positive and negative considerations about the issue at hand. Ambivalent respondents might express an opinion if they are forced to do so, but might otherwise choose the DK response as a way to avoid sorting through the considerations they have in mind. The same is also true of respondents who, while not ambivalent, rarely express a certain attitude. Their feelings are not necessarily accessible at the top of their minds, but they may give a response if forced to think about their attitude. In these instances, the researcher may prefer to force a response rather than allowing respondents to choose DK on topics they are ambivalent about or think about rarely.

- *Sometimes DK/Unsure is of substantive interest.*
  Researchers are often interested in the DK responses themselves. In these instances, of course, DK should be offered.

- *If DK/Unsure is offered, expect people to take it.*

DK is an easy out for many respondents for various reasons, so always expect significant numbers of respondents to choose it if is available.

- *Modality affects DK/Unsure rates.*
  DK selection rates are greater in certain survey modes. Rates tend to be highest in self-administered formats such as web and mail surveys that lack a human interviewer. DK rates tend to be lower in formats like face-to-face and phone where a live interviewer can probe respondents for an answer if they initially choose DK.

- *Never include DK and Unsure as two separate options.*
  DK and Unsure should always be the same selection, though researchers may choose to vary the exact phrasing of the response option. Researchers often develop rationales as to why respondents might choose DK rather than unsure, then include both in the survey as separate options. In practice, respondents do not clearly differentiate between the two answers. Some respondents choose DK for the same reason as others choose Unsure, so there is no clear differentiation between them most of the time. Avoid confusing respondents and making the survey more complicated by combining these.

Middle Positions on Scales

Researchers often debate whether to include a middle option on their response scales. For example, imagine a question that asks to what extent respondents agree or disagree with a statement. The researcher might offer a Likert-type response scale: *agree strongly, agree somewhat, disagree somewhat, and disagree strongly*. Here the researcher might be interested in offering a middle option at the scale midpoint – *neither agree nor disagree*. This midpoint, like most, works similarly to the DK/Unsure option as it gives respondents the freedom not to choose a more substantive answer.

The pros and cons of midpoints are similar to those of DK/Unsure. It may be good for data quality in some instances because it best reflects the attitudes of some respondents. On the other hand, offering a midpoint invites respondents to choose it, and they usually will in significant numbers. Respondents who are ambivalent or who have weak opinions may choose the midpoint, but might easily select a more substantive category were they forced to do so.

Midpoint selection is also affected by the number of response options around it. Respondents are less likely to take the midpoint as the number of other options increases. Thus, midpoint rates will normally be greater on a 5-point scale than on a 7-point scale.

Midpoints are different from DK/Unsure because researchers tend to keep them in analyses rather than discarding these cases as missing data. Think carefully about offering a midpoint on any item because it tends to affect analyses substantially. Is the midpoint of substantive interest? Why will respondents select it? Is it preferable to force respondents to choose other options?

Caution: Researchers sometimes offer both a scale midpoint like *neither agree nor disagree* and a DK/Unsure option on the same question. This can be a dangerous decision. Much like separating DK and Unsure, researchers may develop clear rationales in their minds as to why respondents might choose the midpoint versus DK/Unsure, but respondents do not always make those same distinctions. It is usually the case that some respondents choose the midpoint for the same reason as others choose DK. This conceptual overlap is especially dangerous because analysts usually retain respondents at the midpoint in their analyses. When some respondents are discarded for the same reason as others are systematically retained, this biases the results in unknown ways. It is always advisable to pretest questions of this sort. If respondents are choosing the midpoint and DK for the same reasons, discard one of these options.

Ranking Scales

Researchers often want respondents to rank options, perhaps from most preferred to least preferred or most likely to least likely. Ranking items sound simple on paper, but they are often a source of significant response error from respondents who fail to follow even the simplest ranking directions. A good rule of thumb is that unless a ranking question is very simple or unless respondents can be forced to answer correctly (e.g. in a web survey the respondent cannot proceed with the survey unless options are ranked in a certain way), then consider alternative question formats. Here are some examples of ranking items of varying difficulty:

> *Here are some factors people often consider when deciding where to live. Which is the most important to you?*
> - *Proximity to work*
> - *Quality of schools*
> - *Quality of public parks*
> - *Safety of a neighborhood*
> - *Access to shopping*

This is a pretty simple item. Respondents may have a rank order in mind, but they are only required to select or rank one of the options. Error on this item will likely be low.

Let's change the directions on the same question to this:

> *Please rank the importance of all 5 considerations on a scale from 1 to 5, where 1 means "most important" and 5 means "least important."*

This ranking method is a bit more complicated. Rather than choosing the one consideration that is most important, respondents must now rank all five. Though not terribly complicated, these directions introduce room for error. Respondents may leave some items unranked. Researchers must then decide whether to include these respondents in analyses and how to treat unranked options. Some respondents may also have a hard time deciding how to rank, so they may give two or even three considerations a "1" with the intention of saying that multiple items are equally important. Researchers should also expect higher item nonresponse on this type of ranking scale than on a simpler one where respondents rank only one consideration.

Finally, let's change the ranking directions to another commonly used ranking method:

> *Please rank the 3 considerations that are most important to you, where 1 means "most important" and 3 means "least important."*

These directions complicate the ranking even more. Not only must respondents rank multiple items, but they are explicitly told to leave some considerations unranked. Again, these directions introduce more room for error. Some respondents may rank more or less than three considerations and give equal rankings to some items. Researchers are then in the complicated situation of determining what to do with respondents who express ranking preferences, but who fail to follow directions.

Ranking scales can challenge respondents. Web surveys can usually be programmed to force respondents to follow ranking directions. Interviewers in face-to-face and phone surveys can also correct the respondent if they do not follow the ranking directions. However, complicated ranking directions and multiple attempts at ranking can frustrate the respondent and increase the probability of a break off interview. Researchers should construct ranking scales carefully and always pretest them before including them in a survey instrument.

Other Considerations

Positivity Bias. Respondents tend to overreport positive attitudes and evaluations. Researchers can normally do little about positivity bias itself, but they should try not to exacerbate it. For example, researchers commonly ask respondents to rank objects (e.g. their health, politicians, cars, etc.) on numerical scales. These questions are fine, but the numbers offered to respondents tend to affect their answers. Researchers should avoid asking respondents to use negative numbers in their rankings as the negative sign exaggerates positive reports. The negative sign is a

signal of negativity that turns off many respondents. For example, respondents will rank their health much more positively on a scale from -5 to 5 than on a scale from 0 to 10.

How many points should be on a scale? On ordered scales researchers must strike a balance between offering too many and too few options. We want adequate discrimination between response options, but offering too many can blur the difference between scale points. Research has shown that Likert scales with 7 points and a midpoint tend to be most reliable, though 5-point and 3-point scales are also common. It is always wise to pretest questions to ensure that respondents are able to differentiate response options clearly.

Response options should be clear. Imagine we want to know how many jobs respondents perceive as being available in their towns. The response options are: *many, a lot, some, a few, none*. Are the differences between these options clear? Respondents may be confused by the distinction between "many" and "a lot." *Always pretest questions and check for differentiation.*

Response options should be mutually exclusive. Imagine we want to know the party identification of respondents in a political survey. The response options include: *Democratic, Republican, Green, Libertarian, Third Party, Independent, and Other*. The Green Party and the Libertarian Party *are* third parties, so these options overlap. A libertarian respondent can choose either Libertarian Party or third party and be right in either case. *Always make sure there is zero overlap between response options. Again, pretesting is a good way to check this.*

Response options should anticipate all possibilities. A national survey in 2006 asked respondents whether they felt gay marriage should be legal or illegal. Nearly 20% of respondents volunteered "civil unions" as a third response, refusing to take either the "legal" or "illegal" options the researchers presented. This question failed to anticipate a response that was the genuine position of many respondents. Some respondents may have chosen civil unions had it been offered, but selected one of the other options or DK instead. We cannot know why some chose to volunteer it whereas others did not, but that bias makes this question unusable. *Always try to anticipate the answers that significant percentages of respondents would choose, and include those as response options. It may also be worth offering an "other" response that lets respondents fill in the blank next to it with their preferred answer. Researchers should always pretest questions to ensure that they are not missing a major or obvious response option.*

Vague Quantifiers. Imagine that we want to know how often a political candidate makes a respondent angry. The response options include: *very often, fairly often, occasionally, rarely, and never*. These types of responses are called "vague quantifiers" because they indicate relative frequency without using clear numerical values. Vague quantifiers can be fine as responses as long as the options are clearly differentiated (see example above).

Primacy and Recency Effects. Respondents tend to choose either the first option on a list or the last. These biases are respectively called primacy and recency effects. They are especially problematic with nominal scales in which response options that do not fit in any particular order are listed (e.g. *What is your favorite soda? Coke, Pepsi, Sprite, Mountain Dew, Sunkist*; these sodas could be listed in any order unlike the points on a Likert scale). Researchers can reduce primacy and recency bias by randomizing the order of responses. Of course, randomization is not possible with ordered scales and is difficult in some modes. Research suggests, though, that primacy and recency effects are substantially reduced in self-administered formats where respondents can see and process all of the responses before selecting one. These biases are more severe in human-administered modes because respondents tend to remember the first and last items on a list that is read to them.

Anchoring Effects. The labels of scale endpoints tend to affect how respondents perceive the rest of the scale, something called an "anchoring effect." Endpoints labeled with more extreme language tend to turn off respondents, causing them to choose responses closer to the middle of the scale. For example, imagine that we want to know how worried a respondent is about the state of the economy and we are using a 7-point scale where lower values indicate less worry and higher values indicate more. Respondents are less likely to choose a seventh point labeled "extremely worried" than one labeled "very worried." The word "extremely" sounds more severe than "very," and will cause respondents to select an option that indicates less worry. At the low end of the scale, an endpoint labeled "not at all worried" will likely get fewer responses than one labeled "not worried." These two labels mean the same exact thing, but "not at all" makes the low end sound more extreme than it is. *Anchoring effects are almost always present as there is usually a less extreme way to label the high and low ends of a scale. Researchers should be cautious in how they label endpoints and consider pretesting alternative endpoint labels to test for the severity of anchoring bias.*

Author: Patrick R. Miller, DISM Survey Research Associate