RESPONSE FORMATS

Allyson L. Holbrook
Associate Professor of Public Administration and Psychology
& Associate Research Professor at the Survey Research Laboratory of the University of Illinois at Chicago

HOUSEKEEPING

Hold questions until the end – chat function.

Slides will be made available on the SRL site where you registered (www.srl.uic.edu).

Please raise your hand to acknowledge that you can hear me.
CLOSED-ENDED QUESTIONS WITH SCALED RESPONSE OPTIONS

Questions with a list of response options that fall along a dimension (either unipolar or bipolar).

Produce at least ordinal level variable(s) which are often treated as interval for the purposes of analysis.

Response options are presented (either visually or orally) to respondents.

CHARACTERISTICS OF RESPONSE SCALES TO CONSIDER

• Unipolar v. bipolar scales
• Number of scale points
• Labeling of scale points
• Scales and midpoints
• Use of branching
UNIPOLAR VERSUS BIPOLAR SCALES

Unipolar scales/constructs:
- One end of the scale is conceptually a zero point.
- The midpoint of the scale is a moderate level of the dimension being measured.
- Unipolar constructs include frequency, amount, importance, intensity.

Bipolar scales
- The midpoint of the scale is conceptually a “zero” or neutral point representing neither endpoint, status quo.
- The endpoints represent two opposites (e.g., strongly like and strongly dislike).
- Bipolar constructs include evaluations (like, dislike), perceived or desired change (increased, decreased)

Scales versus constructs:
- Some constructs are inherently unipolar or bipolar (e.g., bipolar scale measuring frequency)
- Some constructs can be measured using either bipolar or unipolar scales (e.g., satisfaction)

EXAMPLES OF CLOSED-ENDED QUESTIONS WITH SCALE RESPONSE OPTIONS:

Unipolar scales/constructs:
How often do you pay for your purchases with cash? Would you say never, occasionally, sometimes, usually, or always?

How important is being an American to you? Would you say not at all important, slightly important, somewhat important, very important, or extremely important?

Bipolar scales/constructs:
What is your opinion about the Affordable Care Act, sometimes called “Obamacare”? Would you say you strongly favor it, somewhat favor it, slightly favor it, neither favor nor oppose it, slightly oppose it, somewhat oppose it, or strongly oppose it?
UNIPOLAR VERSUS BIPOLAR SCALES:

<table>
<thead>
<tr>
<th></th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unipolar scales</td>
<td>• Less cognitively difficult</td>
<td>• Usually less sensitive than bipolar</td>
</tr>
<tr>
<td></td>
<td>• Less likely to force respondents to choose between two “non-opposites”</td>
<td>• Sometimes fails to take into account both ends of a spectrum</td>
</tr>
<tr>
<td>Bipolar scales</td>
<td>• Usually more sensitive than bipolar</td>
<td>• More cognitively difficult</td>
</tr>
<tr>
<td></td>
<td>• More balanced (takes both ends of the spectrum into account)</td>
<td>• Assumes that two opposing points are opposites on a single scale (e.g.,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ambivalence versus neutrality)</td>
</tr>
</tbody>
</table>

WHAT ARE THE GOALS IN CONSTRUCTING A SCALE?

- Covers the entire measurement continuum
- No overlap in the meanings of adjacent points
- Respondents assign precise and stable meaning to each point
- Shared meaning for respondents and researchers
- Enough points to differentiate respondents as much as possible.
- Few enough points that the meaning of each point does not become ambiguous.
COMMON QUESTIONS WHEN CONSTRUCTING RESPONSE SCALES:

- How many scale points should I include?
  - Unipolar v. bipolar

- How should I label scale points?
  - Verbal v. numbers v. pictures
  - Endpoints v. endpoints and midpoint v. fully labeled

- Should I include a midpoint?

NUMBER OF SCALE POINTS:

Survey questions use a wide range of number of scale points, varying from:

Two response options:
- Do you favor or oppose eliminating collective bargaining rights that are used by unions to negotiate salaries and benefits for all state employees, except those in health and safety fields, such as law enforcement officers and nurses?

To 101 - response options:
**NUMBER OF SCALE POINTS:**

Goals:

- Sensitivity
- Interpretability

Many people have conducted experiments to test the impact of number of scale points on the validity/reliability of data.

- E.g., Masters (1974)
- See Krosnick and Presser (2010) for a review
NUMBER OF SCALE POINTS:

Current evidence about best practices for maximizing data quality:

- Unipolar scales: 5 scale points:
  - During the last week, how often have you felt energetic? Would you say never, occasionally, sometimes, often, or always?

- Bipolar scales: 7 scale points:
  - Thinking about your current nursing job, would you say you like it very much, like it somewhat, like it a little, neither like nor dislike it, dislike it a little, dislike it somewhat, or dislike it very much?

- May depend upon the goals of the research:
  - Scales evaluated in terms of validity and reliability
  - A smaller number of points may be sufficient if the goal is to report simple frequencies or percentages (e.g., results reported in media surveys)

LABELING OF SCALE POINTS:

Choices:

- Endpoints, endpoints and midpoint, or all scale points
- Numbers versus verbal labels (or some combination)
- How to choose verbal labels.
- Using pictures as labels.
Examples: Thinking about your current nursing job, how would you evaluate it:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Like</td>
<td>Dislike</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very Much</td>
<td>Very Much</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Like</td>
<td>Neither Like</td>
<td>Dislike</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very Much</td>
<td>Nor Dislike</td>
<td>Very Much</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Like</td>
<td>Like</td>
<td>Like A</td>
<td>Neither Like</td>
<td>Dislike</td>
<td>Dislike</td>
<td>Dislike</td>
</tr>
<tr>
<td>Very Much</td>
<td>Somewhat</td>
<td>Little</td>
<td>Nor Dislike</td>
<td>A Little</td>
<td>Somewhat</td>
<td>Very Much</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Like</td>
<td>Like</td>
<td>Like A</td>
<td>Neither Like</td>
<td>Dislike</td>
<td>Dislike</td>
<td>Dislike</td>
</tr>
<tr>
<td>Very Much</td>
<td>Somewhat</td>
<td>Little</td>
<td>Nor Dislike</td>
<td>A Little</td>
<td>Somewhat</td>
<td>Very Much</td>
</tr>
</tbody>
</table>
GOALS OF SCALE LABELS:

What are the goal of labels?
- Clearly understood
- Understood consistently across individuals
- Evenly spaced (easier to communicate with numbers)
- Numerical labels or unlabeled scale points are open to interpretation (that we may not want)
- Numbers assigned to scale points may unintentionally influence responses
- Unlabeled scale points are even more ambiguous and potentially open to interpretation
- Interpreting unlabeled scale points takes additional cognitive effort

WHY THE SPECIFIC CHOICE OF NUMBERS IN SURVEY RESPONSES CAN INFLUENCE RESPONSES:

Conversational norms: (Gricean norms)
- Assumptions of conversation that allow speakers to communicate more than just the content of what is said
- Example: Maxim of relation – a speaker tries to be relevant and say things that are pertinent to the conversational topic
SURVEY RESPONDENTS INFER MEANING FROM ASPECTS OF THE SURVEY IN UNINTENDED WAYS (SCHWARZ ET AL., 1991):

How successful have you been in life so far?
- Half of respondents were randomly assigned to rate from 0 to 10 with endpoints labeled “not at all successful” and “extremely successful”
- Half of respondents were randomly assigned to rate from -5 to +5 with endpoints labeled “not at all successful” and “extremely successful”
- Only 9% of respondents rated their success either a “9” or a “10” on the first scale, but 28% rated their success either a “+4” or a “+5” on the second.

Evidence suggests that using fully labeled scales (without numbers) is the best approach
- Krosnick and Presser (2010); Menold et al. (2014)

Examples:

Evaluation:
- Like a great deal
- Like somewhat
- Like a little bit
- Neither like nor dislike
- Dislike a little bit
- Dislike somewhat
- Dislike a great deal

Frequency:
- Never
- Occasionally/Rarely
- Sometimes
- Often
- Always

Extent or amount:
- None at all
- Less than half
- About half
- More than half
- All

Qualifiers:
- Not at all
- Slightly
- Somewhat
- Very
- Extremely
HOW TO SELECT SPECIFIC LABELS:

Use existing “standard” scales for dimensions like evaluations, frequency, amount, etc.

Work by Saris (1988) suggests that the use of more “fixed reference points” increases consistency of interpretation across respondents.

- Verbally labeled points
- Specific labels (not just “good” or “bad” but “very good” and “very bad”) are better
- Suggests that it is not a good idea to use “unqualified” response options:
  - Not at all important, slightly important, important, very important, extremely important
  - Versus
  - Not at all important, slightly important, somewhat important, very important, extremely important

CONCERNS ABOUT INTERPRETATION OF SCALE LABELS

Concern that response options are not evenly spaced

Concern that response options are not equivalently interpreted across groups.

Ask respondents (or multiple subsets of respondents) to rate response options in a pretest

- Assign values from 0 to 100 to each response option.
- E.g. Johnson et al. (1996) – interpretation of health rating scale (Excellent, very good, good, fair, or poor) across racial and ethnic groups

Give values to responses based on results of pretest either to adjust for uneven spacing or different interpretation across groups.
USE OF PICTURES TO LABEL SCALES: PAIN (UNIPOLAR)

EVALUATION (BIPOLAR EXAMPLE OF A VISUAL ANALOG SCALE)

How would you describe your education at Berry College?
USING FACES OR GRAPHICS:

Most appropriate when:
- Surveying populations with limited literacy or cognitive skills (e.g., children, elderly)
- See e.g., Herr et al. (1998)

Things to consider:
- Comparable interpretation of stimuli
- Are faces/expressions universal? (see Jack, et al., 2012)

RESPONSE SCALE: SHOULD I INCLUDE A MIDPOINT?

Why might people select the midpoint?
- To avoid giving an undesirable response.
- To avoid doing the cognitive work necessary to answer the question carefully and completely.
- In lieu of a “don’t know” response option
- Their true response is at the midpoint.

Evidence that:
- People do not interpret a “neither” midpoint as equivalent to a “don’t know” option (Dillman, 2000)
- Being neutral is different from not knowing one’s opinion
- Selecting midpoint seems to be the result of a different process than selecting an explicitly offered don’t know or no opinion option.
Little consistent empirical evidence of value of excluding midpoint.

- Midpoint scale increases reliability (Krosnick and Tahk, unpublished)
- Midpoint selection unrelated to cognitive abilities (Narayan and Krosnick, 1996)
- Somewhat less midpoint responding when a “don’t know” response is offered (Schuman and Presser, 1981)
- Evidence that adding midpoints increases reliability and validity (O’Muircheartaigh et al., 2000; although see Saris and Gallhofer, 2007)
- Evidence may be mixed because different subgroups of respondents are attracted to the midpoint for different reasons

General recommendation: Include a midpoint unless there are convincing reasons not to.

Exception to this interpretation (see de Bruin et al, 2002):

- Open-ended probability judgments
- People seem to select 50% as a proxy for “don’t know”
- Phrase “fifty-fifty” denotes uncertainty
- In general, probability judgments that ask for a specific value are problematic:
  - Level of precision
  - Lack of understanding of probability.
  - Avoid altogether
MIDPOINT LABELING:

Unipolar – moderate position

Bipolar – midpoint should have a label like “neither like nor dislike” (e.g., continuum from like strongly to dislike strongly) or “no change” (e.g., continuum from increase a great deal to decrease a great deal) or “equal to” (e.g., continuum from much greater than to much less than).

Midpoint should *not* be labeled “don’t know”.

Midpoint labels clearer if construct specific rather than “neutral.” (e.g., neither like nor dislike)

USE OF BRANCHING IN BIPOLAR SCALES

Bipolar scales:
- Thinking about your current nursing job, would you say you like it very much, like it somewhat, like it a little, neither like nor dislike it, dislike it a little, dislike it somewhat, or dislike it very much?

Concern in modes where response options are orally presented:
- A lot of response options to remember
- One option is to use a showcard (if possible)
- Another option is to use a series of branching items
BRANCHING AS AN ALTERNATIVE: OPTION #1

Thinking about your current nursing job, would you say you like it, dislike it, or neither like nor dislike it?

IF LIKE: Would you say you like it very much, somewhat, or a little?

IF DISLIKE: Would you say you dislike it very much, somewhat, or a little?

Construct response with 7 options from these three variables.

---

BRANCHING AS AN ALTERNATIVE: OPTION #2

Thinking about your current nursing job, would you say you like it, dislike it, or neither like nor dislike it?

IF LIKE: Would you say you like it very much, or somewhat?

IF DISLIKE: Would you say you dislike it very much or somewhat?

IF NEITHER LIKE NOR DISLIKE: If you had to choose, would you say you like it a little or dislike it a little?
BRANCHING VERSUS 7-POINT SCALE PRESENTATION:

Krosnick and Berent (1993):
- Branching results in more reliable and valid responses
- A series of branching doesn’t take longer than a single 7-point bipolar scale.

Krosnick, Malhotra, and Thomas (2009)
- First branching approach (where like and dislike are followed up by a three-point response and there is no follow-up question for those who select the midpoint) is optimal

RESPONSE ORDER EFFECTS

Definition: When the order in which response options are offered influences the distribution of responses.

Two kinds:
1) Primacy — when response options are selected more frequently when presented near the beginning of a list or scale than when they are presented near the end of the list or scale

Recency — when response options are selected more frequently when presented near the end of a list or scale than when they are presented near the beginning of the list or scale

Measuring response order effects: Typically assessed using a between subjects experiment in which respondents are randomly assigned to receive response options in one of two or more possible orders.
WHAT LEADS TO RESPONSE ORDER EFFECTS?

1. Survey satisficing (Krosnick, 1991)
   * When respondents lack the ability or motivation to carefully complete the cognitive steps necessary to answer survey questions, they may look for a strategy to satisfice, or give a satisfactory answer without fully going through all the cognitive steps.
   * Choosing the first satisfactory response they consider is one strategy respondents use to satisfice.

2. Contrast effects (Schwarz and Hippler, 1991; Schwarz, Hippler, and Noelle-Neumann, 1992)
   * Response options may influence the interpretation of later response options.

HOW TO ASSESS RESPONSE ORDER EFFECTS IN RESPONSE SCALES:

Half of respondents: How satisfied are you with the ability to advance within the nursing profession? Would you say not at all satisfied, slightly satisfied, somewhat satisfied, very satisfied, or extremely satisfied?

Half of respondents: How satisfied are you with the ability to advance within the nursing profession? Would you say extremely satisfied, very satisfied, somewhat satisfied, slightly satisfied, or not at all satisfied?
RESPONSE ORDER EFFECTS IN QUESTIONS WITH RESPONSE SCALES

Predominantly primacy effects

Don't need to hear all the response options to form a judgment.
Also may be affected by anchoring and adjustment.

See example: Chan, 1991; Yan & Keusch, 2015

RESPONSE OPTION ORDER RECOMMENDATIONS:

Be aware of possible response option order effects

Conduct experiments to rotate response option order across respondents:
  • Allows researchers to estimate response order effects
  • Allows researchers to control for response order effects

Are there instances where you do not want to rotate response options?
How should you select an order if you do not want to or cannot rotate response order?
FINAL POINTS

Webinar evaluation will be sent out.

Upcoming Spring, 2016 Webinars:
- Fundamentals of Survey Data Set Construction
- Wednesday, April 13, noon CST

THANK YOU!

QUESTIONS?

ALLYSON@UIC.EDU