

Survey Research

Survey Research

- The process of collecting data by asking people questions and recording their answers
- Two purposes of surveys
 1. Estimation of population parameters
 - The percentage of people who hold supporting and opposing positions on social issues
 2. Hypothesis testing
 - Testing specific research questions, such as by correlating survey responses to participants' characteristics

- Surveys can be classified into two broad categories, based on recruitment procedures
 1. Sample surveys: Selected to produce a respondent sample with characteristics similar to the population of interest
 - Examples: National opinion polls, General Social Survey, World Values Survey
 2. Convenience surveys: Members come from a group of people who are convenient to the researcher
 - Used to estimate population parameters or to explore relationships between variables

Reasons for Developing Scale

- Update established scale
- Revise existing scale based on more refined psychometric properties (CFA)
 - For categorical data: optimal scaling, PCA/FA, Item response theory (IRT)
- Shorten
- Contextualize to specific subgroups
- Make it more relevant to construct
- New scale to fill specific research purpose

Specifying the Construct

- Multi-faceted
 - Components can be measured separately and distinguishable but related
 - E.g., 3 Factor model of Narcissism
- Multi-dimensional
 - Components not related to each other
 - E.g., Depression
 - Sleep, mood, activity level
 - Combine facets when interested in latent variable (e.g., Dark Triad)
 - Better predictor of DV
 - More appropriate level of abstraction

Scale Development

- Narrow traits can be measured with ~ 8 items
- Broader traits require more items (12–14) to represent different aspects of content domain
- Consider different situations in which trait may be expressed
- Keep length to 30 min or less
 - < 200 items

Scale Development

Sample

- ▶ Give to group of interest (avoid double dipping)
 - Separate independent samples for reliability and validity
- ▶ Give to 5X the number of people as there are items in your scale



Reliability

- ▶ Crucial when designed to assess individual differences **constructs**, to characterize individuals, or evaluate hypotheses about construct relations (e.g., SE and LoC)
- ▶ Less crucial when used to represent average or composite score for group (e.g., change in alienation in high school students over time)
- ▶ When scale is online measure of ability rather than trait



Scale Development

Reliability

- ▶ No single, universal coefficient for any measure
- ▶ If not reliable – not valid
- ▶ Reliability is a minimum, but not sufficient criterion for validity
 - Why?
- ▶ As reliability decreases, validity usually decreases



Reliability cont.

- ▶ Internal consistency
 - Represented by coefficient alpha (Cronbach's)
 - α should be at least .75 (.80 to begin)
 - Item analysis
 - Corrected item-total correlation
 - Correlation between that item and sum of all other items combined
 - Look for .40 or higher
 - Alpha if item deleted
 - What α would be if item on that row deleted



Validity

- ▶ Content
- ▶ Criterion-related
 - Predictive
 - Concurrent
- ▶ Construct
 - Convergent
 - Discriminant
- ▶ Differential



Types of Questions

- ▶ Open-ended: Allow respondents to say anything they want in their own words
 - Example: Describe how you see yourself as a student
- ▶ Closed-ended: Require respondents to select from choices provided by researcher
 - Usually preferred because responses are easier to quantify
- ▶ Advantages and Disadvantages?



Example Topic for Discussion

- ▶ New Theory of Mind measure
- ▶ Other topics for discussion?

Types of Questions

Open-ended questions are more useful when

- research is preliminary
 - If so, researcher might be unsure about which response options are most appropriate
 - Act Nomination
- assessing frequency of sensitive or socially disapproved behaviors

Constructing a Good Questionnaire

Principles of Good Question Writing

- ▶ **Avoid**
 - Jargon/slang/abbreviations
 - ASP – Association Surfing Professionals, Astronomical Society of the Pacific, Appalachia Service Project
- ▶ People will usually answer questions, even if they don't understand them
 - Complex words – use Flesch–Kincaid grade level
- ▶ **Word**
 - Click the **File** tab, and then click **Options**.
 - Click **Proofing**.
 - Under **When correcting spelling and grammar in Word**, make sure the **Show readability statistics** box is checked.
 - Under review tab, click Spelling & Grammar
 - Once complete, readability tab should appear
 - Aim for grade level of 7 or 8
 - Readability should be between 60 and 70

Constructing a Good Questionnaire

Principles of Good Question Writing

- ▶ **Avoid**
 - Ambiguity/confusion/vague

Where do you like to shop?
 - Emotional language
 - Prestige bias

How do you feel about the President's unconstitutional attempt to grant amnesty to illegal immigrants?
 - Double-barreled questions

Domestic violence and AIDS are the most serious problems plaguing America today

Do you think that the U.S. deficit is too big and that spending should be cut?

Constructing a Good Questionnaire

Principles of Good Question Writing

- ▶ **Avoid**
 - Overlapping/unbalanced response categories
 - All positively worded items
 - Leading questions
 - "Wouldn't you be happy if taxes were reduced"
 - Questions with low variance
 - "I like to talk to some people more than others"
 - False premises
 - Beyond respondents' knowledge
 - Distant future intentions
 - Threatening questions
 - response bias

Writing Survey Questions

Avoid

Confusing phrasing

- Use shorter words and questions
- Avoid double negatives
- Minimize risk of bias
- Frequency of scale

Does it seem possible or impossible to you that the Nazi extermination of the Jews never happened?

Do you agree that people who do not have proper identification should not be allowed to vote?

Bad ?s

- Age?
- Rank these 25 topics in order of importance
- How many times in the past year did you eat out at a restaurant?
- When you travel overseas, do you stay in hotel chains or privately owned facilities, such as bed and breakfasts?
- Working in medicine can be stressful and demand many long hours. How would you describe your ability to cope with these work conditions?
- Is therapy helpful?
- Do you disagree that the government should not increase funding for education?
- It's ok to feed dolphins in the wild
 - Yes No Unsure
- In captivity, it is better to train dolphins than to leave them alone
 - Yes No Unsure

Good Question Wording

- Be specific
 - People's belief systems are complex
 - Beliefs about general categories can be different from beliefs about specific categories
 - People probably respond differently to these items:
 - "I like young children"
 - "I would like to teach young children"
 - Or
 - "Faith is a good thing"
 - "Believing in a Christian God benefits society"

Good Question Wording

- Avoid making assumptions
 - People will often respond to questions even if they are uninformed about the topic
 - Example: "The federal government spends too much money on aid to foreign nations"
 - Assumes the respondent knows how much is spent
 - Be aware that people's backgrounds, interests, and experiences differ
 - Not everyone lives with two parents, is heterosexual, is employed, wants to have children, etc.

Good Question Wording

- Address sensitive topic sensitively

What are your feelings about abortion?

Strongly in favor ...Somewhat in favor...Neutral....Somewhat opposed...Strongly opposed

 - Failing to do so can lead people to
 - underreport undesirable behaviors
 - over-report desirable behaviors

Constructing a Good Questionnaire

Getting Honest Answers

EXTENSION 10.5

Example of a Contingency Question

QUESTION VERSION 1 (NOT CONTINGENCY QUESTION)
In the past year, how often have you used a seat belt when you have ridden in the backseat of a car?

QUESTION VERSION 2 (CONTINGENCY QUESTION)
In the past, have you ridden in the backseat of a car?
No [Skip to next question]

Yes → When you rode in the backseat, how often did you use a seat belt?

| Results | Always Use | Never Use |
|-----------|------------|-----------|
| Version 1 | 30% | 24% |
| Version 2 | 42 | 4 |

During pilot testing, researchers learned that many respondents who answered "never" to Version 1 did not ride in the backseat of a car. Version 1 created ambiguity because respondents who never rode in the backseat plus those who rode there but did not use a seat belt both answered "Never." Version 2 using a contingency question format clarified the question.

Source: Adapted from Presser, *Reassessing Survey Questionnaires*, Hoboken, NJ: Wiley (2004). Reprinted by permission of John Wiley & Sons, Inc.

Contingency questions

- Partially open question

Sleeper Question

- **Example:** Take a look at the images of prime ministers below and rate your knowledge for each as a, b, or c.
 - a. I recognize this prime minister and know his or her name
 - b. I recognize the prime minister but can't remember his or her name
 - c. I don't recognize this prime minister



Levels of Measurement

- ▶ Measurement consists of applying sets of rules to assign numbers to variables
- ▶ Variables can be measured at four levels of information content

Nominal Level Measurement

- ▶ Sorts people into categories based on common characteristics
 - Characteristics represent different aspects of a variable
 - Usually reported as
 - frequency counts
 - percentage of cases in each category
 - Arbitrary numbers are usually assigned to categories for data analysis
 - My ethnicity is (check one)
 - 1) African American
 - 2) Caucasian (non-Latino)
 - 3) Latino
 - 4) Asian
 - 5) Native American

Ordinal Level Measurement

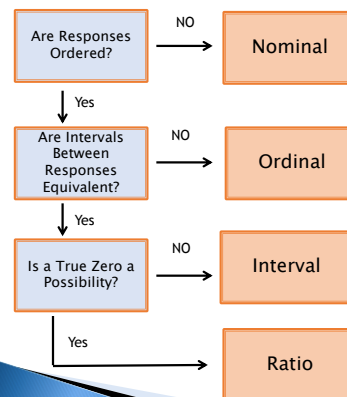
- ▶ Places people or things in a series based on increases or decreases in the magnitude of the variable
- ▶ Provides comparative information
 - Comparison is not quantified
 - Data cannot be mathematically transformed
 - Inferential statistics are not appropriate for this level of measurement
- ▶ The service on this flight was (check one)
 - 1) Better than the last flight I took
 - 2) About the same as the last flight I took
 - 3) Worse than the last flight I took

Interval Level Measurement

- ▶ Assigns numbers to people and things
- ▶ The difference between one set of adjacent scores is assumed to be the same as the difference between two other adjacent scores
 - Respondents assume intervals to be equal
- ▶ can be added to or subtracted from each other
- ▶ can be mathematically transformed using constants
- ▶ lack a true zero point, so cannot be multiplied or divided by each other
 - Researchers must interpret scores cautiously
 - Example: A person with an IQ score of 160 is *not* twice as smart as a person with an IQ score of 80
- ▶ E.g., Likert

Ratio Level Measurement

- ▶ Like interval level measurement, assigns equal interval scores to people or things
- ▶ Ratio-level measures have true zero points
 - So can be multiplied and divided by each other
 - Person who recalls 20 items on a memory test does twice as well as the person who recalls ten items
- ▶ Relatively uncommon in psychological research



Identify the Scale of Measurement

- ▶ Hair color
- ▶ Number of times a student uses blackboard
- ▶ Military rank
- ▶ Ratings of liking for yogurt on a 7-point scale
- ▶ Number of Girl Scout cookies sold by the average Girl Scout Troop
- ▶ Ranked scores on Homework
- ▶ Summed scores on an introversion measure
- ▶ Time taken to solve a Sudoku puzzle

Why is Scale of Measurement Important?

- ▶ Higher level measures contain more information
- ▶ Cannot move from lower level to higher level scale of measurement
 - If only know if people are proficient or not, cannot estimate how long they took to solve puzzle
- ▶ Should collect data using highest level of measurement possible
 - Possible conclusions are restricted by level of measurement
 - Higher level measures tend to be more sensitive to effects of IVs

Hypothetical Responses to Privatizing Social Security

| Nominal Level Data | | |
|--------------------|---------|--------------|
| | Liberal | Conservative |
| Favor | 50% | 50% |
| Oppose | 50% | 50% |

Conclusion: No differences in opinion by political group membership

Hypothetical Responses to Privatizing Social Security

| Interval Level Data | | |
|---------------------|----------|---------------|
| | Liberals | Conservatives |
| Strongly Favor | 2 | 16 |
| Moderately Favor | 2 | 2 |
| Slightly Favor | 16 | 2 |
| Slightly Oppose | 2 | 16 |
| Moderately Oppose | 2 | 2 |
| Strongly Oppose | 16 | 2 |

Conclusion: Liberals are more likely to be opposed to privatization than are conservatives

Why is Scale of Measurement Important?

- ▶ Different assumptions underlie different statistical tests
 - Example: *t*-test and ANOVA assume scores between comparison groups are normally distributed
 - Nominal data violate this assumption
 - Ordinal data may or may not violate this assumption

Response Formats

- ▶ Comparative scaling: Respondents compare a set of stimuli on some characteristic
 - Ten candidates for school board are compared on knowledge of issues
- ▶ Resulting data are at ordinal level of measurement

Comparative Rating Scales

- ▶ Two comparison methods:
 - Paired comparison: All possible pairings are presented and participants select the most knowledgeable candidate from each pair
 - Can become unwieldy with large number of pairs to rate
 - Use for sets of around ten stimuli (around 45 pairs) or fewer



Comparative Rating Scales

- Method of rank order: Respondents rank list of stimuli from highest to lowest
- Can also have participants choose the most and least preferred of ten candidates
 - Then select the most and least preferred of the remaining eight
 - Continue with remaining six and so forth until all pairs are rated



Comparative Rating Scales

- ▶ To provide reliable and valid results, three conditions must be met
 1. Respondents must be familiar with all the stimuli they compare
 2. The characteristics being rated must be unidimensional
 - Otherwise, you have no way of knowing what dimension is being rated
 3. Respondents must completely understand the meaning of the characteristic being rated



Itemized Rating Scales

- ▶ Multiple choice questions
- ▶ Item stem is in the form of a question
- ▶ Respondents choose an answer from the options the researcher provides
- ▶ Usually provide data at nominal level of measurement



Itemized Rating Scale

- ▶ Example: What is your level of education?
 - ___ Some high school
 - ___ High school diploma or GED
 - ___ Some college
 - ___ Bachelor's degree
 - ___ Some graduate work
 - ___ Master's or professional degree
 - ___ PhD



Itemized Rating Scale

- ▶ Can also be used to assess hypothetical constructs
 - Respondents read about a hypothetical situation where another needs help. They choose the response most similar to the one they would make in this situation



Itemized Rating Scale

- ▶ Must be careful that all relevant options are included
 - If none of the response options apply to them, respondents might
 - skip the item
 - choose an inaccurate response
 - Can lead to biased research results
 - In the education example, people with MDs, law degrees, or EdDs may not know which response to choose

Graphic Rating Scales

Responses are indicated pictorially rather than by choosing a category or providing a numerical response

Example: Indicate your response by marking an X on the part of the line that best represents your opinion

What is your opinion about the importance of voting in the election for school board members?

It is _____ It is
essential _____ unnecessary

- Scales are scored by measuring the distance of the respondent's mark from one end of the line
- Graphic scales, such as the Smiley Faces Scale, can be used with children



Numerical Rating Scales

- ▶ Respondents assign numerical values to their responses
- ▶ Meanings of values are defined by verbal labels called *anchors*
- ▶ Anchors are placed at the ends of the measure
- ▶ Can also use intermediate anchors
 - Advantages include
 - clarifying meaning of scale for respondents
 - increasing the level of measurement for a scale
 - Helps ensure that interval sizes are perceived as psychologically equal

- ▶ Number of scale points is determined by
 - Required sensitivity of measurement
 - Scales with more rating points are more sensitive
 - Allow fine gradations of judgment
 - More sensitive measures can detect large differences
 - Usability of the scale
 - Having too many scale points can be overwhelming for respondents
 - In general, people prefer scales with 5 to 9 points

How satisfied are you with our service?



Which Response Format is Best?

- ▶ If purpose is
 - determining relative position (ranking), use comparative format
 - locating an individual or stimulus on a dimension, use numerical rating scale with equal intervals
 - placing people or stimuli into specific categories, use itemized scale

Multi-Item Scales

- ▶ Composed of two or more items in a rating scale format
 - Each is designed to assess the same variable
 - Respondent's scores on items are combined to form an overall score
 - Can be composed of two or more subscales

Advantages of Multi-Item Scales

- ▶ Can assess multiple aspects of a construct
 - Each subscale measures a different component of the construct
 - Subscales can be analyzed separately or combined into an overall score
- ▶ Compared to single-item scales, multi-item scales
 - have greater reliability and validity
 - provide greater sensitivity of measurement

Types of Multi-Item Scales

- ▶ Likert scales: Sets of statements about a person, thing, or concept
 - Respondents' level of agreement or disagreement is assessed
 - Same numerical scale is used for all items
 - Response bias is controlled by having equal numbers of positively and negatively worded statements
 - Negatively worded items are reverse scored so that higher numbers always have the same meaning, such as more of a trait

Likert Scale

- ▶ Term is frequently misused to refer to any numerical rating scale
- ▶ Underlying assumption is that scale is unidimensional
 - Every item on scale or subscale measures just one construct
 - If assumption doesn't hold, cannot easily determine meaning of overall score
- ▶ Assumption can be tested with factor analysis
- ▶ Can assess multidimensional constructs through use of subscales

Likert Scale

INDEX OF EQUAL OPPORTUNITY FOR WOMEN

Questions

- Women have less opportunity than men to get the education they need to be hired in top jobs.
- Many qualified women cannot get good jobs; men with the same skills have less trouble.
- Our society discriminates against women.

| Strongly Agree | Somewhat Agree | Somewhat Disagree | Disagree a Great Deal | Don't Know |
|----------------|----------------|-------------------|-----------------------|------------|
| 1 | 2 | 3 | 4 | 5 |

Scoring: For all items, Strongly Agree = 1, Somewhat Agree = 2, Somewhat Disagree = 4, Disagree a Great Deal = 5, Don't Know = 3.

Highest Possible Index Score = 15, respondent feels opportunities for women are equal

Lowest Possible Index Score = 3, respondent feels opportunities are not equal

SELF-ESTEEM INDEX

Questions

- On the whole, I am satisfied with myself.
- At times, I think I am no good at all.
- I sometimes feel that I'll tell men do not take my opinion seriously.

| Strongly Agree | Somewhat Agree | Somewhat Disagree | Disagree a Great Deal | Don't Know |
|----------------|----------------|-------------------|-----------------------|------------|
| 1 | 2 | 3 | 4 | 5 |

Scoring: Items 1 and 3: 1 = Disagree, 2 = Don't Know, 3 = Agree, Item 2: 1 = Disagree, 2 = Don't Know, 3 = Agree.

Highest Possible Index Score = 9, high self-esteem

Lowest Possible Index Score = 3, low self-esteem

Steps for Creating a True Likert Scale

- ▶ Write a large number of items to represent the variables being measured
 - Items should represent the extremes of the variables
 - Select a numerical response format with intermediate anchors
- ▶ Administer the items to a large number of respondents
 - Use at least 100 respondents or 10 per item, whichever is greater

Steps for Creating a True Likert Scale

- ▶ Conduct an item analysis
 - This indicates which items best discriminate between high and low scorers
 - Examine item-total correlations
 - Throw out items that poorly discriminate
- ▶ Final scale is comprised of items with highest item-total correlations
 - Internal consistency coefficient should be 0.75 or greater

Advantages of Likert Scales

- Relatively easy to construct
- Tend to have high reliability
- Are highly flexible
 - Can assess attitudes, personality, person perception, etc.



Thurstone Scales

- Based on items representing the entire range of attitudes
 - Highly positive
 - Neutral
 - Highly negative

Example Statements

1. I think that the death penalty is cruel and unnecessary punishment.
2. Without the death penalty, there would be many more violent crimes.
3. I believe that the death penalty should be used only for a few extremely violent crimes.
4. I do not think that anyone was ever prevented from committing a murder because of fear of the death penalty.
5. I do not think that people should be exempt from the death penalty if they committed a murder even if they are insane.
6. I believe that the Bible justifies the use of the death penalty.
7. The death penalty itself is not the problem for me, but I believe that electrocuting people is a cruel way to put them to death.



Thurstone Scales

- Judges independently sort items into 11 categories
 - Represents the degree of favorability the items express
 - Interval-level sorting procedure used
 - Difference in favorability between any two adjacent categories should equal that of any other two adjacent categories



Thurstone Scales

- Criteria for items in final scale
 - Must represent the entire range of attitudes
 - Must have low variance in their judged favorability
- Respondents are presented with a list of items in random order
 - Without seeing favorability ratings, they check those they agree with
 - Score is the mean of the favorability ratings of checked items



Thurstone Scales

- Are rarely used because:
 - It is assumed that the attitude being measured is unidimensional
 - Subscales cannot be assessed
 - Compared to Likert scales,
 - are laborious to construct
 - tend to produce less reliable scores
 - Judges' attitudes can influence the values assigned to items
 - Values assigned to items can change over time
 - Limits reusability of scale



Guttman Scales

- Respondents are presented with a set of ordered attitude items
- Scale is designed so that a respondent will agree with all items up to a point
 - Will disagree with all items after that point
 - Total score determines response to all items
- Rarely used because few variables lend themselves to such ordering
- Mokken scale = non-parametric, probabilistic version



Guttman Scaling

| | FORM OF PROTEST | | | | |
|---------------------------------------|-----------------|----------------|----------|--------|--------|
| | Petitions | Demonstrations | Boycotts | Strike | Sit-In |
| <i>Guttman Patterns</i> | | | | | |
| | N | N | N | N | N |
| | Y | N | N | N | N |
| | Y | Y | N | N | N |
| | Y | Y | Y | N | N |
| | Y | Y | Y | Y | N |
| | Y | Y | Y | Y | Y |
| <i>Other Patterns (examples only)</i> | | | | | |
| | N | Y | N | Y | N |
| | Y | N | Y | Y | N |
| | Y | N | Y | N | N |
| | N | Y | Y | N | N |
| | Y | N | N | Y | Y |

Bogardus Social Distance Scale

Please give me your first reaction, yes or no, whether you personally would feel comfortable having an exchange student from (name of country):

- _____ As a visitor to your college for a week
- _____ As a full-time student enrolled at your college
- _____ Taking several of the same classes you are taking
- _____ Sitting next to you in class and studying with you for exams
- _____ Living a few doors down the hall on the same floor in your dormitory
- _____ As a same-sex roommate sharing your dorm room
- _____ As someone of the opposite sex who has asked you to go out on a date

Semantic Differential

Respondents Rate a Concept on Sets of Bipolar Adjective Pairs

- Format is a set of segmented graphic rating scales
- Items are balanced so that positive/negative adjectives appear equally on right/left side
- Each item is scored on a 7-point scale
 - Usually ranges from +3 to -3
- Respondents' score are sums of their item scores
 - Items are reverse scored so that all responses represent positive response

Semantic Differential

Please read each pair of adjectives below, then place a mark on the blank space that comes closest to your first impression feeling. There are no right or wrong answers.

How do you feel about the idea of divorce?

| | | | | | | | |
|--------|-------|-------|-------|-------|-------|-------|-------------|
| Bad | _____ | _____ | _____ | _____ | _____ | _____ | Good |
| Deep | _____ | _____ | _____ | _____ | _____ | _____ | Shallow |
| Weak | _____ | _____ | _____ | _____ | _____ | _____ | Strong |
| Fair | _____ | _____ | _____ | _____ | _____ | _____ | Unfair |
| Quiet | _____ | _____ | _____ | _____ | _____ | _____ | Loud |
| Modern | _____ | _____ | _____ | _____ | _____ | _____ | Traditional |
| Simple | _____ | _____ | _____ | _____ | _____ | _____ | Complex |
| Fast | _____ | _____ | _____ | _____ | _____ | _____ | Slow |
| Dirty | _____ | _____ | _____ | _____ | _____ | _____ | Clean |

How do you feel about the idea of marriage?

| | | | | | | | |
|--------|-------|-------|-------|-------|-------|-------|-------------|
| Bad | _____ | _____ | _____ | _____ | _____ | _____ | Good |
| Deep | _____ | _____ | _____ | _____ | _____ | _____ | Shallow |
| Weak | _____ | _____ | _____ | _____ | _____ | _____ | Strong |
| Fair | _____ | _____ | _____ | _____ | _____ | _____ | Unfair |
| Quiet | _____ | _____ | _____ | _____ | _____ | _____ | Loud |
| Modern | _____ | _____ | _____ | _____ | _____ | _____ | Traditional |
| Simple | _____ | _____ | _____ | _____ | _____ | _____ | Complex |
| Fast | _____ | _____ | _____ | _____ | _____ | _____ | Slow |
| Dirty | _____ | _____ | _____ | _____ | _____ | _____ | Clean |

Semantic Differential

- ▶ Is a ready-made scale
 - Can be used to measure attitudes toward almost anything
 - No need to select scale items
- ▶ Validity is well-documented
- ▶ Be careful that adjective pairs are relevant to attitude object
- ▶ Is widely used due to its flexibility
- ▶ Developed by Osgood and colleagues
 - Found that any concept could be described in terms of
 - evaluation (goodness/badness)
 - activity (active/passive)
 - potency (strength/weakness)

Semantic Differential

Here are a number of statements that could be used to describe WalMart. For each statement tick (/) the box that best describes your feelings about WalMart.

| | | | | | | | |
|-------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|---------------------------|
| Modern Store | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Old-fashioned store |
| Low prices | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | High prices |
| Unfriendly staff | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Friendly staff |
| Narrow product range | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Wide product range |
| Sophisticated customers | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Unsophisticated customers |

Semantic Differential

| Privatizing Social Security | | |
|-----------------------------|-------|----------|
| Good | _____ | Bad |
| Unfair | _____ | Fair |
| Wise | _____ | Foolish |
| Unpleasant | _____ | Pleasant |
| Valuable | _____ | Useless |

Response Bias

- Exists when a person responds to an item for reasons unrelated to the item itself
- Is a form of measurement error
 - Obscures a person's true score on measure

Question-Related Biases

Scale ambiguity

- Occurs when numerical or graphic rating scales are used to assess frequency or amount
- Terms such as "frequently" or "a little" have different meanings for different people
 - People evaluate terms using personal standards
- Possible solution is using an ordered itemized rating scale

Question-Related Biases

Category anchoring

- Occurs when people base their responses on the range of scale values
 - Low-range scales start at low values
 - High-range scales start at high values
 - Both increase until a "greater than" value is reached
 - People tend to give lower responses to low-range scales and higher responses to high-range scales

Example:

- "This happened to me just today when I was purchasing a new pair of shoes from a running specialty store. The clerk asked me a series of questions that she was inputting into the computer system to help me select the right running shoe for my needs. She asked how many miles I run a week. As I was thinking, I glanced at the computer screen and saw the choices were: a. 1–10 miles b. 11–20, or c. 21 or more miles. I usually run 9 miles a week at the most, but having run cross-country all through high school, and considering myself 'a runner', I somehow didn't feel good about selecting what I considered to be the "beginner" response. I thought to myself 'Well, I know I CAN run more than 9 miles if I want to, it's just that I'm also spending time cycling and swimming.' So I told her to select 11 or above. Why? I knew she wasn't judging me, I knew no one else would even see my responses, and I knew the questions were only meant to help me select a sneaker I would be happy with; but somehow I just didn't want to tell her the truth'.

Question-Related Biases

Category anchoring

- People also might believe endorsing some response options would make them look bad (e.g., "more than average")
- People might use scale range to interpret item stem
 - Assume low-range scales refer to rare events
 - Assume high-range scales refer to common events

Respondent Interpretation of Numeric Scales

- ▶ People may also use numerical value of scale to interpret item stem
 - Scale ranges with only positive numbers convey a unipolar dimension
 - Also, if "0" is a scale endpoint, people may interpret it as absence of a characteristic
 - Using a range of numbers from negative to positive conveys a bipolar dimension
 - Numeric values chosen should match intended meaning



Estimation Biases

- ▶ Open-ended questions asking for frequency and amount are subject to bias
 - Memory can be inaccurate, especially for past events
 - Members of different groups may use different estimation techniques
 - Can lead to the appearance of differences that do not exist



Cultural Response Sets

- ▶ A cultural tendency to respond in a certain way on tests or response scales
 - What appear to be cultural differences are actually due to response sets
 - Therefore reflect a measurement confound or systematic error
 - Can also result from a reference group effect
 - People rate themselves relative to the average of their group or culture



Person-Related Biases

- ▶ Reflect people's tendency to respond in a biased manner
- ▶ Social desirability response bias: Tendency to respond in a way that makes the respondent look good
 - Most likely to affect responses when making a good impression is important
 - Less likely when responses are anonymous or when experimenter is not present



Social Desirability Response Bias

- ▶ Takes two forms
 1. Self-deceptive positivity: People believe they are responding honestly, but put themselves in overly positive light
 - Is a personality trait, so is consistent across time and situations
 2. Impression management: People deliberately respond in a way that creates the most positive social image
 - Extent to which this is done varies across time and situation



Acquiescence Response Bias

- ▶ The general tendency to agree or disagree with statements
- ▶ More likely for respondents who lack the skill or motivation to think about their answers
 - "Would you say that you are in favor of increasing funds for education?"
- ▶ Can be controlled by
 - writing questions that are clear and easy to answer
 - using a balanced measure, where for half of the items
 - agreement leads to higher score
 - agreement leads to lower score



Other Biases

- ▶ Neutral positions
 - Satisficing
 - 1st acceptable
- ▶ Floaters
 - Choose substantive answer when uninformed on issue
 - Standard-format question
 - Quasi-filter question ('don't know' option)
 - Full-filter question (contingency)
- Recency effect
- ▶ Selective refusals



Extremity Response Bias

- ▶ Tendency to give extreme responses, such as using only scale endpoints
- ▶ Difficult to control for
 - Is primarily a respondent characteristic
- ▶ Awareness of possibility is important as some groups of people are more likely to engage in this bias



Halo Bias

- ▶ Halo bias: Occurs when a respondent sees a stimulus as being positive or negative on an important characteristic
 - This perception affects how they rate other characteristics
 - Most likely to occur when characteristic of interest is
 - not readily observable
 - not clearly defined
 - involves relations with other people
 - is of moral importance



Leniency Bias

- ▶ Occurs when respondent's
 - positive characteristics are overestimated
 - negative characteristics are underestimated



Interpreting Responses

- ▶ Biases and other factors can lead to a literal interpretation fallacy
 - Occurs when the meanings of scale anchors are taken at face value
 - Anchors should be viewed as simple markers that help respondents determine which scale point to mark



Writing Survey Questions (cont.)

- Minimize fence-sitting and floating
 - Do not force neutral individuals to choose between opposites
 - Give individuals a "Don't know" or "no strong opinion" option
 - Ask for examples
 - Follow with measure of strength in belief
 - Omit neutral options when participants may have opinions they are reluctant to commit to



Constructing a Good Questionnaire

Questionnaire Design Issues

- ▶ Length of questionnaire
- ▶ Question order/sequence
 - Organization
 - Order effects
 - Context effects
 - Funnel sequence
 - Progressively narrower

EXPANSION BOX 10.7
Question Order Effects

QUESTION 1
"Do you think that the United States should let Communist newspaper reporters from other countries come in here and send back to their papers the news as they see it?"

QUESTION 2
"Do you think a Communist country like Russia should let U.S. newspaper reporters come in and send back to America the news as they see it?"

PERCENTAGE SAYING YES

| Heard First | Yes to Question 1 (Communist Reporter) | Yes to Question 2 (American Reporter) |
|-------------|-------------------------------------------|------------------------------------------|
| Question 1 | 54% | 75% |
| Question 2 | 64 | 82 |

The context created by answering the first question affects the answer to the second question.

Source: Adapted from Schuman and Presser (1981). *Questions and Answers in Attitude Surveys: Experiments in Question-Form, Wording, and Context*, p. 29. New York: Academic Press. With permission from Elsevier.

Question Sequencing

- ▶ Start with questions that have as many of these characteristics as possible:
 - easy to answer
 - interesting to respondents
 - clearly important
 - related to stated research purpose
- ▶ Demographic questions are easy to answer, but boring
 - Place them at the end

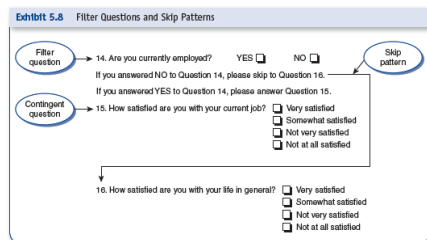
Question Sequencing

- ▶ Group questions together by topic
 - Helps keep respondents focused
 - Enhances recall
- ▶ Make transitions between topics clear
 - Transitions should emphasize the relevance of the new topic
- ▶ Within a topic, questions should move from the general to the specific
 - General questions
 - are seen as easier to answer
 - provide a context for the specific questions

Question Sequencing

- ▶ Pay attention to whether a question is contingent on a previous question
- ▶ Give detailed instructions on what respondent should do next
 - Example: Did you study today? Yes or No
 - If answer is YES, go on to Questions 11–13
 - If answer is NO, skip to Question 14 on p.3

Filter Questions and Skip Patterns



Source: Schutt, R. K. (2009). *Investigating the social world: The process and practice of research* (6th ed., p. 264). Thousand Oaks, CA: Pine Forge Press.

Questionnaire Layout

- ▶ Physical layout of questionnaire affects ease of use
- ▶ General guidelines include:
 - use closed-ended questions to greatest extent possible
 - use a consistent item format
 - use a vertical item format if possible
 - do not crowd the page
 - do not split questions or response options between pages

Instructions

- Instructions can affect the validity of the data
- Confusing or difficult-to-understand instructions often result in errors or non-response
- Should include repetition and examples to enhance understanding of the task
- Should provide an example of how to use each response format



Three categories of instructions

1. General instructions
 - Tell respondents how to answer both close-ended and open-ended questions
 - Explain the response formats and scales used
2. Introductions to topics
 - Explain the topic areas
 - Tie sections to overall research purpose
 - Function as transitions between topics
3. Specific instructions
 - Introduce questions that use a new response format



Instructions

- Avoid phrasing instructions in ways that might bias the results
 - Example: This questionnaire addresses the reasons why people give money to charities
 - Is biased because it assumes the respondents *do* or *should* give money to charities
 - Less biased: This questionnaire explores the factors influencing people's decision to donate or not donate to charities



Using Existing Measures

- Is a good idea because
 - reliability and validity of existing measures is known
 - it is easier to compare your results to those of other researchers
 - developing a new measure is a long and arduous process



Context Effects

- Completing one measure can influence scores on later measures
 - Can lead to spurious correlations between the measures
- Control possible bias by
 - psychologically separating questionnaires by indicating they are for two different studies
 - having two data collection sessions
 - counterbalancing order of presentation of questionnaires
 - randomly intermixing the items from each measure within the questionnaire

