

## The Role of Question Characteristics in Designing and Evaluating Survey Questions

Jennifer Dykema  
University of Wisconsin Survey Center  
University of Wisconsin-Madison

Nora Cate Schaeffer  
University of Wisconsin Survey Center  
Department of Sociology  
University of Wisconsin-Madison

Dana Garbarski  
Department of Sociology  
Loyola University Chicago

AAPOR Webinar  
January 21, 2016



© 2016. Materials may not be reproduced without permission of the author.



---

---

---

---

---

---

---

---

---

---

### Question writers focus on question characteristics ... question length

Do you feel you have ever been treated unfairly because of your race or ethnicity?

- <1> YES
- <2> NO

→ Word Count = 15

In general, would you say that your health is excellent, very good, good, fair, or poor?

- <1> EXCELLENT
- <2> VERY GOOD
- <3> GOOD
- <4> FAIR
- <5> POOR

→ Word Count = 16

During the past week, on about how many days did you have at least one drink of any alcoholic beverage such as beer, wine, a malt beverage, or liquor?

\_\_\_ NUMBER OF DAYS (0-30)

→ Word Count = 29

---

---

---

---

---

---

---

---

---

---

### Question writers focus on question characteristics ... difficulty level

Do you feel you have ever been treated unfairly because of your race or ethnicity?

- <1> YES
- <2> NO

→ Grade Level = 8.4

In general, would you say that your health is excellent, very good, good, fair, or poor?

- <1> EXCELLENT
- <2> VERY GOOD
- <3> GOOD
- <4> FAIR
- <5> POOR

→ Grade Level = 6.0

During the past week, on about how many days did you have at least one drink of any alcoholic beverage such as beer, wine, a malt beverage, or liquor?

\_\_\_ NUMBER OF DAYS (0-30)

→ Grade Level = 12.2

---

---

---

---

---

---

---

---

---

---

**Question writers focus on question characteristics ... response format**

Do you feel you have ever been treated unfairly because of your race or ethnicity?

<1> YES  
<2> NO

Yes-No Response Format

In general, would you say that your health is excellent, very good, good, fair, or poor?

<1> EXCELLENT  
<2> VERY GOOD  
<3> GOOD  
<4> FAIR  
<5> POOR

Selection Response Format

During the past week, on about how many days did you have at least one drink of any alcoholic beverage such as beer, wine, a malt beverage, or liquor?

\_\_\_ NUMBER OF DAYS (0-30)

Discrete Value Response Format

---

---

---

---

---

---

---

---

---

---

**Where we are now ... making recommendations**

- Recommendations for writing questions are
  - formulated around question characteristics
  - based on research about impact on outcomes

5

---

---

---

---

---

---

---

---

---

---

**Where we are now ... knowledge has cumulated**

- Know a lot about effects of some characteristics on data quality

Alwin 2007; Bradburn et al. 2004; Dillman et al. 2014; Fowler 1995; Fowler & Cosenza 2008; Krosnick & Presser 2010; Presser et al. 2004; Saris & Gallhofer 2007; Schaeffer & Presser 2003; Schaeffer & Dykema 2011, 2015; Schuman & Presser 1996; Tourangeau et al. 2000

6

---

---

---

---

---

---

---

---

---

---

Where we are now ... still under development

- Developing a comprehensive typology in which
  - characteristics are cataloged and organized
  - effects on respondents' and interviewers' cognitive processing are understood
  - effects on data quality are documented

7

---

---

---

---

---

---

---

---

Outline and objectives

Overview of approaches

Taxonomy of features

Case studies

Summary and future directions

8

---

---

---

---

---

---

---

---

Notes to help frame our goals for today



Not presenting a list of guidelines



Self-documenting slides



9

---

---

---

---

---

---

---

---

**Outline and objectives**

- Overview of approaches
- Taxonomy of features
- Case studies
- Summary and future directions

10

---

---

---

---

---

---

---

---

**General approaches to identify and evaluate question characteristics**

|                       | Experimental  | Observational   |
|-----------------------|---|---|
| <b>Scope</b>          | Limited   | Wide-ranging  |
| <b>Design</b>         | Write alternative question forms that incorporate the characteristic<br><br>Other characteristics held constant | Code questions along the dimensions of all characteristics<br><br>Other characteristics vary across questions |
| <b>Administration</b> | Rs randomly exposed to a subset of characteristics  | Rs exposed to all questions and characteristics   |
| <b>Evaluation</b>     | Assess effect on predetermined criterion  | Assess impact on outcome available for all questions  |

11

---

---

---

---

---

---

---

---

**Two kinds of observational approaches**

- Ad-hoc
- System-based

12

---

---

---

---

---

---

---

---

### Ad-hoc observational approaches: Kinds of characteristics

- Examine individual question characteristics
  - General characteristics
    - Question type
    - Question length
    - Response format
    - Instructions to respondents
  - Specific characteristics
    - Ambiguous terms
    - End-point only labeling
    - Mismatch between question text and response categories

13

---

---

---

---

---

---

---

---

### Ad-hoc observational approaches: Outcomes

- Outcomes
  - Response times, item-missing responses, respondent and interviewer behaviors
- Some representative studies
  - Couper & Kreuter 2013; Holbrook et al. 2006; Johnson et al. 2015; Kasabian et al. 2014; Knauper et al. 1997; Olson & Smyth 2015; Yan & Tourangeau 2008

14

---

---

---

---

---

---

---

---

### System-based observational approaches

- Code questions using established scheme or system
  - Identify problems through different sources
    - Observation in cognitive interviews
    - Speculation based on response process model
- 
- ```
graph LR; A(Comprehension) --> B(Retrieval); B --> C(Judgment); C --> D(Formatting Response)
```
- Past analysis of effect of characteristics on data quality
  - Representative studies: Forsyth et al. 2004 (CCS); Saris & Gallhofer 2007 (SQP); Willis 2005 (QAS)

15

---

---

---

---

---

---

---

---

| Outline and objectives        |  |
|-------------------------------|--|
| Overview of approaches        |  |
| Taxonomy of features          |  |
| Case studies                  |  |
| Summary and future directions |  |

---

---

---

---

---

---

---

---

| Taxonomy of features                                          |                                                                           |
|---------------------------------------------------------------|---------------------------------------------------------------------------|
| Classes of Characteristics                                    | Examples of individual features                                           |
| Question topic                                                | Health, politics                                                          |
| Question type                                                 | Event or behavior, evaluation or judgment, classification                 |
| Response dimension                                            | Occurrence, frequency, intensity, valence                                 |
| Conceptualization and operationalization of the target object | Labels for target object and response dimension                           |
| Question structure                                            | Filter and follow-up question, battery                                    |
| Response format or question form                              | Yes/no, selection, discrete value, field-coded open, record-verbatim open |
| Response categories                                           | Type, number, and labeling                                                |
| Question wording                                              | Length, readability                                                       |
| Question implementation                                       | Mode, orientation of scale on screen, instructions to interviewers        |

---

---

---

---

---

---

---

---

| Taxonomy of features                                          |                                                                           |
|---------------------------------------------------------------|---------------------------------------------------------------------------|
| Classes of Characteristics                                    | Examples of individual features                                           |
| Question topic                                                | Health, politics                                                          |
| Question type                                                 | Event or behavior, evaluation or judgment, classification                 |
| Response dimension                                            | Occurrence, frequency, intensity, valence                                 |
| Conceptualization and operationalization of the target object | Labels for target object and response dimension                           |
| Question structure                                            | Filter and follow-up question, battery                                    |
| Response format or question form                              | Yes/no, selection, discrete value, field-coded open, record-verbatim open |
| Response categories                                           | Type, number, and labeling                                                |
| Question wording                                              | Length, readability                                                       |
| Question implementation                                       | Mode, orientation of scale on screen, instructions to interviewers        |

---

---

---

---

---

---

---

---

**Question characteristics and decisions:  
Two examples**

- Decisions for questions about events and behaviors
- Decisions for questions about evaluations and judgments
- Two illustrative example decisions
  - Filter questions
  - Response category labels

19

---

---

---

---

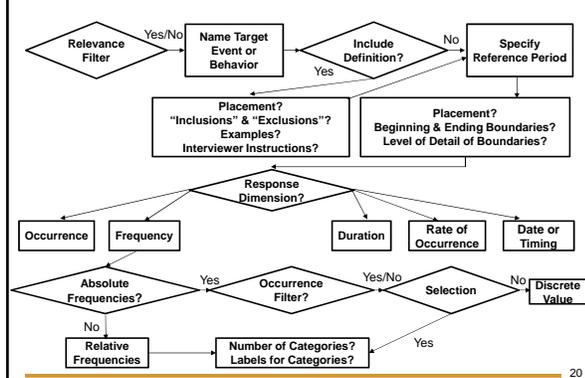
---

---

---

---

**Question about events and behaviors –  
decisions (Schaeffer & Dykema 2011)**



20

---

---

---

---

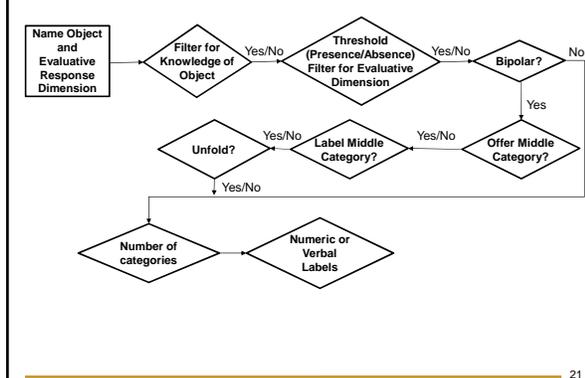
---

---

---

---

**Question asking for evaluations and judgments  
– decisions (Schaeffer & Dykema 2011)**



21

---

---

---

---

---

---

---

---

**Decision: Filter questions**

**Events**

**Structure 1 – With Filter**

During the last twelve months, did you go to a museum?

**IF YES:** During the last twelve months, how many times did you go to a museum?

**Structure 2: Without filter**

During the last twelve months, how times, if any, did you go to a museum?

**Evaluations & Judgements**

**Structure 1 – with Filter**

Are you concerned about pesticides in your food?

**IF YES:** How concerned are you about pesticides in your food: a little bit, somewhat, very, or extremely?

**Structure 2: Without filter**

How concerned are you about pesticides in your food: not at all, a little bit, somewhat, very, or extremely?

22

---

---

---

---

---

---

---

---

---

---

**Comparison: Event question and evaluation question**

- Differences
  - Target objects
  - Response dimensions
  - Cognitive processes
- Implications
  - Meaning differs
  - Impact on quality may differ

23

---

---

---

---

---

---

---

---

---

---

**Decisions and analytic challenges**

- The consequences of item characteristics may vary by question type or response format
  - There may be statistical interactions
  - Estimating interactions with statistical power requires designing appropriate combinations of question characteristics in sample

24

---

---

---

---

---

---

---

---

---

---

**Decision: Category labels**

| Events<br>Form: Discrete value                                                                                                 | Evaluations & Judgements<br>Form: Selection with Ordered Categories                                                                                                                                                                                                        |
|--------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Version 1 – Unlabeled</b><br>In the last 12 months, how many times, if any, did you visit a museum?<br>____ NUMBER OF TIMES | <b>Version 1 – End-points labeled</b><br>Do people in these groups tend to be unintelligent or intelligent?<br><br>A. Where would you rate whites in general on this scale?<br>B. Blacks? |
| <b>Version 2: Labeled</b><br>--Null--                                                                                          | <b>Version 2: Labeled</b><br>Do people in these groups tend to be unintelligent or intelligent?<br><br>A. Where would you rate whites in general on this scale?<br>B. Blacks?             |

---

---

---

---

---

---

---

---

---

---

---

---

**Decisions and analytic challenges**

- The consequences of item characteristics may depend on question type or response format
  - There may be statistical interactions
  - Estimating interactions with statistical power requires designing appropriate combinations of question characteristics in sample
- Some question characteristics are relevant only for some question types or response formats
  - There are structural dependencies or empty cells
- Implications
  - Consider possible interactions and empty cells when designing analyses
  - Consider possible interactions and empty cells when generalizing from or applying results from studies

26

---

---

---

---

---

---

---

---

---

---

---

---

**Outline and objectives**

- Overview of approaches
- Taxonomy of features
- Case studies
- Summary and future directions

27

---

---

---

---

---

---

---

---

---

---

---

---

## Case studies: Overview of design and methods

|                  | Wisconsin Longitudinal Study (WLS)                |
|------------------|---------------------------------------------------|
| Sample           | 1/3 random sample of WI high school class of 1957 |
| Mode             | Telephone                                         |
| Dates            | 2003-2005                                         |
| Question topics  | Health                                            |
| Question types   | Events and behaviors                              |
| Outcomes         | Q-reading accuracy;<br>Any "problem" behaviors    |
| Unit of analysis | Question-answer sequence<br>(n = 8,150)           |
| Analytic method  | Cross-classified random effects logistic models   |
| Further details  | Dykema et al. 2013                                |

28

---

---

---

---

---

---

---

---

---

---

## Question characteristics and coding systems

- General question characteristics
  - Response format
  - Question wording
    - Question length
    - Readability
    - Comprehension difficulty (QUAID)
- Coding systems
  - Problem Classification Coding Scheme (CCS)
  - Question Appraisal System (QAS)
  - Survey Quality Predictor (SQP)

29

---

---

---

---

---

---

---

---

---

---

## General characteristics: Response format

- Format projected by question for responding
  - Yes-No
    - *Have you been able to see at all?*
  - Selection
    - *In general, would you say your health is excellent, very good, good, fair, or poor?*
- Predictions
  - Interviewer question-asking accuracy
    - Selection question → reading error more likely
  - Respondent problem behaviors
    - Selection question → problem behavior more likely

30

---

---

---

---

---

---

---

---

---

---

**General characteristics: Question wording measures**

- Question length
  - Count of total number of words
- Question readability
  - Flesch-Kincaid Grade Level score
- Question comprehension difficulty
  - Question Understanding Aid (QUAID) (Graesser et al. 2006)

31

---

---

---

---

---

---

---

---

**General characteristics: Question wording predictions**

- Interviewer less likely to read question exactly with
  - Longer
  - Harder-to-read
  - Higher comprehension difficulty
- Respondent more likely to display problem behavior with
  - Longer
  - Harder-to-read
  - Higher comprehension difficulty

32

---

---

---

---

---

---

---

---

**Overview of coding systems**

- Problem Classification Coding Scheme (CCS) (Forsyth et al. 2004)
- Question Appraisal System (QAS) (Willis 2005)
- Survey Quality Predictor (SQP) (Saris & Gallhofer 2007)

33

---

---

---

---

---

---

---

---

### Predictions for coding schemes

- Interviewer less likely to read question exactly with
  - Higher CCS scores
  - Higher QAS scores
  - Lower SQP scores
- Respondent more likely to display problem behavior with
  - Higher CCS scores
  - Higher QAS scores
  - Lower SQP scores

34

---

---

---

---

---

---

---

---

---

---

### Values for measures for two questions

|    | Question Length | Question Readability | Comprehension Difficulty QUAID | CCS Problems | QAS Problems | SQP Data Quality Prediction |
|----|-----------------|----------------------|--------------------------------|--------------|--------------|-----------------------------|
| Q1 | Long            | Low                  | Medium                         | High         | High         | Low                         |
| Q2 | Short           | High                 | Medium                         | Low          | Low          | Medium                      |

Q1. *Because of any impairment or health problem, do you need the help of other persons in handling your routine needs, such as everyday household chores, doing necessary business, shopping or getting around for other purposes?*

Q2. *Have you ever been diagnosed with a mental illness?*

35

---

---

---

---

---

---

---

---

---

---

### Results from bivariate cross-classified random effects logistic models of interviewer and respondent outcomes on question characteristics

| Question Characteristics & Systems         | Interviewer Exact Reading (Odds) | Respondent Problems (Any) (Odds) |
|--------------------------------------------|----------------------------------|----------------------------------|
| Response format [Yes/No] Selection         | ↓                                | ↑***                             |
| Question length                            | ↓***                             | ↑***                             |
| Question readability                       | ↓                                | ↑                                |
| Comprehension difficulty (QUAID)           | ↓*                               | ↑***                             |
| Problem Classification Coding System (CCS) | ↓**                              | ↑*                               |
| Question Appraisal System (QAS)            | ↓                                | ↑                                |
| Survey Quality Predictor (SQP)             | ↑**                              | ↓                                |

\*=p<0.05; \*\*=p<0.01; \*\*\*=p< 0.001

36

---

---

---

---

---

---

---

---

---

---

**Case studies: Overview of design and methods**

|                  | Wisconsin Longitudinal Study (WLS)                | General Social Survey (GSS)                                          |
|------------------|---------------------------------------------------|----------------------------------------------------------------------|
| Sample           | 1/3 random sample of WI high school class of 1957 | National area probability sample, noninstitutionalized US population |
| Mode             | Telephone                                         | Face-to-face                                                         |
| Dates            | 2003-2005                                         | 2006, 2008, and 2010                                                 |
| Question topics  | Health                                            | Varied, public opinion                                               |
| Question types   | Events and behaviors                              | Varied, 2/3 are evaluations                                          |
| Outcomes         | Q-reading accuracy; Any "problem" behaviors       | Individual question reliabilities                                    |
| Unit of analysis | Question-answer sequence (n = 8,150)              | Questions (n = 196)                                                  |
| Analytic method  | Cross-classified random effects logistic models   | Regression                                                           |
| Further details  | Dykema et al. 2013                                | Schaeffer et al. 2015                                                |

37

---

---

---

---

---

---

---

---

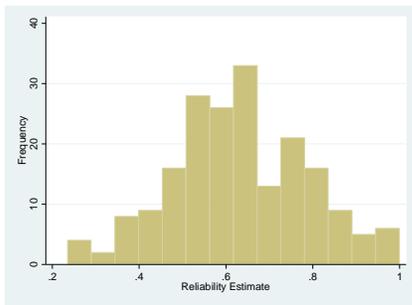
---

---

---

---

**Histogram of distribution of reliability, n=196 GSS core items, 2006-2010**



38

---

---

---

---

---

---

---

---

---

---

---

---

**General Social Survey – an item with high reliability**

IF CURRENTLY MARRIED OR WIDOWED:  
 A. Have you ever been divorced or legally separated?

Yes      1  
 No        2

39

---

---

---

---

---

---

---

---

---

---

---

---

**General Social Survey – an item with low reliability**

Do people in these groups tend to be unintelligent or intelligent?

|                     |               |   |   |   |   |   |             |              |
|---------------------|---------------|---|---|---|---|---|-------------|--------------|
| HAND<br>CARD<br>A15 | Unintelligent |   |   |   |   |   | Intelligent | DONT<br>KNOW |
|                     | 1             | 2 | 3 | 4 | 5 | 6 | 7           |              |
| RATING              |               |   |   |   |   |   |             |              |

- A. Where would you rate whites in general on this scale?
- B. Blacks?

40

---

---

---

---

---

---

---

---

**Case Study: General Social Survey**

- Operationalizing complexity of question wording
- Decisions about batteries

41

---

---

---

---

---

---

---

---

**Battery**

- Two or more items
  - Presented together
  - With the same response categories
- The first target question is usually preceded by an introduction.

42

---

---

---

---

---

---

---

---

**General Social Survey: The nat\* battery**

1. First I would like to talk with you about some things people think about today. We are faced with many problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount. First (**READ ITEM A**)... are we spending too much, too little or about the right amount on (ITEM)?  
 READ EACH ITEM; CODE ONE FOR EACH.

|  | Too much | Too little | About right | DON'T KNOW |
|--|----------|------------|-------------|------------|
|--|----------|------------|-------------|------------|

- A. The space exploration program      3      1      2      8
- B. Improving and protecting the environment      3      1      2      8
- C. Improving and protecting the nation's health      3      1      2      8

43

---

---

---

---

---

---

---

---

---

---

**General Social Survey: The nat\* battery – item A**

1. First I would like to talk with you about some things people think about today. We are faced with many problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount. First (**READ ITEM A**)... are we spending too much, too little or about the right amount on (ITEM)?  
 READ EACH ITEM; CODE ONE FOR EACH.

|  | Too much | Too little | About right | DON'T KNOW |
|--|----------|------------|-------------|------------|
|--|----------|------------|-------------|------------|

- A. The space exploration program      3      1      2      8
- B. Improving and protecting the environment      3      1      2      8
- C. Improving and protecting the nation's health      3      1      2      8

44

---

---

---

---

---

---

---

---

---

---

**General Social Survey: The nat\* battery – item B**

1. First I would like to talk with you about some things people think about today. We are faced with many problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount. First (**READ ITEM A**)... are we spending too much, too little or about the right amount on (ITEM)?  
 READ EACH ITEM; CODE ONE FOR EACH.

|  | Too much | Too little | About right | DON'T KNOW |
|--|----------|------------|-------------|------------|
|--|----------|------------|-------------|------------|

- A. The space exploration program      3      1      2      8
- B. Improving and protecting the environment      3      1      2      8
- C. Improving and protecting the nation's health      3      1      2      8

45

---

---

---

---

---

---

---

---

---

---

**General Social Survey: The nat\* battery**

1. First I would like to talk with you about some things people think about today. We are faced with many problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount. First (READ ITEM A)... are we spending too much, too little or about the right amount on (ITEM)?  
 READ EACH ITEM; CODE ONE FOR EACH.

|  | Too much | Too little | About right | DON'T KNOW |
|--|----------|------------|-------------|------------|
|--|----------|------------|-------------|------------|

|                                                 |   |   |   |   |
|-------------------------------------------------|---|---|---|---|
| A. The space exploration program                | 3 | 1 | 2 | 8 |
| B. Improving and protecting the environment     | 3 | 1 | 2 | 8 |
| C. Improving and protecting the nation's health | 3 | 1 | 2 | 8 |

46

---

---

---

---

---

---

---

---

---

---

---

---

**Question wording: Measures of the complexity of language**

- Number of words
  - All words needed to answer the questions
    - In batteries, we included the preamble in
    - calculating the word count for second and later items
  - Words actually read to the respondent – according to the script
  - Difference between “all words” and “read words”
- QUAID measures (Graesser et al. 2006@)
  - Number of left-embedded words - words before main verb
  - Sum of number of problem words – technical or vague – and other problems
- Flesch measures

47

---

---

---

---

---

---

---

---

---

---

---

---

**General Social Survey: The nat\* battery – item B**

1. First I would like to talk with you about some things people think about today. We are faced with many problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount. First (READ ITEM A)... are we spending too much, too little or about the right amount on (ITEM)?  
 READ EACH ITEM; CODE ONE FOR EACH.

|  | Too much | Too little | About right | DON'T KNOW |
|--|----------|------------|-------------|------------|
|--|----------|------------|-------------|------------|

|                                                 |   |   |   |   |
|-------------------------------------------------|---|---|---|---|
| A. The space exploration program                | 3 | 1 | 2 | 8 |
| B. Improving and protecting the environment     | 3 | 1 | 2 | 8 |
| C. Improving and protecting the nation's health | 3 | 1 | 2 | 8 |

48

---

---

---

---

---

---

---

---

---

---

---

---

**General Social Survey: The nat\* battery – item B**

1. First I would like to talk with you about some things people think about today. We are faced with many problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount. First (READ ITEM A)... are we spending too much, too little or about the right amount on (ITEM)?  
 READ EACH ITEM; CODE ONE FOR EACH.

|  | Too much | Too little | About right | DON'T KNOW |
|--|----------|------------|-------------|------------|
|--|----------|------------|-------------|------------|

|                                                 |   |   |   |   |
|-------------------------------------------------|---|---|---|---|
| A. The space exploration program                | 3 | 1 | 2 | 8 |
| B. Improving and protecting the environment     | 3 | 1 | 2 | 8 |
| C. Improving and protecting the nation's health | 3 | 1 | 2 | 8 |

49

---

---

---

---

---

---

---

---

---

---

---

---

**Question wording: Measures of the complexity of language**

- Number of words
  - All words needed to answer the questions
    - In batteries, we included the preamble in
    - calculating the word count for second and later items
  - Words actually read to the respondent – according to the script
    - Difference between "all words" and "read words"
- QUAID measures (Graesser et al. 2006@)
  - Number of left-embedded words - words before main verb
  - Sum of number of problem words – technical or vague – and other problems
- Flesch measures

50

---

---

---

---

---

---

---

---

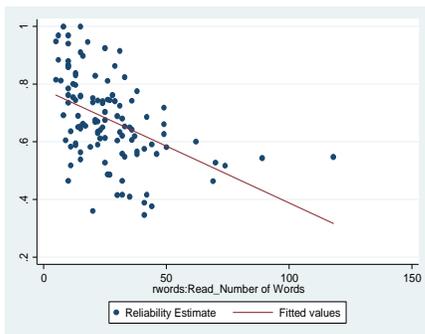
---

---

---

---

**Reliability by the number of words read to the respondent, items not in a battery, N=115**



51

---

---

---

---

---

---

---

---

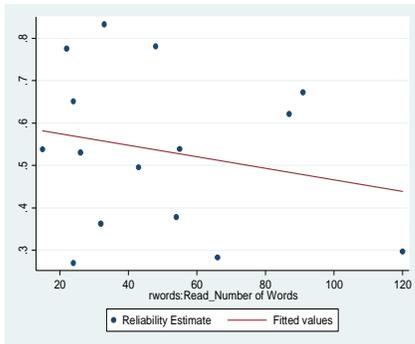
---

---

---

---

**Reliability by the number of words read to the respondent, items first in a battery, N = 15**



52

---

---

---

---

---

---

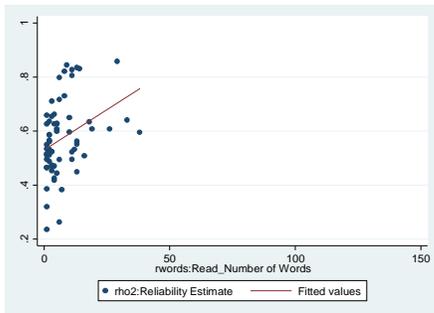
---

---

---

---

**Reliability by the number of words read to the respondent, items second or later in a battery, N = 66**



53

---

---

---

---

---

---

---

---

---

---

**Regression of item reliability on measures of question wording and position in battery, N=196**

| Variable                                  | Effect |
|-------------------------------------------|--------|
| Left-embedded words (QUAID)               | _***   |
| Problem words & language problems (QUAID) | -      |
| Number of words read                      |        |
| First item [Not in battery omitted]       |        |
| First item*Number of words                |        |
| Second or later item                      |        |
| Second or later item*Number of words      |        |

\*=p<0.10, \*\*=p<0.05, \*\*\*=p<0.001

54

---

---

---

---

---

---

---

---

---

---

**Regression of item reliability on measures of question wording and position in battery, N=196**

| Variable                                  | Effect |
|-------------------------------------------|--------|
| Left-embedded words (QUAID)               | -.***  |
| Problem words & language problems (QUAID) | -      |
| Number of words read                      | -.***  |
| First item [Not in battery omitted]       | -.**   |
| First item*Number of words                | +*     |
| Second or later item                      | -      |
| Second or later item*Number of words      | +***   |

\*=p<0.10, \*\*=p<0.05, \*\*\*=p<0.001

55

---

---

---

---

---

---

---

---

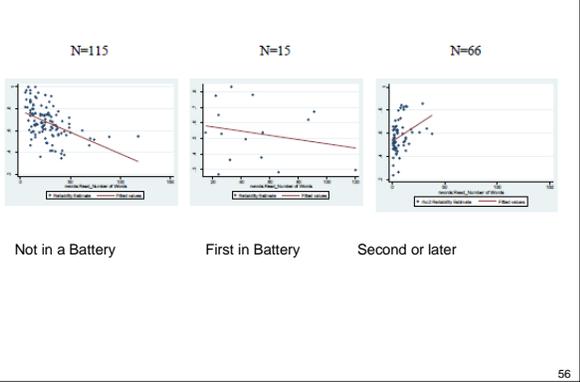
---

---

---

---

**Reliability estimates by the number of words read to the respondent, by placement in a battery**



56

---

---

---

---

---

---

---

---

---

---

---

---

**Question type – Type of target object + response dimension**

- Asking for an evaluation or judgment (“Evaluation questions”) – N=150
- Asking about events or behaviors (“Event questions”) – N = 27
- Event-based classifications and calculations (“Event-based classifications), N = 19

57

---

---

---

---

---

---

---

---

---

---

---

---

**Regression of reliability on language measures, position in battery, and question type, N = 196**

| Variable                                  | Effect |
|-------------------------------------------|--------|
| Question wording                          |        |
| Left-embedded words (QUAID)               | -.***  |
| Problem words & language problems (QUAID) | +      |
| Number of words read                      | -.***  |
| First item [Not in battery omitted]       | -      |
| First item*Number of words                | +      |
| Second or later item                      | +      |
| Second or later item*Number of words      | +.**   |
| Question type [Evaluation omitted]        |        |
| Event question                            | +***   |
| Event-based classification                | +***   |

\*=p<0.10, \*\*=p<0.05, \*\*\*=p<0.001

58

---

---

---

---

---

---

---

---

---

---

**Case study: Comments on the GSS**

- Results depend on the sample of items and distribution of item characteristics in the GSS
- Some evaluation questions in the GSS have very good reliabilities
- Evaluation and event questions are not interchangeable in measurement
- Topic is confounded with other characteristics

59

---

---

---

---

---

---

---

---

---

---

**Outline and objectives**

- Overview of approaches
- Taxonomy of features
- Case studies
- Summary and future directions

60

---

---

---

---

---

---

---

---

---

---

## Summary

- Comments on the case studies
- New coding system
- Importance of a criterion
- Future directions

---

---

---

---

---

---

---

---

Thank You!

For questions about this webinar or more information,  
contact:

Jennifer Dykema  
[dykema@ssc.wisc.edu](mailto:dykema@ssc.wisc.edu)

Nora Cate Schaeffer  
[schaeffe@ssc.wisc.edu](mailto:schaeffe@ssc.wisc.edu)

Please visit us at:  
[www.uwsc.wisc.edu](http://www.uwsc.wisc.edu)



---

---

---

---

---

---

---

---

## References

- Alwin, Duane F. 2007. *Margins of Error: A Study of Reliability in Survey Measurement*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Alwin, Duane F. and Jon A. Krosnick. 1991. "The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes." *Sociological Methods and Research* 20:139-81.
- Bradburn, Norman M., Seymour Sudman and Brian Wansink. 2004. *Asking Questions: The Definitive Guide to Questionnaire Design*. New York: Wiley.
- Couper, Mick P. and Frauke Kreuter. 2013. "Using Paradata to Explore Item Level Response Times in Surveys." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176:271-86.
- Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, Fourth Edition*. Hoboken, NJ: John Wiley & Sons, Inc.
- Dykema, Jennifer, Nora Cate Schaeffer, Dana Garbarski, Rick Nordheim, and Kristen Cyffka. 2013. "Effects of Question, Respondent, and Interviewer Characteristics on Interactional Indicators of Respondent and Interviewer Processing of Health-Related Questions." Paper presented at the Interviewer-Respondent Interaction Workshop, May, Boston, MA.
- Fowler, Jr., Floyd J., ed. 1995. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage.
- Fowler Jr., Floyd J., and Carol Cosenza. 2008. "Writing Effective Questions." Pp. 136-60 in *International Handbook of Survey Methodology*, edited by Edith D. de Leeuw, Joop J. Hox and Don A. Dillman. New York: Lawrence Erlbaum Associates, 2008.

---

---

---

---

---

---

---

---

## References (continued)

- Forsyth, Barbara, Jennifer M. Rothgeb, and Gordon B. Willis. 2004. "Does Pretesting Make a Difference? An Experimental Test." In *Methods for Testing and Evaluating Survey Questionnaires*, edited by S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer, 525-46. New York: Springer-Verlag.
- Graesser, Arthur C., Zhiqiang Cai, Max M. Louwse, and Frances Daniel. 2006. "Question Understanding AID (QUAID): A Web Facility That Tests Question Comprehensibility." *Public Opinion Quarterly* 70:3-22.  
<http://quaid.cohmetix.com>
- Holbrook, Allyson, Young Ik Cho, and Timothy Johnson. 2006. "The Impact of Question and Respondent Characteristics on Comprehension and Mapping Difficulties." *Public Opinion Quarterly* 70:565-95.
- Johnson, Timothy, et al. 2015. "Sources of Interactional Problems in a Survey of Racial/Ethnic Discrimination." *International Journal of Public Opinion Research* 27:244-263.
- Knauper, Barbel, Robert F. Belli, Daniel H. Hill, and A. Regula Herzog. 1997. "Question Difficulty and Respondents Cognitive Ability: The Effect on Data Quality." *Journal of Official Statistics* 13:181-99.
- Kasabian, Allan, Jolene Smyth, and Kristen Olson. 2014. "The Whole is More than the Sum of its Parts: Understanding Item Nonresponse in Self-Administered Surveys." Paper presented at the annual meeting of the American Association for Public Opinion Research, May, Anaheim, CA.

64

## References (continued)

- Krosnick, Jon A., and Stanley Presser. 2010. "Question and Questionnaire Design." In *Handbook of Survey Research*, 2nd ed., edited by P.V. Marsden and J.D. Wright, 263-313. Bingley, UK: Emerald.
- Olson, Kristen and Jolene D. Smyth. 2015. "The Effect of Cati Questions, Respondents, and Interviewers on Response Time." *Journal of Survey Statistics and Methodology* 3:361-396.
- Presser, Stanley, Jennifer M. Rothgeb, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer. 2004. *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, NJ: John Wiley & Sons, Inc.
- Rodgers, Willard L., Frank M. Andrews and A. Regula Herzog. 1992. "Quality of Survey Measures: A Structural Modeling Approach." *Journal of Official Statistics* 8:251-75.
- Saris, Willem E., and Irmtraud Galhofer. 2007. *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. New York: Wiley.  
<http://snp.upf.edu/>
- Schaeffer, Nora Cate, Minne Chen, Jennifer Dykema, Dana Garbaski and Michael A. Hout. 2015. "Question Characteristics and Item Reliability." Paper presented at the annual meeting of the Midwest Association for Public Opinion Research, November, Chicago, IL.
- Schaeffer, Nora Cate, and Jennifer Dykema. 2011. "Questions for Surveys: Current Trends and Future Directions." *Public Opinion Quarterly* 75:909-61.  
<http://poq.oxfordjournals.org/content/75/5/909.full.pdf+html>
- Schaeffer, Nora Cate, and Jennifer Dykema. 2015. "Question Wording and Response Categories." In *International Encyclopedia of the Social and Behavioral Sciences*, edited by James D. Wright. Oxford: Elsevier.

65

## References (continued)

- Schaeffer, Nora Cate, and Stanley Presser. 2003. "The Science of Asking Questions." *Annual Review of Sociology* 29:65-88.  
[http://ajournals.annualreviews.org/eprint/rtU4UOoizjXROhjkRIS/full/10.1146/annurev\\_soc.29.110702.110112](http://ajournals.annualreviews.org/eprint/rtU4UOoizjXROhjkRIS/full/10.1146/annurev_soc.29.110702.110112)
- Schuman, Howard, and Stanley Presser. 1996. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context*. Orlando, FL: Academic Press.
- Stempel, Arthur, Rex H. Warland and Robert O. Herrmann. 1994. "Do Surveys Overstate Public Concern?" *Public Opinion Quarterly* 58:255-63.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, England: Cambridge University Press.
- Willis, Gordon B. 2005. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage.
- Yan, Ting, and Roger Tourangeau. 2008. "Fast Times and Easy Questions: The Effects of Age, Experience, and Question Complexity on Web Survey Response Times." *Applied Cognitive Psychology* 22:51-68.

66