You Can’t Fix by Analysis what You’ve Spoiled by Design:
A Workshop in Survey Design for Medical Educators

Jeffrey La Rochelle, MD, MPH
Amy Shaheen, MD

Academic Internal Medicine Week
2013 CDIM National Meeting, New Orleans, LA
October 4, 2013

Learning Objectives

- Recognize how to use a systematic, 7-step process as a framework for survey design
- Demonstrate how to develop an appropriate set of items to characterize the construct being measured
- Identify common item-writing pitfalls in survey design
- Define the purpose of expert validation, cognitive interviews, and pilot testing

The presenters are U.S. Government employees. The views expressed in this presentation are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, Department of Defense, nor the U.S. Government.
Consider this…

- The puppy problem

- The poodle has 9 puppies.
- The collie has 5 puppies.
- How many more puppies does the poodle have?

- Students’ common response…
  “None”

- Why?
  “It said she had 9 puppies, but it didn’t say she had any more, so it’s none.”

Revised item…

- The poodle has 9 puppies.
- The collie has 5 puppies.
- How many more puppies does the poodle have than the collie?

And this…

Your opinion is that the global economy is the second most important issue in the world today.

The global economy is the most important issue in the world today.

<table>
<thead>
<tr>
<th>strongly disagree</th>
<th>disagree</th>
<th>neither agree nor disagree</th>
<th>agree</th>
<th>strongly agree</th>
</tr>
</thead>
</table>

VS.

How important is the issue of the global economy in the world today?

<table>
<thead>
<tr>
<th>not at all important</th>
<th>slightly important</th>
<th>moderately important</th>
<th>quite important</th>
<th>extremely important</th>
</tr>
</thead>
</table>
Principle #1: You can’t fix by analysis what you’ve bungled by design.

- When creating a survey, it’s important to get it right the first time

- We often use surveys to measure “fuzzy” constructs (e.g., beliefs, attitudes, opinions), which is already quite difficult
  - *Bad data only makes it that much harder*

Principle #1: You can’t fix by analysis what you’ve bungled by design.

- Why worry about survey design problems?

  **On the one hand:**
  - A self-report survey is a “blunt instrument”
  - There is bound to be some “slop”
  - To the extent survey design IS rocket science, we’re content to just hit the moon, somewhere…

  **On the other hand:**
  - Of the various types of errors afflicting surveys – sampling error, non-response error, processing error…
  - *Response Error* (e.g., bias or unreliability) is often the largest category
Outline

- Survey Background
  - Uses of surveys
  - Survey language
- Survey Design
  - 7-Step Process
  - Small-group activity
- Common Item-Writing Pitfalls
- Survey Pretesting
  - Expert Validation
  - Cognitive Interviews
  - Pilot Testing
- Questions

Good for...

- Abstract ideas/concepts
  - Opinions
  - Beliefs
  - Attitudes
- Behaviors
  - Assuming that...
    - observing behaviors is impractical
    - people might reasonably report on their own behaviors

Academic Medicine (2009)
- Survey Methods: 56%
- Other Methods: 44%

Medical Education (2010 to 2012)
- Survey Methods: 23%
- Other Methods: 77%

Journal of Graduate Medical Education (2010 to 2012)
- Survey Methods: 13%
- Other Methods: 87%
Not so good for…

- Clinical outcomes
  - Better to measure perceptions (ex: health questionnaires)

- Populations with cognitive impairment, severe disease
  - Very sick patients may have difficulty with surveys

- Tasks of high cognitive load/burden
  - “How many hours did you use the Internet last year?”
  - “What did you eat for dinner on Wed, Jan 11, 2012?”

Survey Language

- Construct:
  - A model, idea, or theory
    - e.g., resilience, confidence, patient satisfaction, motivation, perceived barriers, interest, procrastination, health-related stigma

- Items (or “indicators”):
  - Individual questions/statements on the survey

- Scale:
  - 3 or more items intended to measure a construct
Survey Language

- Response anchors (aka, “response options” or “the response scale”):
  - All the named points along the response scale

For example:

<table>
<thead>
<tr>
<th>not at all important</th>
<th>slightly important</th>
<th>moderately important</th>
<th>quite important</th>
<th>extremely important</th>
</tr>
</thead>
<tbody>
<tr>
<td>never true</td>
<td>rarely true</td>
<td>sometimes true</td>
<td>often true</td>
<td>true nearly all of the time</td>
</tr>
<tr>
<td>almost never</td>
<td>once in a while</td>
<td>sometimes</td>
<td>often</td>
<td>almost all the time</td>
</tr>
</tbody>
</table>

- Satisficing:
  - Occurs when respondents compromise standards to expend less energy
    - i.e., they don’t put forth effort to answer truthfully or thoughtfully

Customer Service Questionnaire

Please help us improve our services by answering the questions below. We are interested in your opinions about the recent service you received from the Reading Fire Department.

Please rate your response:

<table>
<thead>
<tr>
<th>Very Satisfied</th>
<th>Satisfied</th>
<th>Adequate</th>
<th>Unsatisfied</th>
<th>Very Unsatisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

1. Were our personnel polite and courteous?  5 4 3 2 1
2. Did our personnel take care of you in a professional manner?  5 4 3 2 1
3. Did we explain the services you needed in an understandable manner?  5 4 3 2 1
4. Did we answer all of your questions in an understandable way?  5 4 3 2 1
5. Overall, how satisfied were you with the service you received from us?  5 4 3 2 1

A. Did you have any other concerns, related to your emergency, that you felt was not addressed by our personnel?

B. Please tell us the single most important action we took that made you feel better.

C. What could we have done differently that might have made your experience more positive?
Principle #2:
The questions guide the answers.

9) What topic(s) of study are you most interested in pursing while at university? (Total N = 11)

Ver 1: Lots of Space
- Financing of health care
- Global health, joint operations
- Policy development with regard to military and operational
- Health policy, health economics
- Health care admin and policy
(N = 5)

Ver 2: Small Amount of Space
- Public health
- International health
- (blank)
- Health insurance
- Policy
(N = 6)

Total Word Count = 25
Mean Word Count = 5.0
Cohen's $d = 2.62$
$t(9) = 4.63, p < .001$

Principle #2:
The questions guide the answers.

$N = 91$ Faculty

% Answer = Fire Professor

Pearson $\chi^2(1) = 4.90, p < .05$
**Survey Design: 7-Step Process**

- **Step 1**: Literature Review
- **Step 2**: Interviews & Focus Groups
- **Step 3**: Synthesize
- **Step 4**: Develop Items
- **Step 5**: Expert Validation
- **Step 6**: Cognitive Interviewing
- **Step 7**: Pilot Test
Step 1: Literature Review

Goal: Ensure the construct is relevant in the field

- Critically evaluate the literature
  - How is the construct defined in prior studies?
  - Has the construct been evaluated sufficiently?

- Identify existing scales
  - What items/scales currently exist?
  - Appraise quality

Step 2: Interviews & Focus Groups

Goal: Ensure construct is what “real” people experience

- Interview experts
- Create focus groups from target population

- Apply open-ended questions
  - Avoid yes/no, multiple-choice questions
Step 2: Interviews & Focus Groups

Example: Basic Clinical Skills Self-Efficacy

- **Interview experts**
  - How do “experts” define basic clinical skills self-efficacy?
    - Experienced medical educators
    - Medical education researchers who have studied self-efficacy

- **Focus groups from the target population**
  - How does the target population understand basic clinical skills self-efficacy?
    - Current students and recent graduates

Step 2: Practical exercise

**Group discussion:**
- If you wanted to create a survey to measure…

  *Basic Clinical Skills Self-Efficacy*

- What aspects of this “fuzzy” construct would you need to include/address/ask about?
Step 3: Synthesize Literature & Interviews

Goal: Arrive at consensus/agreement

- Literature
- Target Population
- Experts

Step 4: Develop Items

Goal: Develop items using vocabulary your target population can understand

- Considerations
  - Vocabulary and wording
  - Response anchor selection
    - Ratings vs. rankings; Likert-scale items; yes/no items?
  - Item formatting
    - Visual design, item order, instructions, etc.
Step 4: Develop Items (examples)

Course Importance (a belief; the full scale = 6 items)
1. How important was it for you personally to perform well in this course?
2. How important were the practical applications of the information provided in this course?
3. How important was the content of this course?
4. How important was it for you to learn the material in this course?

response anchors

<table>
<thead>
<tr>
<th>not at all important</th>
<th>slightly important</th>
<th>moderately important</th>
<th>quite important</th>
<th>extremely important</th>
</tr>
</thead>
</table>

Perceived Barriers to Mental Health Care
(a belief; the full scale = 12 items)

Rate each of the possible concerns that might affect your decision to receive mental health counseling or services if you ever had a problem:
1. I don't trust mental health professionals.
2. I don't know where to get help.
3. I don't have adequate transportation.

response anchors

<table>
<thead>
<tr>
<th>strongly disagree</th>
<th>disagree</th>
<th>neutral</th>
<th>agree</th>
<th>strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

(from Hoge et al., 2004)
Principle #3: A survey is a conversation between you and your respondents.

4) To what extent do you favor or oppose the university’s ‘Maximally Accessible Materials’ (MAM) policy to make all printed materials at the school available upon request in enlarged font form for the visually impaired?

<table>
<thead>
<tr>
<th></th>
<th>Strongly Oppose</th>
<th>Moderately Oppose</th>
<th>Slightly Oppose</th>
<th>Neither Favor nor Oppose</th>
<th>Slightly Favor</th>
<th>Moderately Favor</th>
<th>Strongly Favor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Total N = 17
Mean = 5.1

8 said “neither favor nor oppose”
6 said “slightly favor” or “moderately favor”
3 said “strongly favor”!!

NO MISSING DATA!!!!!

Principle #3: A survey is a conversation between you and your respondents.

A note about providing a reason

“because…” or “so that…”

8) As some of you may know, the university is debating whether to move some parts of the university to a new section of campus in Rockville. Do you think the university should move to Rockville?

Yes = 12.5%

No

8) As some of you may know, the university is debating whether to move some parts of the university to a new section of campus in Rockville. Do you think the university should move to Rockville so that the school can have more space?

Yes = 55.6%

No

Pearson $\chi^2(1) = 3.44, p = .06$
Today’s Small-Group Activity

• Form groups

• Using a construct definition, develop a scale of 5 items to measure the construct

**Construct = Basic Clinical Skills Self-Efficacy**

• Definition:
  ◦ A student’s confidence in his/her ability to perform the basic clinical skills expected of a graduating medical student.

Today’s Small-Group Activity

• **Construct:**

  **Basic Clinical Skills Self-Efficacy**

  • Definition:
    A student’s confidence in his/her ability to perform the basic clinical skills expected of a graduating medical student.

  • **Tell us your items:**

    1. ________________________________
    2. ________________________________
    3. ________________________________
    4. ________________________________
    5. ________________________________

  • **Tell us your response anchors:**

Common Pitfalls

• Creating double-barreled items
  ◦ Example Item: “How effective was the classroom and hands-on instruction?”
    • What if one is good and the other is bad?
  ◦ Solution: split into two items
    • “How effective was the classroom instruction?”
    • “How effective was the hands-on instruction?”

Common Pitfalls

• Creating double-barreled items
  ◦ Construct = Elaboration
  ◦ Item 40. When I study for this course, I write brief summaries of the main ideas from the readings and online discussions
    ◦ Cronbach’s alpha = 0.546
Common Pitfalls

- **Creating negatively worded items**
  - Unnecessary cognitive burden
  - Promotes satisficing
    - “In an average week, how often are you unable to start rounds on time?” (rarely-often)
    - “I can’t stop thinking about the war in Afghanistan” (rarely-often)
  - Solution: make sure “yes” means “yes” and “no” means “no”
    - “In an average week, how often do you start rounds on time?”
  - Afghanistan:
    - Slight Improvement: “I keep thinking about the war in Afghanistan”
    - Better: “How often do you think about the war in Afghanistan?”

- **Using statements instead of questions**
  - Example Item: “I am confident I can do well on this rotation.”
  - Statements are not very “conversational”
    - People are more practiced at answering questions
  - Solution: use questions
    - “How confident are you that you can do well in this rotation?”
    - Use construct-specific confidence response anchors
Common Pitfalls

- **Using agreement response anchors**
  - Example Item: “The global economy is the most important issue in the world today.”
  - Agreement response anchors don’t just measure the construct of interest
    - Confounded by how “agreeable” respondents are
    - Respondents often “agree” just because
  - Solution: avoid agreement response anchors; maintain focus on construct by using construct-specific anchors
    - “How important is the issue of the global economy in the world today?”

Common Pitfalls

- **What does it mean to “strongly agree” anyway?**

  **Section II:** In this section, each question will ask you to indicate how you understand a commonly used phrase by marking an “X” at the appropriate place on the line.

  25) When you say that you “strongly agree” with somebody else, what do you mean? Indicate on the line below where “strongly agree” is by marking an “X” on the line.
Strongly Disagreeable Ranges

Common Pitfalls

- **Using too few or too many response anchors**
  - Influences reliability within a set of survey items
    - Too few (<4) → less reliable
    - Too many (>7-9) → diminishing return; false impression of precision
  - Example Item: “How useful was the rotation in emergency medicine?”

<table>
<thead>
<tr>
<th>Not at all useful</th>
<th>Moderately useful</th>
<th>Very useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all useful</td>
<td>Slightly useful</td>
<td>Moderately useful</td>
</tr>
<tr>
<td>Not at all useful</td>
<td>Minimally useful</td>
<td>Slightly useful</td>
</tr>
</tbody>
</table>
Common Pitfalls

- **Using jargon and/or unclear language**
  - Instead, use language that is simple, direct, comprehensible, and unambiguous
  - Instead of *exhausted*... consider *tired*
  - Instead of *leisure time*... consider *free time*
  - Instead of *due to the fact that*... consider *because*
  - Instead of *at this point in time*... consider *now*

Today’s Small-Group Activity

Debrief

- **Construct = Basic Clinical Skills Self-Efficacy**

  **Definition:**
  - A student’s confidence in his/her ability to perform the basic clinical skills expected of a graduating medical student.

  **Tell us your items:**
  1. ____________________________
  2. ____________________________
  3. ____________________________
  4. ____________________________
  5. ____________________________

  **Tell us your response anchors:**
  [ ] [ ] [ ] [ ] [ ] [ ]
Survey Design: 7-Step Process

- Step 1: Literature Review
- Step 2: Interviews & Focus Groups
- Step 3: Synthesize
- Step 4: Develop Items
- **Step 5: Expert Validation**
- **Step 6: Cognitive Interviewing**
- **Step 7: Pilot Test**

Step 5: Expert Validation (aka, content validation)

**Goal:** Make sure the items “ring true” to experts

- Depending on your needs, experts can consider the following for each of your survey items…
  - Clarity
  - Construct relevance
  - Language level
  - Missing facets/aspects
Step 6: Cognitive Interviewing

**Goal:** Make sure respondents understand the items as intended by you (the developer)

- Recruit members of the targeted population
  - e.g., students, teachers, patients, locals, etc.
- Conduct one-on-one interviews, in “laboratory” or other location
- THEN: Make informed decisions, with cognitive interview as one source of input

---

Example

“In the last year, have you been bothered by pain in the abdomen?”

- What problems do you anticipate?
  - What time period are you thinking about?
  - What does “bothered by pain” mean to you?
  - Where is your “abdomen?”
Step 6: Cognitive Interviewing

Example

“Please look at this diagram. During the past 12 months, have you had pain in this area (the area shaded on the diagram)?

Step 7: Pilot Testing

**Goal:** Collect evidence of survey's reliability (score reproducibility) and validity (are you measuring what you intend to measure?)

- Collect data from a small sample
- “Get to know” your descriptive statistics
  - Are the item scores normally distributed?
Step 7: Pilot Testing

- “Get to know” your descriptive statistics
  - Do individual survey items “hang together”?
    - Factor analysis and reliability analysis
    - Scale #1
      - Item 1
      - Item 4
      - Item 7
      - Item 8
      - Item 10
      - Item 15

  - calculate a mean score

Step 7: Pilot Testing

- How does your scale score relate to other variables?
  - Dimensions of Student Motivation

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Test Value</th>
<th>Cronbach's Alpha</th>
<th>Extroverted</th>
<th>Anxious</th>
<th>Direction</th>
<th>Age</th>
<th>Course</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Cronbach's Alpha</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>N</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.05 level (2-tailed).

*Correlation is significant at the 0.01 level (2-tailed).
Final Example

19) Parking is the most important issue at the university today.

19) How important is the issue of parking at the university today?

Cohen’s $d = 1.16$
$t(15) = 2.31, p < .05$

Questions?

If you remember nothing else, remember…

**Principle #1:** You can’t fix by analysis what you’ve bungled by design.

**Principle #2:** The questions guide the answers.

**Principle #3:** A survey is a conversation between you and your respondents.

Anthony Artino, PhD  Jeffrey La Rochelle, MD, MPH
anthony.artino@usuhs.edu  jeffrey.la-rochelle@usuhs.edu

The presenters are U.S. Government employees. The views expressed in this presentation are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, Department of Defense, nor the U.S. Government.
AM Last Page: Survey Development Guidance for Medical Education Researchers

Hunter Gehlbach, PhD, assistant professor of Education, Harvard University; Anthony R. Artino, Jr, PhD, assistant professor of Preventive Medicine and Biometrics, Uniformed Services University of the Health Sciences; and Steven J. Durning, MD, professor of Medicine, Uniformed Services University of the Health Sciences.

Medical education researchers frequently rely on survey data. For example, of Academic Medicine’s 141 research articles from 2009, over half (56%) used surveys. Yet, the literature provides limited guidance on which processes best facilitate the development of surveys—particularly in the design of survey scales (i.e., several items that assess a single underlying construct such as physician empathy or teaching self-efficacy; see example below). This flowchart presents seven steps to facilitate the construction of valid and reliable survey scales.

**Step 1**
Conduct a **literature review** both to ensure that your construct definition aligns with relevant prior research and to identify extant survey scales or items that might be used or adapted for your research context.

**Step 2**
Conduct **interviews** and/or **focus groups** to learn how your population of interest conceptualizes and describes your construct of interest.

**Step 3**
**Synthesize** the literature review and interview/focus group data so that the conceptualization of the construct makes theoretical sense to scholars in the field and uses language that your population of interest understands. For example, a scale assessing teaching self-efficacy (i.e., confidence in one’s teaching ability) should use words like “confidence in trying out new teaching techniques,” not “efficaciousness in experimenting with novel pedagogies.”

**Step 4**
**Develop items** in accordance with current best practices in survey design.\(^1\),\(^2\) For example, the sample scale below uses response anchors that refer to the specific construct (rather than numbers or agree/disagree response anchors).

**Step 5**
Conduct an **expert validation** to assess the items’ clarity and relevance to the construct.\(^3\),\(^4\)*

**Step 6**
Conduct cognitive **pretesting** through which participants restate each item aloud in their own words as they answer it. This step helps ensure that respondents interpret items in the manner that you intend.\(^5\),\(^6\)*

**Step 7**
**Pilot-test** your items to check for adequate item variance, reliability, and convergent/discriminant validity with respect to other measures.\(^7\)*

*Note: After you complete each of these final steps, you may need to revise items and/or repeat steps from this part of the process.*

---

**Sample Items From a Teaching Self-Efficacy Scale**

1. How confident are you that you can help students remember what they learned in your class?
2. When you need to teach less interesting topics, how confident are you that you can keep all students engaged?
3. How confident are you that you can help students learn when they are unmotivated?
4. How confident are you that you can get through to the most difficult students?

S-5 point, Likert-type response scale:

<table>
<thead>
<tr>
<th>Not at all confident</th>
<th>Slightly confident</th>
<th>Moderately confident</th>
<th>Quite confident</th>
<th>Extremely confident</th>
</tr>
</thead>
</table>

---

**References**

### AM Last Page: Avoiding Five Common Pitfalls of Survey Design

Anthony R. Artino, Jr, PhD, assistant professor of preventive medicine and biometrics, Uniformed Services University of the Health Sciences, Hunter Gehlbach, PhD, assistant professor of education, Harvard University, and Steven J. Durning, MD, professor of medicine and pathology, Uniformed Services University of the Health Sciences

Writing good survey items is both an art and a science. Over the last 30 years, scholars have amassed a great deal of scientific evidence on which questionnaire designers can rely.1-5 The guidelines below present some of the more frequently ignored, but more important, of these survey-design basics.

<table>
<thead>
<tr>
<th>Pitfall</th>
<th>Survey example(s)</th>
<th>Why it’s a problem</th>
<th>Solution(s)</th>
<th>Survey example(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creating a double-barreled item</td>
<td>How often do you talk to your nurses and administrative staff when you have a problem?</td>
<td>Respondents have trouble answering survey items that contain more than one question (and thus could have more than one answer).1 In this example, respondents who talk to nurses often talk to administrative staff infrequently will struggle to answer this question. Survey items should address one idea at a time.1</td>
<td>When you have multiple questions/preambles within a given item, either (1) create multiple items for each question that is important or (2) include only the more important question. Be especially wary of conjunctions in your items.1,4</td>
<td>How often do you talk to your nurses when you have a problem? How often do you talk to your administrative staff when you have a problem?</td>
</tr>
<tr>
<td>Creating a negatively worded item</td>
<td>In an average week, how many times are you unable to start class on time? The chief resident should not be responsible for denying admission to patients.</td>
<td>Negatively worded survey items are challenging for respondents to comprehend and answer accurately. Double-negatives are particularly problematic and increase measurement error.1 If a respondent has to say “yes” in order to mean “no” (or “agree” in order to “disagree”), the item is flawed. Make sure “yes” means yes and “no” means no. This generally means wording items positively.1</td>
<td>In an average week, how many times do you start class on time? Should the chief resident be responsible for admitting patients?</td>
<td></td>
</tr>
<tr>
<td>Using statements instead of questions</td>
<td>I am confident I can do well in this course. • not at all true • a little bit true • somewhat true • mostly true • completely true</td>
<td>A survey represents a conversation between the surveyor and the respondents. To make sense of survey items, respondents rely on “the tacit assumptions that govern the conduct of conversation in everyday life.”2 Only rarely do people engage in rating statements in their everyday conversations. Formulate survey items as questions. Questions are more conversational, more straightforward, and easier to process mentally. People are more practiced at responding to them.1,4</td>
<td>How confident are you that you can do well in this course? • not at all confident • slightly confident • moderately confident • quite confident • extremely confident</td>
<td></td>
</tr>
<tr>
<td>Using agreement response anchors</td>
<td>The high cost of health care is the most important issue in America today. • strongly disagree • disagree • neutral • agree • strongly agree</td>
<td>Agreement response anchors do not emphasize the construct being measured and are prone to acquiescence (i.e., the tendency to endorse any assertion made in an item, regardless of its content).3 In addition, agreement response anchors may encourage respondents to think through their responses less thoroughly while completing the survey.4 Use construct-specific response anchors that emphasize the construct of interest. Doing so reduces acquiescence and keeps respondents focused on the construct in question. Doing so results in less measurement error.1,4</td>
<td>How important is the issue of high health care costs in America today? • not at all important • slightly important • moderately important • quite important • extremely important</td>
<td></td>
</tr>
<tr>
<td>Using too few or too many response anchors</td>
<td>How useful was your medical school training in clinical decision making? • not at all useful • somewhat useful • very useful</td>
<td>The number of response anchors influences the reliability of a set of survey items.5 Using too few response anchors generally reduces reliability. There is, however, a point of diminishing returns beyond which more response anchors do not enhance reliability.5 Use five or more response anchors to achieve stable participant responses. In most cases, using more than seven to nine anchors is unlikely to be meaningful to most respondents and will not improve reliability.5</td>
<td>How useful was your medical school training in clinical decision making? • not at all useful • slightly useful • moderately useful • quite useful • extremely useful</td>
<td></td>
</tr>
</tbody>
</table>

**References:**

**Disclaimers:**
The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the Department of Defense. Dr. Steven Durning coauthored this Last Page prior to becoming assistant editor. AM Last Page.
AM Last Page: Avoiding Four Visual-Design Pitfalls in Survey Development

Anthony R. Artino, Jr., PhD, associate professor, Preventive Medicine and Biometrics, Uniformed Services University of the Health Sciences, and Hunter Gehlbach, PhD, associate professor, Harvard Graduate School of Education

A previous AM Last Page presented five common pitfalls of survey design as well as several solutions. This AM Last Page presents four visual-design and layout pitfalls and offers solutions.

### Pitfall: Explanation and Example

<table>
<thead>
<tr>
<th>Labeling only the end points of your response options</th>
<th>Solution: Explanation and Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeling only the end points leaves the meaning of the unlabeled options open to respondents' interpretation. Different respondents can interpret the unlabeled options differently. This ambiguity increases measurement error.</td>
<td>Verbally label each response option&lt;br&gt;Labeling each response option increases consistency in the conceptual spacing between response options and increases the likelihood that all respondents will interpret the response options similarly. Additionally, the visual weight of the response options are comparable, so that respondents' eyes are not drawn to certain options.</td>
</tr>
</tbody>
</table>

### Solution: Explanation and Example

**How interesting did you find this clinical reasoning course?**

<table>
<thead>
<tr>
<th>not at all interesting</th>
<th>slightly interesting</th>
<th>moderately interesting</th>
<th>quite interesting</th>
<th>extremely interesting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Labeling response options with both numbers and verbal labels

Because of the additional information respondents must process, providing both numbers and verbal labels extends response time. The implied meaning of negative numbers can be particularly confusing, and may introduce additional error. For example, in the item below, learning "a little bit" seems incongruous with learning the amount of "-1."

**How much did you learn in today’s workshop?**

<table>
<thead>
<tr>
<th>almost nothing</th>
<th>a little bit</th>
<th>some</th>
<th>quite a bit</th>
<th>a tremendous amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

### Unequally spacing your response options

The visual spacing between options can attract respondents to certain options over others, which in turn might cause them to select these options more frequently. In addition, unbalanced spacing of the response options can shift the visual midpoint of the scale.

**How much did you learn from your peers in this course?**

<table>
<thead>
<tr>
<th>almost nothing</th>
<th>a little bit</th>
<th>some</th>
<th>quite a bit</th>
<th>a tremendous amount</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Placing non-substantive response options together with substantive response options

Placing non-substantive response options such as “don’t know,” “no opinion,” or “not applicable” together with the substantive options can shift the visual and conceptual midpoint of the response scales, thereby skewing the results.

**How satisfied are you with the quality of the library services?**

<table>
<thead>
<tr>
<th>not at all satisfied</th>
<th>slightly satisfied</th>
<th>moderately satisfied</th>
<th>quite satisfied</th>
<th>extremely satisfied</th>
<th>not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Use additional space to visually separate non-substantive response options

Using additional space to visually separate non-substantive response options from the substantive options will align the visual midpoint with the conceptual midpoint thereby reducing measurement error. This recommendation is an important exception to the guidance above about maintaining equal spacing between response options.

**How satisfied are you with the quality of the library services?**

<table>
<thead>
<tr>
<th>not at all satisfied</th>
<th>slightly satisfied</th>
<th>moderately satisfied</th>
<th>quite satisfied</th>
<th>extremely satisfied</th>
<th>not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the U.S. Department of Defense.

### References


Author contact: Anthony.Artino@usuhs.mil
Reliability is the extent to which the scores produced by a particular measurement tool or procedure are consistent and reproducible. Reliability answers the question, “Does the assessment yield the same scores at different times, from different raters, or from different items?”

Validity is the degree to which an assessment measures what investigators want to measure, all of what they want to measure, and nothing but what they want to measure. Validity answers the question, “Does the assessment provide information that is relevant to the inferences that are being made from it?” An assessment, such as a test or questionnaire, does not have validity in any absolute sense. Instead, the scores produced are valid for some uses and not valid for others.

A target provides a metaphor for the relationship between reliability and validity. The true score (or value) for the concept the researcher is attempting to measure is at the center of the target, and the observed score the investigator gets from each person assessed is a shot at the target. Reliability is a necessary but insufficient condition for validity. To be valid, scores must first be at least moderately reliable. However, scores that are reliable may be devoid of validity for the application the researcher has in mind.

Many methods of assessing reliability and validity are available. Each method provides the researcher with slightly different information about the reliability and validity of the assessment.

**Assessing reliability**
- Test-retest
- Equivalent forms
- Interrater
- Internal consistency
  - Split-half
  - Kuder-Richardson
  - Cronbach alpha

**Assessing validity**
- Construct
  - Convergent
  - Discriminant
  - Known-groups
- Criterion-related
  - Predictive
  - Concurrent
  - Postdictive
- Content-related
  - Content
  - Face

References
Survey Development: What Not to Avoid

Surveys are a commonly employed research design method. Developing an effective survey depends on the adequacy of construct development and attention to sampling and design, item construction, data processing, pilot testing, and response rate (Figure 1). The focus of this article is to address construct development, expert validation, cognitive pre-testing, and pilot testing, all of which are critical to ensuring reliability and validity of the data collected.

**Construct Development**

Thorough development of the construct to be measured by items within the survey is an essential first step in survey development. A construct is an abstract concept or idea that is typically not directly measurable or observable (e.g. patient satisfaction or student motivation). Most constructs are not readily assessed using a single survey question. Instead, it is often necessary to create a series of items, referred to as

| **FIGURE 1:** Systematic Eight-Step Process for Survey Development (1) |
|-----------------|---------------------------------------------------------------|
| **Development Step** | **Purpose** |
| 1. Articulate a research question and define the construct(s) of interest | Determine whether or not it makes sense to measure your construct with a survey |
| 2. Conduct a thorough review of the literature | Make sure your construct is relevant to the field of study and coheres with prior research |
| | Determine how the construct is defined by scholars in the field of study and identify existing surveys/items that might be used or adapted to measure your construct |
| 3. Conduct interviews and/or focus groups | Make sure your construct has face validity and is relevant to what your participants experience |
| | Learn how your participants conceptualize and describe your construct |
| 4. Synthesize the data from the literature review and the interviews/focus groups | Make sure your conception of the construct is agreed upon by academics and participants |
| | Ensure the language you plan to use makes theoretical sense yet is understandable to participants |
| 5. Draft a set of survey items | Develop survey items using best practices in survey design |
| | Make sure you have good construct coverage using the vocabulary of participants |
| | Include *extra* items at this stage, as items will get revised and deleted during the next three steps of the survey development process |
| 6. Conduct an expert validation | Make sure the items ring true to experts |
| | Ask experts to consider the clarity of the items, the relevance of items to the construct, the "difficulty" of the items, and the overall representativeness of the construct |
| | Revise items as necessary, based on the expert validation results |
| 7. Conduct several cognitive interviews | Make sure participants understand the items as intended by you (the developer) |
| | Assess the extent to which different participants interpret all the items in the same way and can respond to them accurately |
| | Revise items as necessary, based on the cognitive interview results |
| 8. Pilot test the survey with a small sample of participants | Make sure the items developed have appropriate range and variance |
| | Ensure the scores obtained have adequate internal consistency reliability |
| | Create a composite score for each individual scale and create an inter-scale correlation matrix to demonstrate the overall validity of the survey in measuring the construct(s) of interest |
| | Revise items as necessary, based on the pilot test results |
a scale, to describe the construct. For example, if patient satisfaction with a clinic is measured, it is reasonable to create survey items specific for provider communication, patient trust in provider, and clinic access. Combining these individual items into a scale or series of scales provides a more robust representation of the multi-dimensional nature of an abstract construct like patient satisfaction. In addition, the individual scales may obviate problem areas in specific domains that adversely impact the overall construct of patient satisfaction, thereby creating actionable items not otherwise captured without the use of scales. Literature review, focus groups, and interviews with experts and the population of interest are among the approaches commonly used during construct and scale development.

Expert Validation

Once the construct has been developed and items for the scales written, the next step is expert validation—a formal way of gathering information about a developing survey from experts in the field of interest (2-3). This process involves expert review of each individual survey item using a standard form provided by the survey developer. The standard validation form usually addresses major topics of interest related to the clarity (i.e., whether there are ambiguities or multiple ways to interpret the question or response options), the relevance of items to the specific scale and construct (i.e., the extent to which each item relates to the aspect of the construct that the item is intended to measure), the overall representativeness of the construct (i.e., how completely the items cover the scale and construct), and the “difficulty” of the items. The “difficulty” of an item refers to the extent to which respondents have a hard time endorsing the item.

For example, the average student may find it difficult to strongly endorse the self-confidence item, “I’m confident I can get 100% of the points in biochemistry,” but the same student may find it easier to strongly endorse the item, “I’m confident I can pass biochemistry.” For any given scale, the survey developer should strive to have a range of items with varying levels of difficulty (4). During the process of expert validation, opportunities to improve items, to generate new items that better represent a particular scale, and to identify a previously overlooked dimension of the construct can become apparent.

Cognitive Pretesting

After the experts have an opportunity to refine the survey, it is equally important to understand how the study population will interpret the items and response choices through a process known as cognitive pre-testing or cognitive interviewing (5). On an ideal survey, respondents will interpret all the items the same way as intended by the survey designer. In essence, cognitive pre-testing is a modified pilot test on a small group (five to 15) from the study population using a qualitative approach to elucidate problems with specific questions or responses due to misinterpretations, assumptions, bias, and formatting. Typically, this process involves a face-to-face, scripted interview where a respondent reads each question aloud and uses a “think-aloud” process in determining their response to the question. During this process, survey developers uncover unexpected problems with items and seek input to understand the nature of the issue and a potential solution.

Construct development, expert validation, and cognitive testing add critical information to create more reliable and valid data.

Pilot Testing

Once the items and scales have been modified, the next step is a pilot test of the survey. During pilot testing, members of the target population are asked to complete the survey in the planned delivery format (e.g., paper, web-based, etc.). The obtained data are used to ensure an appropriate range of responses from each item without a skew to one of the extremes. An internal consistency reliability analysis for items within each scale can identify individual items not functioning as expected and may reveal common errors (e.g., questions that address one or more ideas, so-called “double-barreled” questions). A correlation matrix on items within each scale can identify redundant items (items that are too highly correlated) or items that may be unrelated to the other items in the scale. Both of these analyses can cull down the number of overall survey items while still maintaining a high degree of reliability. If there are enough respondents in the pilot study, a factor analysis may provide further validity evidence for the appropriateness of grouping items into specific scales.

In addition to an analysis of individual items, a composite score can also be calculated from each individual scale. These composite scores are then used to create an inter-scale correlation matrix to demonstrate the overall validity of the survey in measuring the construct (or constructs) of interest. This technique will uncover areas of convergent and discriminant validity within the survey tool. For example, if measuring the multi-dimensional construct of student motivation, one might expect the interest dimension would be positively correlated with the enjoyment dimension, and
Survey Development: What Not to Avoid

The enjoyment dimension would be negatively correlated with the anxiety dimension. These basic statistical techniques for both individual items and scales provide valuable information for further refinement of the survey, but as with any process, the psychometric data must be balanced with the underlying theory behind the constructs being measured in the survey.

The processes of construct development, expert validation, cognitive pre-testing, and pilot testing are too often overlooked in the development of survey tools, and yet each step adds critical information toward creating more reliable and valid data. Given the large number of research questions answered through surveys, greater attention to these techniques is worthwhile so that data obtained through survey research methods is of use.

Authors

Jeff LaRochelle, MD
Assistant Professor
Department of Medicine
F. Edward Hébert School of Medicine
Uniformed Services University of the Health Sciences

Andrew R. Hoellein, MD
Assistant Professor
Department of Medicine
University of Kentucky College of Medicine

Liselotte N. Dyrbe, MD
Associate Professor
Department of Medicine
Mayo Clinic College of Medicine

Anthony R. Artino, Jr., PhD
Assistant Professor
Department of Medicine
F. Edward Hébert School of Medicine
Uniformed Services University of the Health Sciences

References


Incorporating Social Media into Medical Education

Transparent about their practices. Ultimately, innovating with social media tools can enhance both the teaching and learning experience for medical educators and medical students. We are bound only by the limits of creativity.

Authors

Carrie Saarinen
Instructional Designer
Brown University

Vineet Arora, MD
Associate Program Director
Internal Medicine Residency Program
University of Chicago Pritzker School of Medicine

Benjamin Ferguson
MD/PhD Candidate
University of Chicago Pritzker School of Medicine

Katherine Chretien, MD
Medicine Clerkship Director
Washington DC VA Medical Center

If you would like to suggest a theme or a list for Ten Tunes, please e-mail Insight Editor Sheila T. Costa at scosta@im.org.
**Survey Design**  
Glossary of Terms

**Acquiescence** is the tendency to endorse any assertion made in a question, regardless of its content (this is really a type of satisficing; see definition below).

**Construct** is a hypothesized concept, model, idea, or theory (something constructed) that we think exists but that we cannot directly observe.

**Content validity** is evidence obtained from an analysis of the relationship between a survey instrument’s content and the construct it is intended to measure.

**Factor** is an “unobserved” variable that statistically explains the variation and co-variation among a larger set of “observed” variables (i.e., the actual items on a survey). Stated another way, factors succinctly represent a larger set of observed variables. Factors often correspond to constructs; although some constructs are made up of multiple factors. Such constructs are often called multi-dimensional constructs.

**Factor analysis** is an analytical technique used to identify factors that statistically explain the variation and co-variation among a set of measures (i.e., a set of survey items). Factor analysis is a data-reduction technique that reduces a large number of overlapping measured variables to a much smaller set of factors.

**Items/Indicators** (observable items, empirical indicators) are the actual questions/statements that make up a survey (or a particular survey scale).

**Optimizing** is the extent to which a respondent performs the necessary cognitive tasks to answer a survey item in a thorough and unbiased manner. These cognitive tasks may include: (1) interpreting a survey item (figuring out its intent), (2) searching memory for relevant information, (3) forming a judgment, and (4) translating the judgment into an answer by summarizing or selecting one of the alternatives offered. These are the tasks we want respondents to do when taking our survey.

**Order effect** is the notion that the order of response alternatives affects the extent to which respondents select those items (primary and recency effects are two types of order effects).

**Primacy effect** is the tendency to remember (and select) answers that appear first (or early) in a list of alternatives (likely because those items were cognitively processes and now reside in long-term member). This effect is more prominent when items are presented visually.

**Recency effect** is the tendency to remember (and select) answers that appear last (or later) in a list of alternatives (likely because they still reside in working memory and so are more accessible). This effect is more prominent when items are presented orally.

**Reliability** is the extent to which the scores produced by a particular measurement procedure or instrument (e.g., a survey) are consistent and reproducible. Reliability is a necessary but insufficient condition for validity.

**Response anchors** are the named points along a set of answer options (e.g., strongly disagree, disagree, neutral, agree, strongly agree).
Response process validity is evidence obtained from an analysis of how respondents interpret the meaning of a survey and its specific survey items.

Satisficing is the extent to which respondents compromise their standards and expend less energy (i.e., they don’t fully optimize).

Scale is two or more items (indicators) intended to measure a construct. Often, however, the word scale is used more generally to refer to the entire survey. As such, many scales are composed of several sub-scales.

Social desirability bias is the tendency to over-report admirable attitudes/behaviors and under-report those that are not socially respected. Stated another way, it is the tendency to lie in order to appear as socially suitable and acceptable as possible.

Strong satisficing is a more dramatic form of satisficing where respondents skip entire cognitive tasks (i.e., comprehension, retrieval, judgment, or response selection) and arbitrarily select an answer (e.g., they may select the first reasonable response; they may accept any assertions made that seem reasonable; they may select “don’t know” or “no opinion” to avoid expending effort; they may randomly select a response from those offered).

Sub-scale is a sub-division of a larger scale. Often, multi-dimensional constructs will be measured with a scale that is made up of several smaller sub-scales.

Weak satisficing is a less serious form of satisficing where respondents are less thorough in comprehension, retrieval, judgment, and response selection (e.g., they may be less thoughtful about a question’s meaning; they may search their memories less comprehensively; they may integrate retrieved information carelessly; they may select a response imprecisely).

Validity is the degree to which evidence and theory support a measure’s intended use.

Validity argument is the process of accumulating evidence to provide a sound scientific basis for the proposed uses of an instrument’s scores.
<table>
<thead>
<tr>
<th>Construct-Specific Response Scales</th>
<th>Construct:</th>
</tr>
</thead>
<tbody>
<tr>
<td>not important</td>
<td>importance</td>
</tr>
<tr>
<td>somewhat important</td>
<td>importance</td>
</tr>
<tr>
<td>important</td>
<td>importance</td>
</tr>
<tr>
<td>very important</td>
<td>importance</td>
</tr>
<tr>
<td>extremely important</td>
<td>importance</td>
</tr>
<tr>
<td>unimportant</td>
<td>importance</td>
</tr>
<tr>
<td>of little importance</td>
<td>importance</td>
</tr>
<tr>
<td>moderately important</td>
<td>importance</td>
</tr>
<tr>
<td>important</td>
<td>importance</td>
</tr>
<tr>
<td>very important</td>
<td>importance</td>
</tr>
<tr>
<td>not at all important</td>
<td>importance</td>
</tr>
<tr>
<td>slightly important</td>
<td>importance</td>
</tr>
<tr>
<td>moderately important</td>
<td>importance</td>
</tr>
<tr>
<td>quite important</td>
<td>importance</td>
</tr>
<tr>
<td>extremely important</td>
<td>importance</td>
</tr>
<tr>
<td>completely unimportant</td>
<td>importance</td>
</tr>
<tr>
<td>unimportant</td>
<td>importance</td>
</tr>
<tr>
<td>neutral</td>
<td>importance</td>
</tr>
<tr>
<td>important</td>
<td>importance</td>
</tr>
<tr>
<td>completely important</td>
<td>importance</td>
</tr>
<tr>
<td>not at all confident</td>
<td>self-efficacy (confidence)</td>
</tr>
<tr>
<td>slightly confident</td>
<td>satisfaction</td>
</tr>
<tr>
<td>moderately confident</td>
<td>satisfaction</td>
</tr>
<tr>
<td>quite confident</td>
<td>satisfaction</td>
</tr>
<tr>
<td>extremely confident</td>
<td>satisfaction</td>
</tr>
<tr>
<td>completely dissatisfied</td>
<td>satisfaction</td>
</tr>
<tr>
<td>moderately dissatisfied</td>
<td>satisfaction</td>
</tr>
<tr>
<td>neutral</td>
<td>satisfaction</td>
</tr>
<tr>
<td>moderately satisfied</td>
<td>satisfaction</td>
</tr>
<tr>
<td>completely satisfied</td>
<td>satisfaction</td>
</tr>
<tr>
<td>not at all satisfied</td>
<td>satisfaction</td>
</tr>
<tr>
<td>slightly satisfied</td>
<td>satisfaction</td>
</tr>
<tr>
<td>moderately satisfied</td>
<td>satisfaction</td>
</tr>
<tr>
<td>quite satisfied</td>
<td>satisfaction</td>
</tr>
<tr>
<td>extremely satisfied</td>
<td>satisfaction</td>
</tr>
<tr>
<td>not at all bored</td>
<td>boredom</td>
</tr>
<tr>
<td>slightly bored</td>
<td>boredom</td>
</tr>
<tr>
<td>moderately bored</td>
<td>boredom</td>
</tr>
<tr>
<td>quite bored</td>
<td>boredom</td>
</tr>
<tr>
<td>extremely bored</td>
<td>boredom</td>
</tr>
<tr>
<td>not at all frustrated</td>
<td>frustration</td>
</tr>
<tr>
<td>slightly frustrated</td>
<td>frustration</td>
</tr>
<tr>
<td>moderately frustrated</td>
<td>frustration</td>
</tr>
<tr>
<td>quite frustrated</td>
<td>frustration</td>
</tr>
<tr>
<td>extremely frustrated</td>
<td>frustration</td>
</tr>
<tr>
<td>strongly prefer x</td>
<td>comparing x to y</td>
</tr>
<tr>
<td>prefer x</td>
<td>comparing x to y</td>
</tr>
<tr>
<td>neutral</td>
<td>comparing x to y</td>
</tr>
<tr>
<td>prefer y</td>
<td>comparing x to y</td>
</tr>
<tr>
<td>strongly prefer y</td>
<td>comparing x to y</td>
</tr>
<tr>
<td>almost no effort</td>
<td>effort</td>
</tr>
<tr>
<td>a little bit of effort</td>
<td>effort</td>
</tr>
<tr>
<td>some effort</td>
<td>effort</td>
</tr>
<tr>
<td>quite a bit of effort</td>
<td>effort</td>
</tr>
<tr>
<td>a great deal of effort</td>
<td>effort</td>
</tr>
<tr>
<td>very poor</td>
<td>quality</td>
</tr>
<tr>
<td>poor</td>
<td>quality</td>
</tr>
<tr>
<td>barely acceptable</td>
<td>quality</td>
</tr>
<tr>
<td>good</td>
<td>quality</td>
</tr>
<tr>
<td>very good</td>
<td>quality</td>
</tr>
</tbody>
</table>
More General Response Scales

<table>
<thead>
<tr>
<th>Completely untrue</th>
<th>Somewhat untrue</th>
<th>Yes and no</th>
<th>Somewhat true</th>
<th>Completely true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all true of me</td>
<td>Slightly true of me</td>
<td>Somewhat true of me</td>
<td>Mostly true of me</td>
<td>Completely true of me</td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly agree</td>
</tr>
<tr>
<td>Or</td>
<td>Neither agree nor disagree</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disagree strongly</td>
<td>Disagree</td>
<td>Tend to disagree</td>
<td>Tend to agree</td>
<td>Agree</td>
</tr>
<tr>
<td>Disagree strongly</td>
<td>Disagree moderately</td>
<td>Disagree slightly</td>
<td>Agree slightly</td>
<td>Agree moderately</td>
</tr>
<tr>
<td>Disagree very strongly</td>
<td>Disagree strongly</td>
<td>Disagree</td>
<td>Agree</td>
<td>Agree strongly</td>
</tr>
<tr>
<td>Completely disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Completely agree</td>
</tr>
<tr>
<td>Completely disagree</td>
<td>Mostly disagree</td>
<td>Slightly disagree</td>
<td>Slightly agree</td>
<td>Mostly agree</td>
</tr>
</tbody>
</table>
### Frequency or "Degree" Response Scales

<table>
<thead>
<tr>
<th>almost never</th>
<th>once in a while</th>
<th>sometimes</th>
<th>often</th>
<th>almost all the time</th>
</tr>
</thead>
<tbody>
<tr>
<td>never</td>
<td>seldom</td>
<td>about half the time</td>
<td>usually</td>
<td>always</td>
</tr>
<tr>
<td>never</td>
<td>little</td>
<td>somewhat</td>
<td>much</td>
<td>a great deal</td>
</tr>
<tr>
<td>not at all</td>
<td>very little</td>
<td>moderately</td>
<td>quite a bit</td>
<td>a tremendous amount</td>
</tr>
<tr>
<td>never</td>
<td>rarely</td>
<td>occasionally</td>
<td>frequently</td>
<td>almost always</td>
</tr>
<tr>
<td>never</td>
<td>seldom</td>
<td>sometimes</td>
<td>often</td>
<td></td>
</tr>
<tr>
<td>never</td>
<td>rarely</td>
<td>sometimes</td>
<td>often</td>
<td>very often</td>
</tr>
<tr>
<td>seldom</td>
<td>occasionally</td>
<td>to a considerable degree</td>
<td>almost always</td>
<td></td>
</tr>
<tr>
<td>never</td>
<td>very rarely</td>
<td>rarely</td>
<td>occasionally</td>
<td>frequently</td>
</tr>
<tr>
<td>never</td>
<td>very rarely</td>
<td>rarely</td>
<td>occasionally</td>
<td>very frequently</td>
</tr>
</tbody>
</table>
Survey Development References

Good General Textbooks & Articles:


Articles on Expert Validation:


Articles on Cognitive Interviewing:


Articles on Reliability and Factor Analyses:


