How to Develop and Utilize Surveys in Research

Workshop Description:

Surveys are commonly used in health professions education. Unfortunately, few educators are familiar with the best practices of survey design. The purpose of this workshop is to provide health professions educators with an introduction to a systematic process for creating valid and reliable surveys that can be used as assessment or research tools.

Upon completion of the workshop, participants will be able to...

1) Recognize how to use a systematic, 7-step process as the framework for survey design;
2) Demonstrate how to develop an appropriate set of items (a survey “scale”) to characterize the educational construct being measured; and
3) Identify common item-writing pitfalls in survey design.

Presenters:

**CDR Anthony R. Artino, Jr.** is Associate Professor of Preventive Medicine and Biometrics at the Uniformed Services University of the Health Sciences (USU). Anthony received his Ph.D. in educational psychology from the University of Connecticut and has published several peer-reviewed articles on survey design in health professions education. He has recently presented several survey-design workshops at national and international meetings.

**LTC Kent J. DeZee** is the General Medicine Fellowship Director and Associate Professor of Medicine at USU. Kent received his MD from The Ohio State University and his MPH from USU. He has published numerous peer-reviewed articles using survey methodology and has recently presented several survey-design workshops at national and international meetings.
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. 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.

**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.

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

*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?

<table>
<thead>
<tr>
<th>5-point, Likert-type response scale:</th>
<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 but 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.</td>
<td>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>
</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.</td>
<td>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>
</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</td>
<td>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>
</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</td>
<td>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>
</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 of Preventive Medicine and Biometrics, Uniformed Services University of the Health Sciences, Bethesda, Maryland and Hunter Gehlbach, PhD, assistant professor of Education, Harvard University, Cambridge, Massachusetts

In a previous AM Last Page, five common pitfalls of survey design were presented and several solutions were provided. In this Last Page, four visual-design and layout pitfalls are presented, and more solutions are offered.

<table>
<thead>
<tr>
<th>Pitfall</th>
<th>Solution(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Labeling only the end points of your response options</td>
<td>Verbally label each response option. Doing so 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>
<tr>
<td>Problematic item:</td>
<td>Improved item:</td>
</tr>
<tr>
<td>How interesting did you find this clinical reasoning course?</td>
<td>How interesting did you find this clinical reasoning course?</td>
</tr>
<tr>
<td>not at all interesting</td>
<td>not at all interesting</td>
</tr>
<tr>
<td>extremely interesting</td>
<td>slightly interesting</td>
</tr>
<tr>
<td>2. Labeling response options with both numbers and verbal labels</td>
<td>In general, use only verbal labels for each response option. Doing so will reduce the cognitive effort required of your respondents and will likely reduce measurement error.</td>
</tr>
<tr>
<td>Problematic item:</td>
<td>Improved item:</td>
</tr>
<tr>
<td>How much did you learn in today’s workshop?</td>
<td>How much did you learn in today’s workshop?</td>
</tr>
<tr>
<td>almost nothing</td>
<td>almost nothing</td>
</tr>
<tr>
<td>-2 almost nothing</td>
<td>some</td>
</tr>
<tr>
<td>-1 a little bit</td>
<td>1 quite a bit</td>
</tr>
<tr>
<td>0 some</td>
<td>2 a tremendous amount</td>
</tr>
<tr>
<td>3. Unequally spacing your response options</td>
<td>Maintain equal spacing between response options. Doing so will reinforce the notion that, conceptually, there is equal space or “distance” between each response option. As a result, the answers will be less biased, thereby reducing measurement error.</td>
</tr>
<tr>
<td>Problematic item:</td>
<td>Improved item:</td>
</tr>
<tr>
<td>How much did you learn from your peers in this course?</td>
<td>How much did you learn from your peers in this course?</td>
</tr>
<tr>
<td>almost nothing</td>
<td>almost nothing</td>
</tr>
<tr>
<td>a little bit</td>
<td>a little bit</td>
</tr>
<tr>
<td>some</td>
<td>some</td>
</tr>
<tr>
<td>quite a bit</td>
<td>quite a bit</td>
</tr>
<tr>
<td>a tremendous amount</td>
<td>a tremendous amount</td>
</tr>
<tr>
<td>4. Placing non-substantive response options together with substantive response options</td>
<td>Use additional space to visually separate non-substantive response options from the substantive options. Doing so will align the visual midpoint with the conceptual midpoint thereby reducing measurement error. This recommendation is a beneficial exception to the guidance above about maintaining equal spacing between response options.</td>
</tr>
<tr>
<td>Problematic item:</td>
<td>Improved item:</td>
</tr>
<tr>
<td>How satisfied are you with the quality of the library services?</td>
<td>How satisfied are you with the quality of the library services?</td>
</tr>
<tr>
<td>not at all satisfied</td>
<td>not at all satisfied</td>
</tr>
<tr>
<td>slightly satisfied</td>
<td>slightly satisfied</td>
</tr>
<tr>
<td>moderately satisfied</td>
<td>moderately satisfied</td>
</tr>
<tr>
<td>quite satisfied</td>
<td>quite satisfied</td>
</tr>
<tr>
<td>extremely satisfied</td>
<td>extremely satisfied</td>
</tr>
<tr>
<td>applicable</td>
<td>applicable</td>
</tr>
</tbody>
</table>

References:
Reliability is the extent to which the scores produced by a particular measurement tool or procedure are consistent and reproducible.\(^1\) 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.\(^1\) 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.\(^1\) However, scores that are reliable may be devoid of validity for the application the researcher has in mind.\(^1\)

Many methods of assessing reliability and validity are available.\(^1\)–\(^4\) Each method provides the researcher with slightly different information about the reliability and validity of the assessment.

**Assessing reliability**

- **Test–retest**
  - Assesses agreement between the same assessment given on two separate occasions.
- **Equivalent forms**
  - Assesses agreement between similar forms of an assessment given at separate times.
- **Interrater**
  - Assesses agreement between two or more coders or raters.
- **Internal consistency**
  - **Split-half**
  - **Kuder–Richardson**
  - **Cronbach alpha**
  - Assesses correlation between the different items on an assessment.

**Assessing validity**

- **Construct**
  - **Convergent**
  - **Discriminant**
  - **Known-groups**
- **Criterion-related**
  - **Predictive**
  - **Concurrent**
  - **Postdictive**
- **Content-related**
  - **Content**
  - **Face**
  - Assesses whether the assessment captures the concept. **Example:** Factor analysis reveals that the survey items “hang together” as expected, relate to a single factor, and are unrelated to another, different set of items.
  - Compares the assessment to a criterion that the researcher thinks is important. **Example:** MCAT scores should be related to medical school performance.
  - Assesses whether the items measure all the important aspects of the construct. **Example:** Items on an exam should assess all of the learning objectives.
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

<table>
<thead>
<tr>
<th>Development Step</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Articulate a research question and define the construct(s) of interest</td>
<td>Determine whether or not it makes sense to measure your construct with a survey</td>
</tr>
<tr>
<td>2. Conduct a thorough review of the literature</td>
<td>Make sure your construct is relevant to the field of study and coheres with prior research</td>
</tr>
<tr>
<td>3. Conduct interviews and/or focus groups</td>
<td>Make sure your construct has face validity and is relevant to what your participants experience</td>
</tr>
<tr>
<td>4. Synthesize the data from the literature review and the interviews/focus groups</td>
<td>Make sure your conception of the construct is agreed upon by academics and participants</td>
</tr>
<tr>
<td>5. Draft a set of survey items</td>
<td>Develop survey items using best practices in survey design</td>
</tr>
<tr>
<td>6. Conduct an expert validation</td>
<td>Make sure the items ring true to experts</td>
</tr>
<tr>
<td>7. Conduct several cognitive interviews</td>
<td>Make sure participants understand the items as intended by you (the developer)</td>
</tr>
<tr>
<td>8. Pilot test the survey with a small sample of participants</td>
<td>Make sure the items developed have appropriate range and variance</td>
</tr>
</tbody>
</table>
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 read each question aloud and use 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.

**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 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.

Continued on page 19
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 LaRochele, MD
Assistant Professor
Department of Medicine
F. Edward Hébert School of Medicine
Uniformed Services University of the Health Sciences

Andrew R. Hoelllein, 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
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).

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

**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 items 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).

**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.

**Response anchors (or response options)** are the named points along a response scale (e.g., strongly disagree, disagree, neutral, agree, strongly agree; never, once in a while, sometimes, often, almost all the time).

**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. Oftentimes, 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, though often unconscious, to lie in order to appear as socially suitable and acceptable as possible.

Sub-scale is a sub-division of a larger scale. Often times, 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).

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).
## Construct-Specific Response Scales

<table>
<thead>
<tr>
<th>not important</th>
<th>somewhat important</th>
<th>important</th>
<th>very important</th>
<th>extremely important</th>
</tr>
</thead>
<tbody>
<tr>
<td>unimportant</td>
<td>of little importance</td>
<td>moderately important</td>
<td>important</td>
<td>very important</td>
</tr>
<tr>
<td>not at all important</td>
<td>slightly important</td>
<td>moderately important</td>
<td>quite important</td>
<td>extremely important</td>
</tr>
<tr>
<td>completely unimportant</td>
<td>unimportant</td>
<td>neutral</td>
<td>important</td>
<td>completely important</td>
</tr>
<tr>
<td>not at all confident</td>
<td>slightly confident</td>
<td>moderately confident</td>
<td>quite confident</td>
<td>extremely confident</td>
</tr>
<tr>
<td>completely dissatisfied</td>
<td>moderately dissatisfied</td>
<td>neutral</td>
<td>moderately satisfied</td>
<td>completely satisfied</td>
</tr>
<tr>
<td>not at all satisfied</td>
<td>slightly satisfied</td>
<td>moderately satisfied</td>
<td>quite satisfied</td>
<td>extremely satisfied</td>
</tr>
<tr>
<td>not at all bored</td>
<td>slightly bored</td>
<td>moderately bored</td>
<td>quite bored</td>
<td>extremely bored</td>
</tr>
<tr>
<td>not at all frustrated</td>
<td>slightly frustrated</td>
<td>moderately frustrated</td>
<td>quite frustrated</td>
<td>extremely frustrated</td>
</tr>
<tr>
<td>strongly prefer x</td>
<td>prefer x</td>
<td>neutral</td>
<td>prefer y</td>
<td>strongly prefer y</td>
</tr>
<tr>
<td>almost no effort</td>
<td>a little bit of effort</td>
<td>some effort</td>
<td>quite a bit of effort</td>
<td>a great deal of effort</td>
</tr>
<tr>
<td>very poor</td>
<td>poor</td>
<td>barely acceptable</td>
<td>good</td>
<td>very good</td>
</tr>
</tbody>
</table>
### More General Response Scales

<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>Disagree strongly</td>
<td>Disagree</td>
<td>Tend to disagree</td>
<td>Tend to agree</td>
<td>Agree</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>little</td>
<td>somewhat</td>
<td>much</td>
<td>a great deal</td>
<td></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:


