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)

<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 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

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 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

10 Tunes
To Make You Think…End of Out by One

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

One (Is the Loneliest Number):
Three Dog Night

One and One:
Robert Miles

You’re Still the One:
Shania Twain

One Love:
Bob Marley and the Wailers

Do That to Me One More Time:
Captain and Tennille

One on One:
Hall and Oates

Baby One More Time:
Britney Spears

Back at One:
Brian McKnight

If You’re Not the One:
Daniel Bedingfield

Bonus: 11:11 by Rufus Wainwright

Academic Internal Medicine Insight  |  2011  |  9:1