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.[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?

5-point, Likert-type response scale:

Not at all confident	Slightly confident	Moderately confident	Quite confident	Extremely confident

References

1. Dillman DA, Smyth JD, Christian LM. Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method. 3rd ed. Hoboken, NJ: Wiley; 2009.
2. Fowler FJ. Survey Research Methods. 4th ed. Thousand Oaks, Calif: Sage Publications; 2009.
3. McKenzie JF, Wood ML, Kotecki JE, Clark JK, Brey RA. Establishing content validity: Using qualitative and quantitative steps. Am J Health Behav. 1999;23:311–318.
4. Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. Objectifying content validity: Conducting a content validity study in social work research. Soc Work Res. 2003;27:94–104.
5. Karabenick SA, Woolley ME, Friedel JM, et al. Cognitive processing of self-report items in educational research: Do they think what we mean? Educ Psychol. 2007;42:139–151.
6. Willis GB. Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, Calif: Sage Publications; 2005.
7. DeVellis RF. Scale Development: Theory and Applications. 2nd ed. Newbury Park, Calif: Sage; 2003.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

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.

Pitfall	Survey example(s)	Why it's a problem	Solution(s)	Survey example(s)
Creating a double-barreled item	How often do you talk to your nurses and administrative staff when you have a problem?	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]	When you have multiple questions/premises 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]	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?
Creating a negatively worded item	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.	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]	In an average week, how many times do you start class on time? Should the chief resident be responsible for admitting patients?
Using statements instead of questions	I am confident I can do well in this course. • not at all true • a little bit true • somewhat true • mostly true • completely true	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]	How confident are you that you can do well in this course? • not at all confident • slightly confident • moderately confident • quite confident • extremely confident
Using agreement response anchors	The high cost of health care is the most important issue in America today. • strongly disagree • disagree • neutral • agree • strongly agree	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]	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
Using too few or too many response anchors	How useful was your medical school training in clinical decision making? • not at all useful • somewhat useful • very useful	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]	How useful was your medical school training in clinical decision making? • not at all useful • slightly useful • moderately useful • quite useful • extremely useful

References:
1. Dillman DA, Smyth JD, Christian LM. Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method. 3rd ed. New York, NY: John Wiley & Sons; 2009.
2. Schwarz N. Self-reports: How the questions shape the answers. Am Psychol. 1999;54:93–105.
3. Krosnick JA. Survey research. Annu Rev Psychol. 1999;50:537–567.
4. Tourangeau R, Rips LJ, Rasinski KA. The Psychology of Survey Response. New York, NY: Cambridge University Press; 2000.
5. Weng L. Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. Educ Psychol Meas. 2004;64:956–972.

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.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

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,[1] 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.

Pitfall	Solution(s)
1. Labeling only the end points of your response options 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.[2]	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.
Problematic item: How interesting did you find this clinical reasoning course? ● not at all interesting ● ● ● ● extremely interesting	**Improved item:** How interesting did you find this clinical reasoning course? ● not at all interesting ● slightly interesting ● moderately interesting ● quite interesting ● extremely interesting
2. Labeling response options with both numbers and verbal labels Because of the additional information respondents must process, including numbers and verbal labels extends response time.[3] 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."	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.[2]
Problematic item: How much did you learn in today's workshop? ● -2 almost nothing ● -1 a little bit ● 0 some ● 1 quite a bit ● 2 a tremendous amount	**Improved item:** How much did you learn in today's workshop? ● almost nothing ● a little bit ● some ● quite a bit ● a tremendous amount
3. 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.[4] In addition, unbalanced spacing of the response options can shift the visual midpoint of the scale.	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.
Problematic item: How much did you learn from your peers in this course? ● almost nothing ● a little bit ● some ● quite a bit ● a tremendous amount	**Improved item:** How much did you learn from your peers in this course? ● almost nothing ● a little bit ● some ● quite a bit ● a tremendous amount
4. 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.[4]	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.[4] This recommendation is a beneficial exception to the guidance above about maintaining equal spacing between response options.
Problematic item: How satisfied are you with the quality of the library services? ● not at all satisfied ● slightly satisfied ● moderately satisfied ● quite satisfied ● extremely satisfied ● not applicable	**Improved item:** How satisfied are you with the quality of the library services? ● not at all satisfied ● slightly satisfied ● moderately satisfied ● quite satisfied ● extremely satisfied ● not applicable

References:

1. Artino AR, Gehlbach H, Durning SJ. Avoiding five common pitfalls of survey design. Acad Med. 2011;86:1327.
2. Krosnick JA. Survey research. Annu Rev Psychol. 1999;50:537-567.
3. Christian LM, Parsons NL, Dillman DA. Designing scalar questions for web surveys. Sociol Meth Res. 2009;37:392-424.
4. Dillman DA, Smyth JD, Christian LM. Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method. 3rd ed. New York, NY: John Wiley & Sons; 2009.

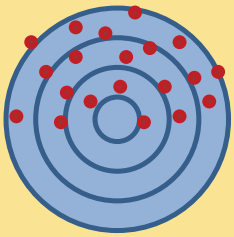
AM Last Page: Reliability and Validity in Educational Measurement

Anthony R. Artino, Jr., PhD, assistant professor of preventive medicine and biometrics, Steven J. Durning, MD, professor of medicine, and Alisha H. Creel, PhD, assistant professor of social and behavioral sciences, Uniformed Services University of the Health Sciences

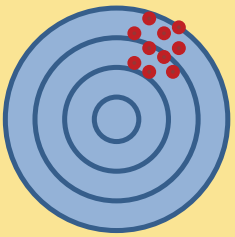
Reliability is the extent to which the scores produced by a particular measurement tool or procedure are consistent and reproducible.[1] Reliability answers the question, "Does the assessment yield the same scores at different times, from different raters, or from different items?"

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.

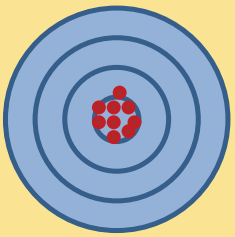
A target provides a metaphor for the relationship between reliability and validity.
The true score (or value) for the concept the researcher is attempting to measure is at the center of the target, and the observed score the investigator gets from each person assessed is a shot at the target.



Neither reliable nor valid



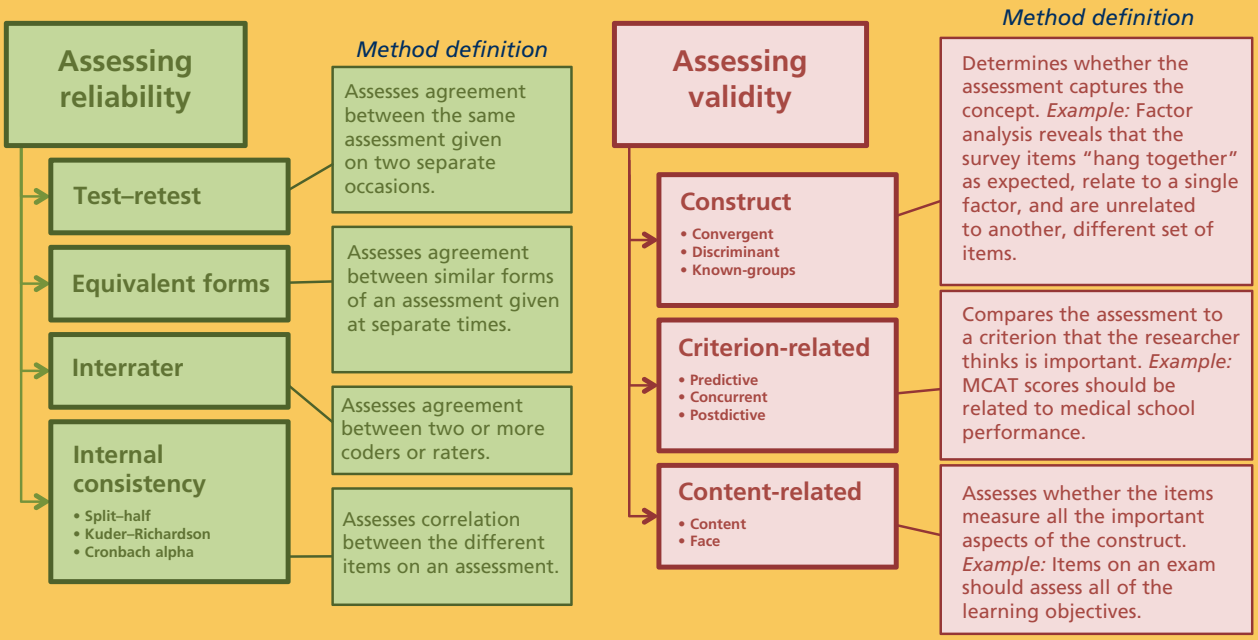
Reliable, but not valid



Both reliable and valid

Reliability is a *necessary but insufficient* condition for validity. To be valid, scores must first be at least moderately reliable.[1–3] 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.



References

1. Thorndike RM. Measurement and Evaluation in Psychology and Education. 7th ed. Upper Saddle River, NJ: Pearson Education; 2005.
2. American Educational Research Association; American Psychological Association; National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 1999.
3. Downing SM. Validity: On the meaningful interpretation of assessment data. Med Educ. 2003;37:830–837.
4. Kane MT. Validation. In: Educational Measurement, 4th ed. Brennan RL, ed. Westport, CT: American Council on Education and Praeger Publishers; 2006:17–64.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

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

Continued on page 19

Continued from page 11

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

1. Gehlbach H, Artino AR, Durning SJ. Last page: Survey development guidance for medical education researchers. *Acad Med.* 2010; 85(5): 925.
2. McKenzie JF, Wood ML, Kotecki JE, Clark JK, Brey RA. Establishing content validity: Using qualitative and quantitative steps. *Am J Health Behav.* 1999; 23(4): 311-318.
3. Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. Objectifying content validity: Conducting a content validity study in social work research. *Soc Work Res.* 2003; 27(2): 94-104.
4. Tourangeau R, Rips LJ, Rasinski KA. *The Psychology of Survey Response.* New York: Cambridge University Press. 2000.
5. Willis GB. *Cognitive interviewing: A Tool for Improving Questionnaire Design.* Thousand Oaks, CA: Sage Publications. 2005.

Continued from page 13

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 Fergusen
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 for 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: U2
One Love: Bob Marley and the Wailers
Do That to Me One More Time: Captain and Teneille
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

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.

Survey Design

Glossary of Terms

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

					Construct:
not important	somewhat important	important	very important	extremely important	importance
unimportant	of little importance	moderately important	important	very important	importance
not at all important	slightly important	moderately important	quite important	extremely important	importance
completely unimportant	unimportant	neutral	important	completely important	importance
not at all confident	slightly confident	moderately confident	quite confident	extremely confident	self-efficacy (confidence)
completely dissatisfied	moderately dissatisfied	neutral	moderately satisfied	completely satisfied	satisfaction
not at all satisfied	slightly satisfied	moderately satisfied	quite satisfied	extremely satisfied	satisfaction
not at all bored	slightly bored	moderately bored	quite bored	extremely bored	boredom
not at all frustrated	slightly frustrated	moderately frustrated	quite frustrated	extremely frustrated	frustration
strongly prefer x	prefer x	neutral	prefer y	strongly prefer y	comparing x to y
almost no effort	a little bit of effort	some effort	quite a bit of effort	a great deal of effort	effort
very poor	poor	barely acceptable	good	very good	quality

More General Response Scales

completely untrue	somewhat untrue	yes and no	somewhat true	completely true

not at all true of me	slightly true of me	somewhat true of me	mostly true of me	completely true of me

strongly disagree	disagree	neutral	agree	strongly agree

or neither agree nor disagree

disagree strongly	disagree	tend to disagree	tend to agree	agree	agree strongly

disagree strongly	disagree moderately	disagree slightly	agree slightly	agree moderately	agree strongly

disagree very strongly	disagree strongly	disagree	agree	agree strongly	agree very strongly

completely disagree	disagree	neutral	agree	completely agree

completely disagree	mostly disagree	slightly disagree	slightly agree	mostly agree	completely agree

Frequency or "Degree" Response Scales

almost never	once in a while	sometimes	often	almost all the time	
never	seldom	about half the time	usually	always	
never	little	somewhat	much	a great deal	
not at all	very little	moderately	quite a bit	a tremendous amount	
never	rarely	occasionally	frequently	almost always	
never	seldom	sometimes	often		
never	rarely	sometimes	often	very often	always
seldom	occasionally	to a considerable degree	almost always		
never	very rarely	rarely	occasionally	frequently	very frequently
never	very rarely	rarely	occasionally	very frequently	always

Survey Development References

Good General Textbooks & Articles:

Bradburn, N. M., Rips, L, J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236*, 157-161.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Newbury Park, CA: Sage.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). New York: John Wiley & Sons.

Fowler, F. J. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537-567.

Schwarz, N. (1999). Self-reports. How the questions shape the answers. *American Psychologist, 54*, 93-105.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. New York: Cambridge University Press.

Articles on Expert Validation:

McKenzie, J.F., Wood, M.L., Kotecki, J.E., Clark, J.K., & Brey, R.A. (1999). Establishing content validity: Using qualitative and quantitative steps. *American Journal of Health Behavior, 23*, 311-318.

Rubio, D.M., Berg-Weger, M., Tebb, S.S., Lee, E.S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research, 27*, 94-104.

Articles on Cognitive Interviewing:

Karabenick, S.A., Woolley, M.E., Friedel, J.M., Ammon, B.V., Blazevski, J., Bonney, C.R., et al. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist, 42*, 139-151.

Willis, G.B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.

Articles on Reliability and Factor Analyses:

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common Errors and some comment on improved practice. *Educational and Psychological Measurement, 66*, 393-416.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*, 350-353.

Thompson, B. (1994). Guidelines for authors reporting score reliability estimates. *Educational and Psychological Measurement, 54*, 837-847.