How to Create a Bad Survey Instrument

Gail M. Sullivan, MD, MPH
Anthony R. Artino Jr, PhD

As educators and education researchers, we often survey trainees, faculty, and patients as a rapid and accurate method to obtain data on outcomes of interest. The Accreditation Council for Graduate Medical Education surveys residents and faculty every year; institutions survey graduating residents and staff regarding learning environments; program directors survey residents about rotation experiences and faculty skills; and researchers use surveys to measure a range of outcomes, from empathy to well-being to patient satisfaction. As editors, we see survey instruments in submitted manuscripts daily. These include questionnaires previously used in other studies, and others that are "homegrown." In a 2012 review of papers submitted to this journal, 77% used a survey instrument to assess at least 1 outcome,[1] and a more recent study of 3 high-impact medical education journals found that 52% of research studies used at least 1 survey.[2] Despite advice from many sources, including this journal,[1,3] we continue to see manuscript submissions with surveys unlikely to yield reliable or valid data.

We suggest that if you want to create a dubious, low-quality survey, follow the tips in bold below (and see the TABLE). If not, read further about each item. Remember that creating a credible survey may take more time and effort, but it is well worth the investment.

To Create a Weak Survey

1. Ask Leading or Biased Questions So That Respondents Will Feel Compelled to Answer in a Particular Way

Questions expressed in a neutral manner are more likely to generate unbiased answers and may be less annoying to respondents. For example, the question "*How interested are you in learning the essential skill of delivering bad news?*" suggests that this skill is important. Or, consider the question "*Should responsible physicians discuss unproven risks of vaccination with parents?*" which implies that these discussions are more likely to occur with physicians who are particularly responsible.

Whenever the question topic concerns *values*, in which respondents may be more likely to choose more socially acceptable answers (social desirability response bias), careful attention to neutral language is critical.

2. Include 2 (or More) Questions in 1 Item So That Respondents Don't Know Which One to Answer

"Double-barreled" questions are common. They place respondents in the awkward position of deciding which part of the question to answer, especially if their answers are not the same for each part. For example, it would be difficult for respondents to rate "*The new educational program on fatigue mitigation strategies is time-efficient and meets the needs of my residents*" because it presents 2 questions in a single statement. The first question concerns the extent to which the educational program is time-efficient, while the second asks whether the program meets the needs of the residents. Such items cause problems for respondents and for the researchers hoping to use the results. Survey writers may develop such questions in an attempt to shorten the survey's length. However, to collect accurate information, these questions must be split into 2 or more parts.

3. Ask Vague Questions

Both open-ended and close-ended questions should be appropriately focused to garner an accurate, thoughtful response. The open-ended question "*What is your assessment of this rotation?*" is quite vague compared to questions that ask about specific facets of the rotation—patients, procedures, faculty availability, faculty teaching skills, and staff supports. Although the latter will require multiple questions, these items are more focused and more likely to elicit specific responses. Vague questions are subject to variable interpretation by respondents, which reduces their usefulness. Although we acknowledge that the investigator's research question should guide the survey design and format, vague questions open to interpretation often result in less-than-useful data.

4. Create Negatively Worded Questions

As with vague questions, negatively worded questions or statements can be confusing to readers. For example, the question "*How many times in the past*

DOI: http://dx.doi.org/10.4300/JGME-D-17-00375.1

TABLE
Common Problems and Alternatives in Survey Design

Problem	Poor Example	Better Alternative
1. Leading or biased questions	Which of the following do you believe is *most* responsible for the high costs of health care? ▪ Physicians ▪ Irresponsible health insurance companies ▪ The federal government	Which of the following do you believe is most responsible for the high costs of health care? ▪ Physicians ▪ Health insurance companies ▪ The federal government
2. Double-barreled questions	Was your attending on time to rounds and knowledgeable about the discussion topic?	Was your attending on time to rounds? Was your attending knowledgeable about the discussion topic?
3. Vague questions	How was your rotation experience?	Please rate the quality of attending rounds: Please rate the variety of patients seen on attending rounds:
4. Negatively worded questions	How often do you fail to start rounds on time?	How often do you start rounds on time?
5. Acronyms, nonspecific, or unfamiliar terms	For your most recent MAX experience, were the CBME Milestones reviewed at the start of the rotation?	For your most recent outpatient medicine rotation, were the competencies (skills) for the rotation reviewed at the start of the rotation?
6. Incomplete range or overlapping answer choices	Approximately what percentage of patients who you cared for in clinic the past month were over age 70 years? Choices: 0%–25%; 25%–50%; 50%–75%; 75%–100%	Choices: 0%–24%; 25%–49%; 50%–74%; 75%–100%; not applicable, I was not in clinic last month
7. Absolute answers, such as *always*	In attending rounds over the past 2 months, about how often did you provide scheduled, midrotation formative feedback to the interns? Choices: always; sometimes; rarely; never	Choices: more than 75% of the time; 51% to 75% of the time; about 50% of the time; 25% to 49% of the time; less than 25% of the time
8. Responses that do not match questions	During clinic rotation last month, did the interdisciplinary team meetings assist you in caring for your patients? Choices: strongly agree; agree; neutral; disagree; strongly disagree	Choices: yes, helpful with many patients; yes, helpful with a few patients; no, the team was not helpful; not applicable, I did not discuss my patients with the interdisciplinary team
9. No content review by experts	Survey may omit key areas or not reflect recent studies	Review literature; modify or create the survey; experts review survey
10. No pretesting with similar individuals	Questions and answer responses may not be interpreted consistently by respondents or as you intended	Pretest survey with sample of subjects similar to your target population, using cognitive interviewing techniques[3]
11. No pilot study to examine score reliability or relation to other variables	Survey scores may not be reliable; scores may not measure what you think they measure	Conduct small or large scale pilot study and begin assessing score reliability and validity evidence
12. Excessively long survey	Survey is pages long, with many unnecessary items that may not be used in the analysis	Pretest survey to determine time required to complete; use analysis plan to guide which questions are necessary and which can be removed

month did you not attend conferences due to clinical care that was not assumed by the covering resident?" is overly difficult to interpret. If a respondent has to *disagree* in order to *agree*, or say *no* in order to mean *yes*, that's a problem. Staying positive is a much better approach.

5. Use Acronyms, Vague, or Unfamiliar Terms

One of our favorite vague terms is *formal teaching*. Does this refer to lectures, small group conferences, online modules, attending sit-down rounds, morning report, some of the above, or all of the above? A key principle in survey design is that questions should be understood and interpreted the same way by all respondents.[1,4] This does not mean that all respondents will *answer* in the same way—for if they did, a survey would not be needed. However, respondents should understand the question the same way and as the survey developer intended. In the example above, if the survey developer defines the term *formal teaching* operationally, there is a greater chance that all respondents will understand the question in the same way (eg, "*Does your program offer small group conferences on pain management?*").

Similarly, using abbreviations that not all respondents know will decrease response accuracy and completeness. For instance, the item "*Did your medical school graduation requirements include achievement of the AAMC EPAs?*" may be clear to some, but may confuse others. Much like the guidelines for writing a manuscript, survey developers should use acronyms sparingly, and define them when first introduced.

6. Provide Answer Choices That Fail to Include the Full Spectrum of Potential Answers, So That Respondents May Be Forced to Choose an Incorrect Option or Skip the Question—Alternatively, Provide Response Options That Are Not Mutually Exclusive

We sometimes see survey response options with answers that overlap, are not mutually exclusive, or are not comprehensive. This error is obvious with quantitative answers, for example, these overlapping options in response to "*How many years since you left residency?*": *0–4*, *4–8*, *8–12*, or *12–16*." Although less obvious, this mistake can also occur with qualitative answer choices. For example, the following response options for the question "*What's your occupational status?*" are not mutually exclusive: "*full-time employment*, *part-time employment*, *full-time student*, *part-time student*, or *unemployed*."

To provide comprehensive choices to the question "*How many years have you been in clinical practice?*" a survey should probably include the answer choice "*None, I am not in clinical practice*." Otherwise, respondents not in clinical practice may skip the question or may choose the lowest number range. In this case, incomplete or inaccurate data are collected. The difficulty comes with figuring out the best way to construct such a question. The solution to this challenge is to pretest your survey.[1,3,4]

7. Provide Absolute Answer Options, Like Always or Never

Always and *never* are tough standards to reach. Some individuals will be more flexible and interpret *always* to be *all but once, a while ago*, and others will interpret this as an absolute. Other than the sun rising each day, absolutes are rare; therefore, these are often unnecessary response options. We see this in duty hour research, such as "*How often do you accurately report duty hour violations to your residency program?*" with answer options "*always*, *usually*, *sometimes*, *seldom*, or *never*." These answer choices are *vague quantifiers*,[4] subject to interpretation; what is *sometimes* to one individual may be *seldom* to another. Defining these vague quantifiers may help. This question exhibits another problem: *violations* is a fairly loaded term, with negative connotations that may influence respondents to answer less accurately.

8. Use the Same Response Options for All Survey Items, Even When They Don't Match the Question

Some surveys utilize the same answer choices for the entire survey. Agreement response options, which ask respondents to agree or disagree with a set of statements, are commonly used. The typical set of response options corresponding to such items are *strongly disagree, disagree, neither agree nor disagree, agree*, and *strongly agree*. For statements that ask about agreement, these options *may* work, but they do not for other types of questions. For example, an agreement response scale makes little sense for a yes/no question like "*Were our administrative staff courteous?*"

Although agreement response options are popular, most survey design experts oppose their use. Agreement response options are susceptible to acquiescence (ie, the tendency for respondents to agree just because they want to be agreeable).[5,6] Experts also argue that such items conflate the degree to which the respondent is agreeable and his actual answer on the construct being measured.[4] Survey questions and corresponding response options should highlight the construct being measured, and agreement response options should be used rarely.[4–7]

9. Create a New or Adapted Survey Tool Without Any Review by Content Experts

An important step in creating a survey is to search the literature for prior studies in the area, not only to inform your education or research project, but also to determine if there are existing surveys that may work for your project.[1,4,6] If there are existing survey instruments, closely review the content to assess if they make sense for your context, as well as evaluate any evidence supporting the reliability and validity of the survey scores.[7,8] If there is no such evidence, which happens with some regularity, you may still wish to adapt the survey items for your own purpose, population, and setting—preferably after discussing the adaptation with the original survey creators. Often there are no existing suitable surveys, and authors must create their own from scratch. An important early step in the development process is for content experts (those with particular expertise in the topic area) to review your survey. Content experts can check the survey items for clarity and relevance, and suggest important facets of the topic of interest that might be missing.[1,7]

10. Administer the New or Adapted Survey Without First Testing It on Individuals Similar to Your Target Audience

No matter how careful you are in crafting questions and response options, it is likely that some items will be confusing, at least to some people. It is a straightforward and effective strategy to sit down with someone similar to your target sample to conduct a cognitive interview.[3] Such interviews can be done using several different techniques, such as a think-aloud protocol that has subjects talk through their thought processes while reading and answering specific survey questions. In some instances, the subject may complete the entire survey. Areas of confusion can be further elicited by asking questions such as "*Why did you pick this option and not the others?*" It is important that respondents understand the questions and answers as intended by the designer. It also is important that they find the questions acceptable, rather than biased, too personal, annoying, or otherwise undesirable.[3]

11. Before Using the Survey, Don't Pilot Test It to See if the Scores Are Reliable and Measure What the Researchers Intended

Piloting the survey, evaluating score reliability, and comparing responses to other available measures can boost its quality and credibility enormously.[1,4,7,8] In addition, this process is essential for estimating the amount of time required to complete the survey. Despite your best efforts, it is often not clear how a survey will function until it is tested under realistic conditions with real people.

Pilot testing is an important part of the overall validity argument for the survey scores and their intended use.[8] Testing is an important way to assess the quality of individual survey items, as well as how the items work together as a coherent whole. As part of pilot testing, survey developers often collect validity evidence by assessing the survey's internal structure (such as through factor and reliability analysis) and relationships with other variables.[1,8]

12. Make the Survey So Long That Respondents Will Answer Carelessly Just to Finish

Survey fatigue[4] is a real problem in survey research and must be considered. Residents, faculty, program directors—everyone is "surveyed out," and most will groan at the sight of yet another survey request. In the days of paper surveys, this fatigue often resulted in respondents circling the same answer for an entire page or stopping after the first page—neither of which produced useful results. In the era of online surveys, this fatigue often results in respondents closing their browsers before the survey is completed. A shorter survey is almost always better than a longer one. Prior pretesting should provide an estimate of how long it takes respondents to read and answer the survey.

In Summary

Too often, educators and researchers rapidly assemble a questionnaire in the hopes of ascertaining attitudes, opinions, and behaviors of their trainees, colleagues, or patients. Frequently, these efforts produce a low-quality survey instrument, with scores that are neither reliable nor valid for their intended use. If you want to create such a survey tool, follow the 12 tips above. If not, we recommend a more systematic approach. Yes, such an approach takes more time and effort, but the benefits far outweigh the costs.

References

1. Rickards G, Magee C, Artino AR Jr. You can't fix by analysis what you've spoiled by design: developing survey instruments and collecting validity evidence. *J Grad Med Educ*. 2012;4(4):407–410.
2. Phillips AW, Friedman BT, Utrankar A, et al. Surveys of health professions trainees: prevalence, response rates, and predictive factors to guide researchers. *Acad Med*. 2017;92(2):222–228.
3. Willis GB, Artino AR Jr. What do our respondents think we're asking? Using cognitive interviewing to improve

medical education surveys. *J Grad Med Educ*. 2013;5(3):353–356.
4. Dillman DA, Smyth JD, Christian LM. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4th ed. Hoboken, NJ: John Wiley & Sons Inc; 2014.
5. Krosnick JA. Survey research. *Ann Rev Psychol*. 1999;50(1):537–567.
6. Saris WE, Revilla M, Krosnick JA, et al. Comparing questions with agree/disagree response options to questions with item-specific response options. *Surv Res Methods*. 2010;4(1):61–79.
7. Gehlbach H, Brinkworth ME. Measure twice, cut down error: a process for enhancing the validity of survey scales. *Rev Gen Psychol*. 2011;15(4):380–387.
8. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119(2):166.e7–e16.

Gail M. Sullivan, MD, MPH, is Editor-in-Chief, *Journal of Graduate Medical Education* (*JGME*), and Professor of Medicine, University of Connecticut; and **Anthony R. Artino Jr, PhD,** is Deputy Editor, *JGME*, and Professor, Department of Medicine, Uniformed Services University of the Health Sciences.

Dr Artino is a military service member. The views expressed in this article are those of the author and do not necessarily reflect the official views of the Uniformed Services University of the Health Sciences, the US Navy, or the Department of Defense.

Corresponding author: Gail M. Sullivan, MD, MPH, University of Connecticut, 253 Farmington Avenue, Farmington, CT 06030-5215, gsullivan@uchc.edu