

EDUCATOR'S BLUEPRINT

Educator's blueprint: A how-to guide for survey design

Jeffery Hill MD MED¹   | Kathleen Ogle MD²   | Sally A. Santen MD, PhD^{1,3}  |
Michael Gottlieb MD⁴   | Anthony R. Artino Jr PhD²  

¹University of Cincinnati College of Medicine, Cincinnati, Ohio, USA

²George Washington University School of Medicine and Health Sciences, Washington, DC, USA

³Virginia Commonwealth University School of Medicine, Richmond, Virginia, USA

⁴Rush University Medical Center, Chicago, Illinois, USA

Correspondence

Sally A. Santen, MD, PhD, Virginia Commonwealth University School of Medicine, Richmond, VA 23298, USA.
Email: santensa@ucmail.uc.edu

Abstract

Surveys are ubiquitous in medical education. They can be valuable for assessment across a wide range of applications and are frequently used in medical education research. This Educator's Blueprint paper reviews the best practices in survey design with a focus on survey development. Key components of the survey design process include determining whether a survey is the right tool, using an intentional approach to content development, and following best practices in item writing and formatting. These processes are meant to help educators and researchers design better surveys for making better decisions.

BACKGROUND

The modern world is replete with surveys. From a pop-up on our phone ("How are you enjoying this app?") to a Press Ganey survey after a visit to a health care provider or a survey used as part of a research project, it is difficult to imagine a day passing when you are not asked to complete a survey. Like any other assessment tool, surveys can be used in both high-stakes and low-stakes settings. When used in higher-stakes environments, such as a grant-funded research study, there is a greater burden on the researcher to ensure they have collected reliability and validity evidence for the survey scores and their intended use. In this series of papers, we will discuss the best practices in survey development and implementation, focusing primarily on surveys used for research purposes.

Although surveys are commonly employed in medical education, there is no single survey design standard. Instead, survey designers must use evidence-informed best practices, which are based on decades of empirical evidence, to guide their efforts. Furthermore, in the many places where the empirical evidence is limited or conflicting, designers should apply theoretical guidance to guide their

survey design efforts. In this first paper, we discuss: (1) the selection of a survey as a measurement tool, (2) an intentional approach to content development, and (3) an evidence-informed approach to question formulation and formatting. Future papers will cover gathering reliability and validity evidence, survey administration, and best practices for reporting.

IS A SURVEY THE RIGHT TOOL?

For the educator, researcher, or program evaluator, the first and often most important question to ask is the following: "Is a survey the right tool to measure my variables of interest and answer my question?" In many cases, the answer to this question is "no," and the designer may find that their outcomes of interest are best measured in other ways. For example, a study evaluating the *effectiveness* of a short, just-in-time procedural education video is likely best accomplished through objective measures of procedural effectiveness.¹ Alternatively, a study examining different types of stressors in a simulated clinical environment may use heart rate variability to assess

Supervising Editor: Dr. Anne Messman.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *AEM Education and Training* published by Wiley Periodicals LLC on behalf of Society for Academic Emergency Medicine.

objective physiologic response to stressors and a survey instrument to assess trainees' subjective perceptions of stress.² Additionally, qualitative methodologies are often ideally suited to deeply explore poorly understood or poorly defined concepts, such as physician shame.³ Ultimately, the prospective survey designer should carefully consider the strengths and limitations of using a survey as a research tool.

Surveys are descriptive tools that employ questions to collect statistical information on some facet of a population.⁴ Surveys are best suited for collecting data on nonobservable human phenomena. These include attitudes, beliefs, and opinions but can also include reports of behaviors and actions that are otherwise unmeasurable (or very hard to measure).⁵ Similar to the process of curriculum development, where methods of teaching should logically align with the defined educational objectives, the use of a survey should logically align with the underlying questions being asked and the variables or outcomes being studied.⁶ What is more, researchers should clearly describe in their articles the rationale for those measures and the subsequent use (or uses) of a survey.^{4,5}

CONTENT DEVELOPMENT

High-quality surveys are rigorously developed assessment tools that have validity evidence in support of their proposed uses. There are various components of validity evidence, which we will cover in more detail in a subsequent paper. Establishing content validity (one source of validity evidence) for a survey begins with an intentional and rigorous approach to content development. The process is again similar to curriculum development where an educator first defines the overarching goal of a curriculum and then writes specific learning objectives, which build to the successful achievement of that goal. In survey development, the designer should first establish the overall goal of the study or survey (i.e., what question is being addressed and what variables are being measured with the survey?). Following that overall goal, the researcher then defines specific constructs that best represent the study variables or outcomes being assessed.⁷

Take, for example, a study examining the impact of virtual didactics on resident education. For such a study, the researcher should first identify the overall goal for their survey (e.g., "This survey will measure the perceived effects of video conferencing software for didactic teaching on learning outcomes in resident education"). There are a number of constructs that could feed into that overarching goal, including task-technology fit, cognitive load, and resident wellness (to name just a few). These constructs then guide the development of individual survey items that are used to assess, for example, the utility of specific features of the video conferencing software, distractions in the home environment, and resident attitudes.

Defining the objectives and constructs is an iterative process that starts with a detailed literature review to identify any previously published surveys used to assess similar or related variables in prior work. While identifying a survey that perfectly matches the needs of

the study purpose or research question is unlikely, it is possible that previous surveys have addressed some aspects of the study and/or constructs of interest. Researchers should ideally seek out published instruments that report robust validity and reliability evidence, understanding, however, that many published survey instruments fail to adequately describe this evidence.⁶ Researchers should be aware, as well, that any alterations to existing surveys, to include using the survey in a new population, may change the validity argument and whether or not the instrument is still appropriate for its (new) intended use.⁸ This process of literature review and thoughtful reflection on the goals and objectives of the survey should leave the researcher with a thorough understanding of the constructs to be assessed and ideas on how to frame their survey questions.

BEST PRACTICES FOR ITEM WRITING

Once it is determined that a survey is the best measurement tool, construction of individual survey items can begin. The manner in which each item is developed requires deliberate consideration and should be informed by the literature. As noted in many different fields, including cognitive psychology and public opinion polling, multiple steps may be required to successfully navigate the challenging process of writing high-quality survey items that all respondents will interpret in the same way (and in the way the survey designer intended). Therefore, the use of best practices in item writing will ultimately enhance the survey designer's ability to capture meaningful data. Herein, we propose some structured guidelines to breakdown survey development into its building blocks, thereby supporting high-quality survey-based scholarship (see [Table 1](#) for examples of each best practice).

Write positively worded questions

Framing of individual survey items should generally use positive language. The use of positively worded items enhances response accuracy because they are easier for respondents to comprehend. On the other hand, many respondents may inadvertently miss words and prefixes like "not" and "un-" in negatively worded items, particularly if respondents are trying to get through the survey quickly (which they often are). Negatively worded items require more cognitive resources to understand and, therefore, can be easily misinterpreted by respondents, leading to inaccurate or otherwise hard-to-interpret data.^{9,10}

Use questions and item-specific response options rather than statements with agree/disagree options

In many ways, a survey is like a conversation between the survey designer and their respondents. By asking questions rather than using statements with agree/disagree response options, a

TABLE 1 Sample questions and recommended phrasing

Best practice	Problematic example	Recommended improvement
Write positively worded questions	How often are you unable to start class on time?	How often do you start class on time?
Use questions and item-specific response options	I enjoyed the lecture. (response options: <i>strongly disagree</i> to <i>strongly agree</i>)	How much did you enjoy the lecture? (response option: <i>not at all</i> to <i>a great amount</i>)
Avoid double-barreled items	How effective was the lecture and hands-on instruction?	How effective was the lecture instruction? How effective was the hands-on instruction? Or, the item could be written at a higher level of abstraction: How effective was the residency instruction?
Choose an appropriate number of response options	Did you like the activity? • Yes • No	How much did you like the activity? • Not at all • A little • A moderate amount • Quite a bit • A lot
Attend to formatting and layout	How satisfied were you with your residency training? 1. Not at all satisfied 2. 3. 4. 5. Extremely satisfied	How satisfied were you with your residency training? • Not at all satisfied • Somewhat satisfied • Moderately satisfied • Quite satisfied • Extremely satisfied
Organize the survey items intentionally	<i>First question on the survey:</i> How often do you take illicit drugs?	<i>First question on the survey:</i> What is your favorite extracurricular activity?

survey conversation can flow more naturally—which can enhance respondent comprehension and help respondents process the information being asked.¹⁰ A recent, comprehensive review of this topic reported that statements with agree/disagree response options are associated with more *undesirable* outcomes (e.g., acquiescence and deleterious response effects).¹¹ The authors recommended that survey designers use item-specific questions instead of agree/disagree items for most purposes, as noted in Appendix S1.

Avoid double-barreled items

Researchers often attempt to combine multiple questions into one in an effort to decrease the length of a survey; however, this approach is problematic in that it often results in a double-barreled (or multibarreled) item. Respondents may be confused about how to respond to double-barreled items, particularly if they have one opinion about one part of the question and another opinion about the other. Respondents will use various strategies to handle this challenge, but the critical point is that the survey designer has no way of knowing which approach each person took, thereby making individual responses uninterpretable. When met with this challenge, researchers can take three different approaches: (1) consider which of the two ideas is most important and ask only that one, (2) create two or more separate survey items, or (3) combine the two facets of

the question in a way that encourages respondents to abstract to a more complex idea.⁹ In the latter case, for example, the double-barreled question “How skilled are you at identifying and accommodating your patients’ different communication styles?” could be abstracted to something like “How skilled are you at adapting to patients’ different communication styles?” By taking one of these three approaches, designers can help their respondents focus on and more easily comprehend individual ideas, rather than collections of ideas which may or may not tie together.

Choose an appropriate number of response options

Since the number of response options in a closed-ended survey item may impact the reliability of the responses, researchers should carefully consider the number of options selected. For most purposes, five to seven response options is often the sweet spot for closed-ended survey items.¹² For items which seek to quantify constructs that are unipolar (i.e., things that go from zero to a larger number, like the frequency of a behavior that goes from “never” to “almost all the time”), five response options is often adequate. On the other hand, for items that seek to assess constructs that are more bipolar (i.e., things that go from a negative amount to a positive amount, like an attitude), seven response options is often ideal. Seven options allow for three negative options, three positive options, and one neutral

option in the middle¹² (see several examples in [Table 1](#)). As a general rule, fewer than five response options is likely to decrease the reliability of survey scores and more than seven typically does not enhance reliability and may overburden respondents.¹⁰

Another consideration is whether to use an odd or even number of responses. Although survey designers often spend hours agonizing over this decision, the limited empirical evidence is equivocal, suggesting that the consequences of this choice are minor.⁹ That said, if it makes sense for the survey item to have a neutral point, then an odd number of response options probably makes the most sense (i.e., if it is reasonable for a respondent to be neutral on an issue). On the other hand, if no natural midpoint exists, then it is reasonable to instead use an even number of response options with no midpoint.

Attend to formatting and layout

At a time when virtually everyone has a smartphone, especially professionals in healthcare, it is important to note that most respondents will use their mobile device to complete most web-based surveys. As such, it is essential to preview the survey layout in multiple formats, including the mobile format. For example, web-based survey applications like SurveyMonkey or Qualtrics typically align response options horizontally across the page for surveys presented on a computer screen, but they also have the capability to format the same options vertically for the mobile user. Thus, it is important that researchers carefully review (and even pretest) the appearance of their survey items in both formats to ensure readability, especially because these different formats can affect the quality of the survey responses provided.⁹

Survey designers may also be inclined to include both verbal labels and numbers on their response options, with the goal of enhancing precision. Surprisingly, however, the limited empirical research that exists suggests that respondents often variably interpret such response options (since the meaning of numbers and the meaning of verbal labels can sometimes be misaligned).¹³ This finding suggests that including both verbal labels and numbers may be an ineffective approach. As such, a best practice is to use verbal labels only (without numbers) and to label all response options (as opposed to just the end points); doing so eases respondents' cognitive burden, thereby encouraging more precise answers.⁹ Appendix S1 outlines a number of example verbal labels for commonly studied topics.

Another important formatting issue is response option spacing. Unevenly spaced response options can make certain options stand out visually more than others. This can have the inadvertent effect of making these options more likely to be noticed and selected by respondents. Researchers should be meticulous in their formatting efforts, ensuring that all response options are equally spaced.⁹

Organize the survey items intentionally

As researchers consider respondent motivation to begin and ultimately complete the survey, the order of the items should

be carefully considered and intentional. Obtaining clear and interpretable information from a survey depends on accurate data collection and, as such, the most important questions addressing the central construct of interest should be situated near the beginning of the survey.⁹ That way, if respondents choose to stop taking the survey, at least some useful data have been collected.

Moreover, it is important to keep sensitive questions and demographic items (which are often considered sensitive) toward the end of the survey. In particular, questions about race and ethnicity that are included early in a survey or other assessments have been shown to induce an effect known as stereotype threat, which can negatively impact response quality and respondent motivation. This effect alone is reason enough to ask demographic items near the end of most surveys.¹⁴ Finally, with regard to sensitive questions, the notion of “rapport” is critical. Once the survey designer has built some rapport and buy-in with the respondents early in a survey, then more sensitive questions can be asked. This rapport-building approach applies as much for surveys as it does for everyday conversations.

In addition, all survey questions should be relevant to each respondent. Irrelevant questions tend to demotivate participants and can negatively impact response quality. With this in mind, designers should consider using branching questions.⁴ For example, when gathering information about the quality of the university's library services, a designer might first ask: “Have you used the institution's library?” and then ask questions about the quality of the library services only to those who answer “yes.” Electronically administered, web-based surveys make it easy for designers to create this type of branching question, and many web-based survey products, such as RedCap, Qualtrics, Google Forms, and SurveyMonkey, have tutorials on how to construct a survey with branching logic.

CONCLUSION

Surveys can be a powerful way to answer otherwise unanswerable questions. By using the best practices described here (and in the broader survey design literature)—early in the process of survey development—designers can help to ensure their survey tools are credible assessments that can be used for making high-quality decisions. The evidence-informed approach to survey design presented in our series of papers includes: (1) developing content in a rigorous way, (2) writing and formatting survey items with clarity, (3) collecting evidence to support the validity of the survey results, (4) administering the survey to maximize response rate, and (5) clearly communicating the results of the survey in the medical literature.

In this first paper, we have focused on the first two steps of this process. By first explicating the goals of the survey and the constructs being investigated, designers can determine the content of survey items. Next, by following evidence-informed guidelines for writing and formatting survey items and responses, designers can support the validity of their survey scores and their intended use. In our next paper, we will discuss the critical aspects of testing and piloting surveys to further establish reliability and validity evidence for

the survey scores and their proposed uses. Subsequently, we will describe approaches to administering and distributing surveys, as well as best practices for reporting survey design and research efforts.

ORCID

Jeffery Hill  <https://orcid.org/0000-0001-8369-3022>

Kathleen Ogle  <https://orcid.org/0000-0002-7658-0895>

Sally A. Santen  <https://orcid.org/0000-0002-8327-8002>

Michael Gottlieb  <https://orcid.org/0000-0003-3276-8375>

Anthony R. Artino Jr  <https://orcid.org/0000-0003-2661-7853>

TWITTER

Jeffery Hill  @_drjeffy

Kathleen Ogle  @DrKittyKat

Michael Gottlieb  @MGottliebMD

Anthony R. Artino  @mededdoc

REFERENCES

1. Alshawkani YY, Orfield NJ, Samuel LT, Kuehl DR, Hagan HJ. An ultrashort video can teach residents to perform a fingertip injury repair. *AEM Educ Train*. 2021, Nov 30;6(1):e10713. doi:10.1002/aet2.10713
2. Joseph M, Ray JM, Chang J, et al. All clinical stressors are not created equal: differential task stress in a simulated clinical environment. *AEM Educ Train*. 2022, Jan 24;6:e10726. doi:10.1002/aet2.10726
3. Deutsch AJ, Sangha H, Spadaro A, et al. Defining well-being: A case-study among emergency medicine residents at an academic center: a qualitative study. *AEM Educ Train*. 2021, Nov 24;5(4):e10712. doi:10.1002/aet2.10712
4. Phillips AW, Durning SJ, Artino AR. In: Phillips AW, Durning SJ, Artino AR, eds. *Survey Methods for Medical and Health Professions Education: A Six-Step Approach*. Elsevier; 2021.
5. Artino AR, Durning SJ, Sklar DP. Guidelines for reporting survey-based research submitted to academic medicine. *Acad Med*. 2018;93:337-340.
6. Kern DE, Thomas PA. In: Hughes MT, ed. *Curriculum Development for Medical Education: A Six-Step Approach*. Johns Hopkins University Press; 2009.
7. Rickards G, Magee C, Jr ARA. You can't fix by analysis what you've spoiled by design: developing survey instruments and collecting validity evidence. *J Graduate Med Educ*. 2012;4:407-410.
8. Phillips AW, Artino A. Lies, damned lies, and surveys. *J Graduate Med Educ*. 2017;9:677-679.
9. Gehlbach H, Artino AR. The survey checklist (manifesto). *Acad Med*. 2018;93(3):360-366.
10. Artino AR, Gehlbach H, Durning SJ. AM last page: avoiding five common pitfalls of survey design. *Acad Med*. 2011;86(10):1327.
11. Dykema J, Schaeffer NC, Garbarski D, Assad N, Blixt S. Towards a reconsideration of the use of agree-disagree questions in measuring subjective evaluations. *Res Social Adm Pharm*. 2022;18(2):2335-2344.
12. Krosnick JA, Fabrigar LR. Designing rating scales for effective measurement in surveys. In: Lyberg L, Biemer P, Collins M, et al., eds. *Survey Measurement and Process Quality*. John Wiley & Sons, Inc.; 1997:141-164.
13. Krosnick JA. Survey research. *Annu Rev Psychol*. 1999;50:537-567.
14. Steele CM, Aronson J. Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol*. 1995;69:797-811.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hill J, Ogle K, Santen SA, Gottlieb M, Artino AR. Educator's blueprint: A how-to guide for survey design. *AEM Educ Train*. 2022;10:e10796. doi: [10.1002/aet2.10796](https://doi.org/10.1002/aet2.10796)