

Guidelines for Reporting Survey-Based Research Submitted to *Academic Medicine*

Surveys are ubiquitous in health professions education (HPE). Phillips et al¹ recently reported that 52% of all original research articles in three high-impact HPE journals included a survey. Yet, as stated in this issue of *Academic Medicine*, Artino et al² found flaws in the design of surveys and the reporting of validity and reliability evidence in those same journals. Educators use surveys to assess the attitudes of their medical students, residents, and faculty, such as reported in the study in this issue by Rassbach and Blankenburg,³ who employed a survey to study feedback from faculty coaches to pediatrics residents. Education researchers often use surveys to query colleagues about the adoption of innovations, like the survey of residency directors about the use of a flipped classroom,⁴ also found in this issue. When designed and implemented carefully, surveys can answer questions that no other research method can.⁵

Academic Medicine receives submissions every day that report the use of surveys as a primary means of data collection. We are sent many outstanding survey studies that are both detailed and easy to read. Unfortunately, we also receive a large number of submissions that provide inadequate descriptions of the surveys themselves and how the survey studies were conducted. The failure to fully describe a survey study—including the survey's construction or adaptation, how it was tested prior to use, and how the data were analyzed—is problematic for several reasons.

First, inadequate reporting challenges our reviewers, making it difficult for them to accurately assess how well the survey might yield reliable, credible data from which valid inferences can be made. Without this understanding, reviewers are more likely to reject a survey study for publication. In this issue, Meyer et al⁶ reviewed manuscripts submitted to *Academic Medicine* that were rejected

without external review and found that the most prevalent theme was “ineffective study question and/or design.” Many survey-based projects never make it to the external review stage because of a flawed survey or inadequate reporting.

Second, inadequate reporting stifles the growth of HPE as a scientific field. Although articles sometimes do get published with incomplete or otherwise poor survey descriptions, such reports are more difficult to interpret and leave readers with many unanswered questions about what was done, how, and to what end.

Finally, inadequate reporting makes survey adoption or adaptation dubious because other scholars cannot appropriately judge the degree to which the survey is trustworthy and relevant to their own medical education contexts.

Because we think surveys are an important tool in HPE—and because we want to encourage the dissemination of high-quality survey research—we focus this editorial on the topic of survey reporting in *Academic Medicine* articles. We recognize, of course, that the details and sophistication of a given survey design and implementation project may vary depending on the purpose and maturity of the ideas being explored and the type of submission (e.g., Research Report, Article, Innovation Report).

Selected Reporting Guidelines for Survey Studies

Small differences in how a survey is designed, formatted, implemented, and analyzed can lead to large differences in survey results.⁷ This fundamental principle has guided four decades of survey research in fields like cognitive psychology and public opinion polling. One of the most powerful examples of the effects of survey design is the infamous “butterfly ballot” used in Palm Beach County, Florida, during the 2000 U.S. presidential election. According to an analysis conducted by Wand et al,⁸ the ballot's double-column format caused

several thousand Al Gore supporters to inadvertently vote for Pat Buchanan, thereby costing Gore the election and, some might argue, changing the course of modern American history.

Although the stakes of a single survey in HPE are rarely as high as choosing the next leader of the United States, the butterfly ballot case emphasizes the importance of survey design and, for the purposes of research dissemination, the significance of accurately and completely reporting a survey's design, implementation, and analysis. In short, accurate and complete reporting are necessary if reviewers and readers are to evaluate the credibility of a survey study.

Below, we outline six selected reporting guidelines to help authors make informed decisions during study design and manuscript preparation, and to help reviewers make informed decisions during manuscript evaluation. These six guidelines, along with additional questions for authors to consider, are also listed in Table 1.

1. Provide a rationale for using a survey

Surveys are ideal for measuring things that are not directly observable, including attitudes, opinions, and beliefs, such as residents' satisfaction with feedback from coaches.³ Surveys can also be used to ask people about their past behaviors or to recall previous facts and events, but these uses are thorny for a variety of reasons. First, asking individuals about their past behavior only works if respondents feel comfortable sharing that information. Often, however, respondents do not want to share data about sensitive topics, including, in some cases, illicit behaviors. At other times, respondents are simply unaware of how they have behaved and why.⁵ And when it comes to recalling past events, recall bias and other memory limitations can make it difficult to gather and estimate accurate responses. In light of these and other limitations, it is important that authors articulate, early in the project, their rationale for using a survey as a data collection method.

Table 1
Checklist of Selected Reporting Guidelines for Studies That Use Surveys

Reporting guideline	Questions to address in the manuscript
Introduction	
1. Provide a rationale for using a survey. ^a	<ul style="list-style-type: none"> • Why is a survey an appropriate data collection method? • How can the research question(s) be answered using a survey?
Method	
2. Describe how the survey was created or adapted from existing survey(s).	<ul style="list-style-type: none"> • How were the survey items developed? • What literature was reviewed? • If applicable, what changes were made to previously published surveys?
3. Describe how the survey was pretested prior to full implementation.	<ul style="list-style-type: none"> • Were experts used to pretest the survey? <ul style="list-style-type: none"> ◦ If so, describe their qualifications, how many were used, and what the review process was like. • Were cognitive interviews conducted? <ul style="list-style-type: none"> ◦ If so, describe the interviewees, how many were interviewed, and what the interviewing procedures were like. • Was a pilot test conducted? <ul style="list-style-type: none"> ◦ If so, describe the sample size, the types of participants, and how the pilot test was conducted.
4. Describe the final survey instrument, including how and when it was administered.	<ul style="list-style-type: none"> • Has the content of the final survey draft been described in detail (e.g., number and types of items and response options)? • Has a complete, formatted copy of the survey been provided for inclusion in the article's appendix? • What was the method of survey administration (e.g., web or paper based), and where and when was the survey administered? • Were survey responses anonymous or otherwise confidential? • How were respondents contacted, and how often? • How long did respondents have to complete the survey? • Were respondents offered incentives for completing the survey?
Results	
5. Describe the respondents, response rate, and how nonresponse bias was assessed.	<ul style="list-style-type: none"> • Who comprises the sample, and how does the sample relate to the population of interest? • What was the response rate, and how was it calculated? • Was nonresponse bias assessed, and if so, what was done to correct for it?
6. Describe how score reliability and validity were assessed. ^b	<ul style="list-style-type: none"> • What processes and statistics were used to assess the reliability of the survey scores? • What sources of validity evidence were collected, and how do they support the intended use of the survey results? <ul style="list-style-type: none"> ◦ At a minimum, content and response process validity should be considered (e.g., through expert reviews and cognitive interviewing). ◦ So-called "face validity" may be included but should be supplemented by other sources of validity evidence. • If applicable, what type of validity framework was used to guide survey development and validation (e.g., Messick's five sources of validity evidence,¹⁵ Kane's framework¹⁶)?

^aThe rationale can also be presented in the Method.

^bReliability and validity evidence are often presented elsewhere in a survey research report (e.g., Introduction, Method, or even Discussion).

Moreover, this rationale should be prominently stated in the manuscript so readers can determine whether a survey is appropriate.

2. Describe how the survey was created or adapted from existing survey(s)

A complete and thorough description of how a survey was created, or how it was adapted from a previously published survey, is a critical component to any survey manuscript. Fortunately, there are guides that describe these processes,^{5,9} including how to create good survey items, how to format and administer a survey, and how to analyze the resulting data. In addition, in this issue, Gehlbach and Artino¹⁰ provide a checklist to assist

authors in preparing surveys. Evidence-based best practices such as those in the checklist should be consulted and followed; doing so is one of the easiest ways to improve the surveys we use in HPE.

Although it is beyond the scope of this editorial to detail all of the important steps to developing or adapting a survey, we note that authors should be as thorough as possible when describing the processes they used to create their surveys. Important steps in the item-development process include following best practices in item writing and asking content experts to review the items for clarity, relevance, and topic coverage. Even surveys that are adapted from the

literature must be closely reviewed by authors and improved upon, where appropriate, since the context for the survey may differ from the circumstances under which it was initially developed. In addition, even instruments pulled from the literature and used "as is" may require additional pretesting prior to use in a new context.

3. Discuss how the survey was pretested prior to full implementation

Some authors do not pretest their surveys prior to use, but they should. Pretesting includes activities like expert reviews, cognitive interviewing, and pilot testing, which can help to establish content and response process validity. Experts can

be particularly helpful in identifying key topics that might be missing from a survey. Cognitive interviewing, on the other hand, is a way to ensure that respondents understand the questions in the way the survey designer intended.¹¹ The basic idea is to try out the survey to ascertain whether the questions are clear, and then to modify those that are not and thoroughly describe how this process was done.

When possible, the final step before survey implementation is to conduct a pilot test. Despite best efforts, one never knows how a survey will function until it is used to collect actual data from real people. Pilot testing entails having members of the target population complete the survey in the planned delivery mode (e.g., web- or paper-based). The data obtained from a pilot test, which often include feasibility information, can then be used to identify items or survey formats that may not be functioning properly. In the description of the survey development in manuscripts submitted to *Academic Medicine*, authors should fully report pretesting activities, including, for example, descriptions of those who helped with the expert reviews and cognitive interviews, as well as any pilot testing that was conducted. Such descriptions should contain the qualifications of the experts and how the reviews and/or cognitive interviews were conducted.

Ultimately, pretesting activities are systematic ways of getting more eyes on a survey prior to full implementation, and they are critical to survey success. Authors should have others carefully review their surveys before sending them out to hundreds of busy colleagues, students, residents, or patients. Not only do journals such as ours expect surveys to be rigorously pretested, and for those processes to be documented in survey reports, but we also believe that these activities demonstrate respect for our community of scholars who review survey research and for the participants who give their time to provide data for analysis.

4. Describe the final survey instrument, including how and when it was administered

A full description of the final survey instrument, including the number of survey items and types of items and response options, is essential for

reader comprehension. Additionally, descriptions of how and when a survey was administered are both key to placing survey findings in context. For example, it would be important to know that a survey about course quality was administered anonymously to students the day after a course ended, as opposed to being administered the day before the final exam in a way that students could be identified. Such differences affect how the data should be interpreted. Important administration details to report include the survey format (web-based, paper-based, or interview), whether responses were anonymous, when the survey was sent to respondents, how much time they were given to complete the survey, how many reminders were sent, and whether incentives were offered. In addition, authors should provide a complete, formatted copy of their survey instrument for inclusion in the article's appendix.

5. Describe the respondents, response rate, and how nonresponse bias was assessed

Full descriptions of the sample, the sample size, and response rate calculations are important for interpreting survey results and ensuring that those results reflect the population of interest. Thus, this information should be included in survey manuscripts. There are a number of ways to calculate a response rate,¹² and *Academic Medicine* does not require that authors achieve a minimum response rate. However, very low response rates are related to greater nonresponse bias and thus are carefully considered by reviewers (although recent research suggests that response rates may not be as strongly associated with survey quality or representativeness as scholars once thought).¹³ *Nonresponse bias* refers to the extent to which survey findings may not represent the population of interest. In other words, if nonresponders had participated, would the results have been markedly different? Thus, where possible, authors should report how they assessed nonresponse bias and, where applicable, whether anything was done to correct for it (e.g., using a stratified sample or analyzing respondents' patterns over time).¹³

6. Describe how score reliability and validity were assessed

At its core, a survey is an assessment tool.¹⁴ Therefore, score reliability and

validity are arguably the two most important factors for authors to consider when developing and using a survey. There are many different ways to assess score reliability (or reproducibility), including consistency between raters (interrater reliability), consistency over time (test–retest reliability), and internal consistency reliability (classically reported as Cronbach's alpha). In reporting reliability, authors should describe whether and how they measured score reliability and whether those reliability statistics are adequate for the proposed use.

Validity can be thought of as the extent to which evidence and theory support the *interpretations* and proposed *uses* of the survey data. Therefore, authors should collect evidence from various sources and construct an argument about why the survey tool is relevant for the intended use and why the resulting data should be believed by the reader. In a very simple sense, validity can be seen as the degree to which the survey measures what the authors intended it to measure. Often, a fair amount of validity evidence is collected during the development of the survey through the inclusion of content experts to review the questions (content) and the testing of the questions by a sample of individuals prior to sending out the survey (response process). Other sources of validity evidence require examination and analysis of information that may already exist related to the results of the survey, such as scores on national tests (relationship to other variables). Ultimately, authors should use some type of validity framework, such as those described by Messick¹⁵ or Kane,¹⁶ to help make a case to readers that their survey and the resulting data are credible. Cook and Beckman¹⁷ and Cook et al¹⁸ provide useful discussions of reliability and validity in the context of assessments in HPE research.

Mitigating Factors to Consider

While we have provided some general guidelines for what we look for in articles that include surveys, we understand that there are mitigating circumstances that may limit the development of a survey. For instance, surveys of disaster victims may need to prioritize timeliness over other aspects of the survey. In such cases, there may not be time to test questions in the way that could occur when time is

not so critical. There also are limitations in funding to carry out many of the elements of survey design that may be difficult to overcome, particularly for surveys developed by trainees. A creative, innovative idea embedded in a survey that has design limitations may still offer important insights. As long as those limitations can be clearly identified, so the results of the survey can be understood in the context of those boundaries, there may be value in sharing the information with the academic community, and we encourage submissions of such articles to *Academic Medicine*.

Concluding Thoughts

Surveys provide important information that can guide our understanding of educational innovations and current challenges facing the academic community. Our goal in writing this editorial was to encourage better survey design and better reporting of survey development and implementation without unduly burdening investigators with superfluous requirements. We want HPE investigators to use surveys to ask important questions, and we want to encourage our students and residents to conduct survey research and share their findings with the community. By following this and other relevant guidance,^{5,9,17,19} we believe the time spent on survey projects will be more productive. What's more, we believe survey-based submissions to this and other journals will have better chances for success. In the end, we hope the outcome of better surveys will be better education

and assessment, better research and evaluation, and better patient care.

Anthony R. Artino Jr, PhD

Assistant editor for AM Last Pages

Steven J. Durning, MD, PhD

Deputy editor for research

David P. Sklar, MD

Editor-in-chief

Editor's Note: The opinions expressed in this editorial do not necessarily reflect the opinions of the AAMC or its members. In addition, those opinions do not necessarily reflect the official policies or positions of the Uniformed Services University of the Health Sciences, the U.S. Navy, the Department of Defense, or the U.S. government.

References

- 1 Phillips AW, Friedman BT, Utrankar A, Ta AQ, Reddy ST, Durning SJ. Surveys of health professions trainees: Prevalence, response rates, and predictive factors to guide researchers. *Acad Med*. 2017;92:222–228.
- 2 Artino AR Jr, Phillips AW, Utrankar A, Ta AQ, Durning SJ. "The questions shape the answers": Assessing the quality of published survey instruments in health professions education research. *Acad Med*. 2018;93:456–463.
- 3 Rassbach CE, Blankenburg R. A novel pediatric residency coaching program: Outcomes after one year. *Acad Med*. 2018;93:430–434.
- 4 Wittich CM, Agrawal A, Wang AT, et al. Flipped classrooms in graduate medical education: A national survey of residency program directors. *Acad Med*. 2018;93:471–477.
- 5 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.
- 6 Meyer HS, Durning SJ, Sklar D, Maggio LA. Making the first cut: An analysis of Academic Medicine editors' reasons for not sending manuscripts out for external peer review. *Acad Med*. 2018;93:464–470.
- 7 Schwarz N. Self-reports: How the questions shape the answers. *Am Psychol*. 1999;54:93–105.
- 8 Wand JN, Shotts KW, Sekhon JS, et al. The butterfly did it: The aberrant vote for Buchanan in Palm Beach County, Florida. *Am Pol Science Rev*. 2001;95:793–810.
- 9 Artino AR Jr, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE guide no. 87. *Med Teach*. 2014;36:463–474.
- 10 Gehlbach H, Artino AR Jr. The survey checklist (manifesto). *Acad Med*. 2018;93:360–366.
- 11 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;5:353–356.
- 12 Phillips AW, Friedman BT, Durning SJ. How to calculate a survey response rate: Best practices. *Acad Med*. 2017;92:269.
- 13 Johnson TP, Wislar JS. Response rates and nonresponse errors in surveys. *JAMA*. 2012;307:1805–1806.
- 14 American Educational Research Association; American Psychological Association; National Council on Measurement in Education; Joint Committee on Standards for Educational and Psychological Testing. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 2014.
- 15 Messick S. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol*. 1995;50:741–749.
- 16 Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas*. 2013;50:1–73.
- 17 Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: Theory and application. *Am J Med*. 2006;119:166.e7–166.16.
- 18 Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: A practical guide to Kane's framework. *Med Educ*. 2015;49:560–575.
- 19 Phillips AW. Proper applications for surveys as a study methodology. *West J Emerg Med*. 2017;18:8–11.