

Lies, Damned Lies, and Surveys

Andrew W. Phillips, MD, MEd

Anthony R. Artino Jr, PhD

“Let’s just do a quick survey.”

—Someone in everyone’s program

Surveys are a common research method used in medical education. For example, a retrospective review of the 3 highest impact journals in the field found that more than half of original research studies included a survey as part of the methods.¹ That same review found that only about half of survey-based studies reported a response rate and provided sufficient paradata (ie, information about the survey design and implementation, such as how it was prepared, the credentials of the content experts, and whether an incentive was offered). Another recent study found that 95% of questionnaires (self-administered, written surveys) broke at least 1 commonly accepted tenet in survey question design, and only 35% and 22% mentioned validity or reliability evidence, respectively.²

This matters because even small differences in how a survey is designed and formatted can change the results in important ways. In this editorial, we share a few examples of how survey design decisions can affect results, and we encourage authors to develop and implement survey instruments with the same scientific rigor used for other study methods.

In this issue of the *Journal of Graduate Medical Education*, an article by Yock and colleagues³ adds to the growing body of concern about survey quality in education research. The authors explored problems associated with vague quantifiers—an issue reported in the cognitive psychology and public opinion literature since the 1970s.⁴ Yet, as the authors found, vague quantifiers are present in a mandated annual resident survey that has important consequences for training programs accredited by the Accreditation Council for Graduate Medical Education–International, making it susceptible to “judgment overlap” across different response options.³

Yock and colleagues’ findings are supported by prior research on survey response options in other populations and fields of study. Krosnick and Berent⁵ reported that simply labeling all response anchors (as opposed to labeling just the end points) can make as much as a

40% absolute difference in test-retest consistency. In their example, the researchers challenged the dogma that Americans’ political affiliations were more persistent over time than their attitudes toward political policies. They demonstrated that the differences seen over time were not the result of attitude change, but were an artifact of the survey design. The political affiliation poll used fully labeled options, whereas the political policy attitude poll used partially labeled options, which led the latter to have low reliability—year after year—for the general US public.

In another example, Alwin and Krosnick⁶ found a 37% relative increase in reliability (Cronbach’s alpha, α) for fully labeled ($\alpha = .78$) compared to partially labeled ($\alpha = .57$) response options when they reviewed 96 different measures of attitude from five, 3-wave US surveys. Once again, a seemingly simple change in response option format resulted in dramatically different results.

The findings specific to response options are consistent with a long history of cognitive psychology research. This research supports the assertion that the question and response wording, survey context, and physical format exert great influence on self-administered survey results, to the point that researchers can affect their results through changes in their survey design.⁷

In addition, there is a fairly consistent difference when respondents are asked to *rate* versus *rank* items. In the first option, respondents are asked to *rate* different items, and researchers then calculate a ranking based on the mean ratings. In the second option, respondents are explicitly asked to *rank* the items directly. Several studies have suggested that having respondents directly rank items provides stronger validity and reliability evidence than the researchers’ corresponding rating calculation.^{8–10} The key point, however, is that the specific research question should drive survey design, which includes such decisions as the use of ratings versus rankings.

In another example of the influence survey design can have on survey responses, Krosnick⁸ found that respondents tend to agree with statements more often than they disagree with them. In other words—all else being equal—respondents want to be agreeable. In a meta-analysis, 52% of people agreed with one assertion, while only 42% disagreed with its exact

opposite.⁸ Moreover, other work reveals that positively and negatively worded items about the same concept showed an average correlation of merely -0.22 .¹¹ One review estimated a 10% acquiescence effect—the effect of respondents simply agreeing because they want to be agreeable—a notable amount of variance to be explained simply by the way the items are worded.⁸

From the above examples—a mere sliver of an expansive literature—it becomes clear that surveys are the tangible embodiment of the myriad of complicating factors in social science research. When surveys are used as research instruments, they should be treated with scientific integrity. To maintain scientific integrity, the above survey design issues, as well as others identified in the past nearly half century, must be central to survey development. Yet, they are rarely part of the current conversation in medical education survey studies. In short, the rigor with which many surveys are created and administered is unsatisfactory,^{12,13} and can alter not only individual study findings, but also the degree to which surveys are accepted by the medical education research and publication community as worthy of dissemination and publication.

Some medical education journals are already beginning to push back against survey research. One emergency medicine journal publicly printed its distaste for surveys in the author instructions:

Annals “. . . only rarely publishes surveys given their potent methodological limitations. To be seriously considered, manuscripts describing surveys must show evidence of rigorous instrument development and testing, a high response rate, and a topic of unusual importance to emergency physicians.”¹⁴

Survey methods, applied to an appropriate research question and developed with proper rigor, can provide insights into human phenomena that other research methods cannot assess.¹⁵ However, there is considerable distrust of surveys currently, and health professions education researchers are primarily to blame.

Mark Twain famously popularized the saying, “There are 3 kinds of lies: lies, damned lies, and statistics.”¹⁶ He referred to his own difficulty in understanding figures, and to the idea that statistics can have persuasive power, even when used inappropriately. Statistics can be—and often are—used to bolster weak arguments. As such, many view statistics with skepticism. The same can be said for surveys. Results from a national survey on an important medical education topic, even if poorly designed and poorly executed, can have considerable persuasive

power. However, such survey studies damage the field by filling the literature, drip by drip, with unsubstantiated claims that may take years to correct.

As a scientific community, we run the risk of valuable data not being published in the foreseeable future because the only method capable of describing the phenomena might no longer be accepted, due to misuse and lack of scientific integrity. Editors, reviewers, and authors should understand that there is a systematic science behind writing, distributing, and analyzing surveys. Moreover, we all must recognize that quality control is essential, since writing poorly designed surveys is easy.

Qualitative research has become more accepted in education research. Its acceptance can be traced back to stricter definitions and clearer methods described in influential texts.^{17,18} Such texts already exist in the social sciences for survey design,^{19,20} and there is at least 1 introductory primer specific to medical education.²¹ Medical education researchers should become familiar with these works and deliberately apply evidence-based practices.

References

1. 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.
2. Artino AR Jr, Phillips AW, Utrankar A, et al. The questions shape the answers: assessing the quality of published survey instruments in health professions education. *Acad Med*. In press.
3. Yock Y, Lim I, Lim YH, et al. *Sometimes means some of the time*: residents' overlapping responses to vague quantifiers on the ACGME-I Resident Survey. *J Grad Med Educ*. 2017;9(6):735–740.
4. Bradburn NM, Miles C. Vague quantifiers. *Public Opinion Q*. 1979;43(1):92–101.
5. Krosnick JA, Berent MK. Comparisons of party identification and policy preferences: the impact of survey question format. *Am J Polit Sci*. 1993;37(3):941.
6. Alwin DF, Krosnick JA. The reliability of survey attitude measurement the influence of question and respondent attributes. *Sociol Methods Res*. 1991;20(1):139–181.
7. Schwarz N. Self-reports: how the questions shape the answers. *Am Psychol*. 1999;54(2):93–105.
8. Krosnick JA. Survey research. *Annu Rev Psychol*. 1999;50:537–567.
9. Miethe TD. The validity and reliability of value measurements. *J Psychol*. 1985;119(5):441–453.
10. Elig TW, Frieze IH. Measuring causal attributions for success and failure. *J Pers Soc Psychol*. 1979;37(4):621–634.

11. Krosnick JA, Fabrigar LR. *Designing Good Questionnaires*. New York, NY: Oxford University Press; 1998.
12. Fowler FJ Jr. *Survey Research Methods*. 5th ed. Thousand Oaks, CA: SAGE Publications Inc; 2013.
13. American Association for Public Opinion Research. Standard definitions. [http://www.aapor.org/Standards-Ethics/Standard-Definitions-\(1\).aspx](http://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx). Accessed September 20, 2017.
14. Annals of Emergency Medicine. Guidelines and preferences for specific research study designs. <http://www.annemergmed.com/content/designs>. Accessed September 20, 2017.
15. Phillips AW. Proper applications for surveys as a study methodology. *West J Emerg Med*. 2017;18(1):8–11.
16. Twain M. *Chapters from My Autobiography*. <http://www.gutenberg.org/files/19987/19987-h/19987-h.htm>. Published 1906. Accessed September 20, 2017.
17. Greenhalgh T, Hurwitz B. *Narrative Based Medicine*. London, UK: BMJ Books; 1998.
18. Hudelson PM. *Qualitative Research for Health Programmes*. Geneva: World Health Organization; 1994.
19. 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.
20. Fink A. *How to Conduct Surveys: A Step-by-Step Guide*. Google Books. 2012.
21. Artino AR, La Rochelle JS, Dezee KJ, et al. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach*. 2014;36(6):463–474.



Andrew W. Phillips, MD, MEd, is Adjunct Assistant Professor, Department of Emergency Medicine, University of North Carolina; and **Anthony R. Artino Jr, PhD**, is Professor and Deputy Director, Division of Health Professions Education, Department of Medicine, Uniformed Services University of the Health Sciences, and Deputy Editor, *Journal of Graduate Medical Education*.

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: Anthony R. Artino Jr, PhD, Uniformed Services University of the Health Sciences, Department of Medicine, Division of Health Professions Education, 4301 Jones Bridge Road, Bethesda, MD 20814, anthony.artino@usuhs.edu