

Threats to Reliability and Validity With Resident Wellness Surveying Efforts

Nital P. Appelbaum, PhD
Sally A. Santen, MD, PhD
Scott Vota, DO

Lauren Wingfield, MD
Roy Sabo, PhD
Nicholas Yaghmour, MPP

ABSTRACT

Background Residency programs and the Accreditation Council for Graduate Medical Education (ACGME) use survey data for the purpose of program evaluation. A priority for many programs is to improve resident wellness, often relying on self-reported surveys to drive interventions.

Objective We tested for result differences on wellness surveys collected through varying survey methodology and identified potential causes for differences.

Methods Aggregated results on the resident wellness scale for a single institution were compared when collected electronically through the ACGME Resident Survey immediately following the program evaluation survey for accreditation purposes and anonymously through an internal survey aimed at program improvement.

Results Across 18 residency programs, 293 of 404 (73%) residents responded to the internal survey, and 383 of 398 residents (96%) responded to the 2018 ACGME survey. There was a significant difference ($P < .001$, Cohen's $d = 1.22$) between the composite wellness score from our internal survey (3.69 ± 0.34) compared to its measurement through the ACGME (4.08 ± 0.30), indicating reports of more positive wellness on the national accreditation survey. ACGME results were also statistically more favorable for all 10 individual scale items compared to the internal results.

Conclusions Potential causes for differences in wellness scores between internal and ACGME collected surveys include poor test-retest reliability, nonresponse bias, coaching responses, social desirability bias, different modes for data collection, and differences in survey response options. Triangulation of data through multiple methodologies and tools may be one approach to accurately gauge resident wellness.

Introduction

Residency training programs rely on survey data to continuously improve medical education. Starting in 2004, the Accreditation Council for Graduate Medical Education (ACGME) annually distributed a web-based survey to accredited residency programs to systematically assess issues such as work hour compliance and adequacy of clinical supervision.¹ For the 2018 distribution of the ACGME Resident Survey, coinciding with revised Section VI of the Common Program Requirements,² and reflecting a growing concern about physician wellness, questions regarding resident wellness were added. Response data from these additional wellness items were provided to program and institutional leadership as a measure of resident wellness in their training programs and as a signaling mechanism for targeted intervention. As illustrated through the annual ACGME survey, graduate medical education uses surveys to assess values, beliefs, and perspectives of trainees. However, survey methodology is vulnerable to bias depending on a variety of methodological and psychological factors.^{3,4}

Prior to the original 2004 distribution of the ACGME survey, a number of researchers surveyed residents to gain insights into wellness, duty hours, retention, and learner and faculty perspectives on residency training.^{5–8} More recently, many residency program sponsoring institutions have surveyed their residents' wellness, morale, burnout, and other related constructs.⁹ Existing evidence suggests resident burnout links to poor patient outcomes, including self-reported medical errors,¹⁰ self-reported suboptimal patient care,¹¹ and changes in brain activity during clinical reasoning.¹² Solutions to improving physician wellness include organizational-level (eg, work hour restrictions) and individual-level (eg, mindfulness training, stress management, small group discussions) interventions.¹³ However, it remains unclear when certain interventions are applicable, for which groups, and in what combination.¹³ Accurate institutional-level and program-level data serve as a foundation to design effective interventions to improve resident wellness.

The objective of this study is to determine if there are differences in responses on the ACGME Resident Survey as compared to an internally administered wellness survey to examine potential threats to validity and reliability that can affect survey responses.

DOI: <http://dx.doi.org/10.4300/JGME-D-19-00216.1>

Methods

Internal Survey

Between May 1 and June 30, 2018, residents across 18 residency programs were internally surveyed at Virginia Commonwealth University Health, a large academic medical center in Central Virginia, to assess facets of culture and context. One of the scales on the internal survey was the 10-item Resident Wellness Scale (RWS),^{14,15} which was also the scale used by ACGME to assess wellness in 2018. Respondents indicated frequency of positive indicators of wellness on a 5-point Likert scale over the past 3 weeks (1, never; 2, seldom [on internal survey] or 2, rarely [on ACGME survey]; 3, sometimes; 4, often; and 5, very often). Both internal and ACGME surveys included a 5-point Likert scale, but the ACGME survey altered a response option from 2 as “seldom” to 2 as “rarely” to remain consistent with other program evaluation item response options, whereas the internal survey kept the original response option as 2 equals seldom.

Residents were recruited to complete the voluntary internal survey via paper format during a regularly scheduled meeting (eg, didactic conference) primarily restricted to residents. A medical education researcher with no supervisory duties or oversight of trainees read aloud a script detailing key information (eg, anonymity in responses, improvement purpose for surveying, protections during data reporting to ensure confidentiality) to recruit participation. Residents submitted completed or blank paper surveys to a large box to allow anonymity in response and participation. Residents were also e-mailed an anonymous electronic Qualtrics (Qualtrics LLC, Provo, UT) survey link, allowing participation from absent members.

ACGME Resident Survey

Between January 15 and April 15, 2018, the ACGME opened a mandatory national survey to all residents, including those who participated in our internal survey, for the purpose of annual program evaluation. Residency training programs were scheduled by ACGME in a staggered manner for data collection within 5-week windows during this time period. Program directors (PDs) were responsible for e-mailing their residents during the 5 weeks that their survey was open for completion and securing a minimum 70% to 100% response rate depending on program size. PDs did not have access to the survey questions or individual resident responses. Residents had discretion over where they completed the ACGME survey.

Following the accreditation-based electronic survey questions on the ACGME Resident Survey, introductory text was displayed explaining that data from the

What was known and gap

Residency programs and the Accreditation Council for Graduate Medical Education use survey data to improve education and gain insight into resident wellness. A more clear understanding of the validity and reliability problems that result from different survey methodologies can help improve the interventions that result from such surveys.

What is new

A comparison of results from a single institution's ACGME Resident Survey wellness scale with those of an internal survey on resident wellness.

Limitations

Both surveys were administered at a single institution, limiting generalizability.

Bottom line

Results from one institution's wellness scale collected as part of the ACGME Resident Survey and an internal survey on wellness varied considerably.

wellness items would not be provided to any ACGME residency review committees and would not be used to make accreditation decisions. Responses to the RWS questions were mandatory to progress and complete the ACGME survey. Items 9 and 10 included a “not applicable” response option on the ACGME survey, which was not offered on the internal survey. The reason ACGME included a “not applicable” option for items 9 and 10 was because some residents may not have interacted with a patient or had a tragic work incident within the past 3 weeks. Responses of “not applicable” were not counted into mean calculations for ACGME scoring.

Two survey administrations allowed the opportunity to examine potential threats to validity and reliability that affect survey responses on resident wellness. Despite both surveys being anonymous, residents may judge wellness items differently when administered through an internal source rather than the ACGME. In addition, variations between the survey administrations may also influence findings and interpretations on the current state of resident wellness.

The Virginia Commonwealth University Institutional Review Board deemed our internal survey and our study to compare program-level data with ACGME data exempt from review.

Analysis

Internal survey data were aggregated to the program level to link with the ACGME RWS results. The results were reported in aggregate by each training program with raw counts for each response option for the 10 RWS items. Since program means were not directly reported, manual calculation was conducted to determine a mean score for each item as well as a composite score based on the number of respondents

TABLE 1
Demographics of Internal Survey Respondents

Characteristic	Internal Survey Count, No. (%)	Population Count, No. (%)
Gender		
Male	168 (57)	
Female	105 (36)	
Unknown	20 (7)	
Race		
White	189 (65)	
Asian	45 (15)	
Black or African American	12 (4)	
Hispanic or Latino	6 (2)	
American Indian	1 (< 1)	
Other	14 (5)	
Unknown	26 (9)	
Training level		
PGY-1	49 (17)	105 (26)
PGY-2	76 (26)	107 (26)
PGY-3	69 (24)	99 (25)
PGY-4	48 (16)	57 (14)
PGY-5	22 (8)	22 (5)
PGY-6	4 (1)	6 (1)
PGY-7	3 (1)	6 (1)
PGY-8	0 (0)	1 (< 1)
Unknown	22 (8)	...

Note: Population comparison counts were from our office of graduate medical education database (n = 403), which did not include gender or comparable race/ethnicity data.

Abbreviation: PGY, postgraduate year.

for each response option and the number of respondents to the entire survey. Scale reliability and principal components analysis was conducted using SPSS 25 (IBM Corp, Armonk, NY), as well as 1-sample *t* tests to detect differences between the internal and ACGME data for both composite scores and each of the 10 scale items. Differences between paper and electronic results with the internal survey were compared through the Mann-Whitney U test.

Results

Of 404 residents, 293 (73%) responded to the internal survey (TABLE 1). The 2018 ACGME Resident

Survey on wellness had a higher response rate (96%, 383 of 398). Unlike the ACGME survey, our internal survey included residents who were in a 1-year preliminary spot without categorical affiliation. The majority of internal survey responses were collected through paper format (82%, 240 of 293) compared to electronic (18%, 53 of 293). There was no significant difference in composite RWS scores across residencies between paper and electronic formats for the internal survey ($P > .05$).

For our internal survey RWS results, the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.89, above the commonly recommended value of 0.60, the Bartlett's test of sphericity was significant ($\chi^2(45) = 1347.33$; $P < .001$; Cronbach's alpha = 0.88). Principal components analysis with varimax rotation results (TABLE 2) suggested adequate construct validity evidence. Similar analysis was not possible with the reported program-level ACGME wellness data.

There was a significant difference between the composite score for wellness collected through our internal survey compared to its measurement through the ACGME survey (TABLE 3). The overall composite score for RWS from ACGME was higher (4.08 ± 0.30), indicating more positive wellness on the national accreditation survey, compared to the overall composite score from our internal survey (3.69 ± 0.34). Resident wellness was also significantly more positive on the ACGME survey for all of the 10 individual RWS items compared to the internal survey results (TABLE 3). Further inspection of effect sizes showed a range of medium to very large effect sizes, suggesting a remarkable magnitude of difference between ACGME and internal survey results (TABLE 3).¹⁶ Mean composite RWS scores broken out by program (deidentified) also showed higher scores on ACGME compared to internal surveys for 15 of the 18 programs (TABLE 4). For the ACGME survey, 8% to 53% of residents within a program chose "not applicable" for item 10, whereas 0% to 44% of residents within a program chose "not applicable" for item 9.

Discussion

We found large differences in wellness scores from 2 anonymous resident surveys using the RWS

TABLE 2
Factor Analysis Results for Construct Validity

Measure	Model	Survey Items	Range of Loadings	Eigenvalue	Percentage of Variance Explained
Resident Wellness Scale	2-component model	1, 2, 4, 5, 6, 10	0.67–0.81	4.92	37.7%
		3, 7, 8, 9	0.56–0.81	1.19	23.4%

Note: Factor analysis was conducted only with individual-level internal survey data.

TABLE 3

Internal and ACGME Survey Comparisons on Assessment of Resident Wellness

Resident Wellness Scale ^{11,12}	Internal Survey Mean $\pm$ SD	ACGME Survey Mean $\pm$ SD	t Test	Cohen's d Effect Size
Item 1: Reflected on how your work helps make the world a better place	3.12 $\pm$ 0.36	3.54 $\pm$ 0.37	t(17) = -4.62; <i>P</i> < .001; 95% CI -0.61, -0.23	1.15 (large)
Item 2: Felt the vitality to do your work	3.54 $\pm$ 0.37	3.96 $\pm$ 0.42	t(17) = -4.25; <i>P</i> = .001; 95% CI -0.63, -0.21	1.06 (large)
Item 3: Felt supported by your coworkers	4.04 $\pm$ 0.27	4.42 $\pm$ 0.27	t(17) = -5.16; <i>P</i> < .001; 95% CI -0.54, -0.23	1.41 (very large)
Item 4: Were proud of the work you did	3.93 $\pm$ 0.37	4.16 $\pm$ 0.34	t(17) = -2.72; <i>P</i> = .014; 95% CI -0.41, -0.05	0.65 (medium)
Item 5: Were eager to come back to work the next day	3.45 $\pm$ 0.49	3.77 $\pm$ 0.51	t(17) = -3.44; <i>P</i> = .003; 95% CI -0.51, -0.12	0.64 (medium)
Item 6: You felt your basic needs are met	3.93 $\pm$ 0.40	4.21 $\pm$ 0.31	t(17) = -4.04; <i>P</i> = .001; 95% CI -0.44, -0.14	0.78 (medium)
Item 7: You ate well	3.52 $\pm$ 0.43	3.95 $\pm$ 0.37	t(17) = -5.25; <i>P</i> < .001; 95% CI -0.61, -0.26	1.07 (large)
Item 8: You felt connected to your work in a deep sense	3.52 $\pm$ 0.50	3.83 $\pm$ 0.40	t(17) = -2.92; <i>P</i> = .009; 95% CI -0.53, -0.09	0.68 (medium)
Item 9: Had an enjoyable interaction with a patient	4.00 $\pm$ 0.49	4.50 $\pm$ 0.31	t(17) = -5.27; <i>P</i> < .001; 95% CI -0.70, -0.30	1.22 (large)
Item 10: Knew who to call when something tragic happened at work	3.90 $\pm$ 0.39	4.44 $\pm$ 0.35	t(17) = -4.98; <i>P</i> < .001; 95% CI -0.76, -0.31	1.46 (very large)
Composite scale score	3.69 $\pm$ 0.34	4.08 $\pm$ 0.30	t(17) = -5.15; <i>P</i> < .001; 95% CI -0.54, -0.23	1.22 (large)

TABLE 4

Program-Level Means for Composite Resident Wellness Scale Score

Training Program	Internal Survey	ACGME Resident Survey
A	3.65	3.80
B	3.24	4.44
C	4.14	4.55
D	3.48	3.79
E	3.55	4.03
F	3.81	4.28
G	3.51	4.35
H	4.31	4.60
I	3.71	4.10
J	3.41	4.00
K	3.70	4.12
L	3.50	3.88
M	3.48	3.90
N	3.68	3.61
O	3.88	3.81
P	4.55	4.43
Q	3.28	3.84
R	3.59	3.87

Abbreviation: ACGME, Accreditation Council for Graduate Medical Education.

instrument with primarily the same subjects at a single institution; one survey was administered in winter/early spring and the other in late spring. The first administration of RWS was electronic through the ACGME Resident Survey, which produced significantly higher scores across all items in comparison with the second internally administered survey of RWS through paper and electronic formats.

Our study suggests a need for scrutiny when analyzing resident wellness data. Without a gold standard of measurement for wellness, it remains unknown which administration accurately represented the true state of wellness in our resident population. Graduate medical education leaders responsible for monitoring and addressing resident wellness should be aware of potential threats to validity and reliability of wellness data.

Factors that may explain our results include test-retest reliability, which refers to the degree in which results are stable when the same measurement tool is administered at 2 different time points with the same sample.¹⁷ There was a time lag between when the ACGME and internal surveys were collected, primarily to avoid survey fatigue. Therefore, one explanation for our findings may be due to poor test-retest reliability. The scale instructions also asked respondents to reflect on the last 3 weeks when responding to items. Our findings may be influenced by instability

in wellness during each administration's study period. In addition, 3 weeks may not be an adequate length of time to study the construct of wellness, which may be situation-dependent with fluctuations over time.¹⁸ Unfortunately, time effects were difficult to address in our study since we were unable to link ACGME and internal survey data at the individual level. Within-person studies of resident wellness would help determine whether external factors and time influence results.

Nonresponse bias, the extent that nonrespondents are different from respondents with surveying efforts, may also explain study findings.¹⁹ Such a bias can be of concern when large portions of the population choose not to participate in surveying efforts; however, low response rates are not always indicative of nonresponse bias.^{19–21} Response rates only explain a small amount of variance in nonresponse bias findings, therefore direct measurement of nonresponse bias is advised (eg, interest-level or passive nonresponse analysis, wave analysis, benchmarking, replication).^{19–22} Nonresponse bias can also be assessed by comparing population and sample characteristics to detect representation.²⁰ The ACGME survey did not provide respondent demographics, but respondents' training year data from our internal survey was similar to the distribution with our resident population (TABLE 1). Nonresponse bias was also minimized with the internal survey because attendance during the data collection sessions was random, and absent members were able to complete the survey via electronic format.

Social desirability, a response bias, refers to the process by which survey responses are influenced by the tendency for respondents to present themselves in a favorable manner, unintentionally or intentionally adjusting answers to the perceived ideal or correct responses.²³ This concept has been identified as a limitation of surveys and source of bias since the 1960s.²⁴ Given that ACGME is an accrediting body, the responses to the wellness items may be particularly susceptible to social desirability bias as respondents attempt to represent themselves and their training programs in a favorable manner. Residents' responses may be perceived as a threat to the accreditation status of their respective programs because the wellness items were collected alongside the accreditation items. New validity evidence needs to be determined with each survey administration considering the potential for unique subjects, settings, and purpose. Even though residents were told the wellness items on the ACGME survey would not be used for accreditation, respondents may not have fully grasped this point. In comparison, the internal

survey collected wellness data for the explicit purpose of program improvement.

In many programs, leadership meets with residents to help them understand the purpose of the ACGME survey. While "coaching" residents to respond favorably is not permitted, program directors have voiced concerns about the potential for ACGME survey items to be misinterpreted.²⁵ Program leadership may, intentionally or unintentionally, exacerbate social desirability bias by discussing the potential consequences to accreditation status based on certain response patterns. By clarifying the intent for surveying wellness with residents for program improvement, leadership can more accurately gauge wellness among their trainees. Emphasizing the purpose and scope of the ACGME wellness items may also moderate social desirability bias and reframe a perceived threat to accreditation.

Finally, and most likely the most influential drivers of our findings, different formatting and modes for data collection and different response options can impact reliability and validity of survey findings. Psychologists and social scientists have long studied the impact that wording, formatting, and response options can have on findings through cognitive and communicative processes when responding to survey items.²⁶ For example, optimizing refers to the series of complex cognitive steps enacted to respond to survey items: (1) Interpret the question and infer its intent; (2) Recall memories for relevant information; (3) Integrate recalled information to form a judgment; and (4) Translate the judgment by selecting a response option.²⁷ Accordingly, cognitive judgments can be largely dependent on what is literally presented to the respondent in the survey. Different modes for data collection and administration (eg, paper, electronic, in-person, voluntary/mandatory participation) can also influence survey responses.^{3,4} While our results found no significant difference in composite RWS scores between our internal paper and electronic survey results, there is still the potential for such differences to influence results. There is also evidence that anchor wording and response choice can influence survey results.²⁸ The internal and ACGME surveys had different response options for one of the scale anchors, and the ACGME survey added a "not applicable" response option for a few items, which could have influenced mean scores.²⁹

Our results caution the reliance on single or limited sources of data on resident wellness. Cross-sectional survey designs provide a snapshot of resident perception, and surveys at other times of the year may produce differing results depending on various factors. Therefore, survey results are only one

indicator of the health of a program. Areas of concern can receive additional clarification in follow-up efforts to collect data through anonymous or confidential methods. Program leadership can facilitate an open discussion with residents on wellness and an action plan for improvement based on multiple sources of data, including internal and ACGME survey results. A feedback loop of results with an action plan also signals an improvement-oriented reason for collecting such data compared to a passive act of annually collecting data on wellness without effort to improve the lives of trainees.

Based on our results, other institutions may want to triangulate their own internal data sources on wellness with ACGME results. Other applications of multi-source, multi-method data³⁰ for improving wellness include correlation between program evaluation data with internal measures of the clinical learning environment³¹ and deconstructing external reports to determine group differences in wellness.³² Next key research steps include integration of qualitative methodologies (eg, cognitive interviewing, focus groups) to understand the motivations behind response patterns, interpretations of scale items, and psychometric stability of the RWS based on specialty type.^{33,34}

Conclusions

Our study found large differences between responses on a wellness instrument collected through the ACGME compared to an internal survey, suggesting potential threats to validity and reliability in the measurement of resident wellness. Potential causes for differences include poor test-retest reliability, nonresponse bias, coaching responses, social desirability bias, different modes for data collection, and differences in survey response options.

References

- Holt KD, Miller RS. The ACGME resident survey aggregate reports: an analysis and assessment of overall program compliance. *J Grad Med Educ*. 2009;1(2):327–333. doi:10.4300/JGME-D-09-00062.1.
- Accreditation Council for Graduate Medical Education. ACGME Common Program Requirements Section VI with Background and Intent. https://www.acgme.org/Portals/0/PFAAssets/ProgramRequirements/CPRs_Section%20VI_with-Background-and-Intent_2017-01.pdf. Accessed August 26, 2019.
- Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J Public Health*. 2005;27(3):281–291. doi:10.1093/pubmed/fdi031.
- McMahon SR, Iwamoto M, Massoudi MS, Yusuf HR, Stevenson JM, David F, et al. Comparison of e-mail, fax, and postal surveys of pediatricians. *Pediatrics*. 2003;111(4 pt 1):e299–e303. doi:10.1542/peds.111.4.e299.
- Baldwin DC, Daugherty SR, Tsai R, Scotti MJ. A national survey of residents' self-reported work hours: thinking beyond specialty. *Acad Med*. 2003;78(11):1154–1163.
- Lieff SJ, Warshaw GA, Bragg EJ, Shaull RW, Lindsell CJ, Goldenhar LM. Geriatric psychiatry fellowship programs in the United States: findings from the Association of Directors of Geriatric Academic Programs' longitudinal study of training and practice. *Am J Geriatr Psychiatry*. 2003;11(3):291–299.
- Wheeler DS, Clapp CR, Poss WB. Training in pediatric critical care medicine: A survey of pediatric residency training programs. *Pediatr Emerg Care*. 2003;19(1):1–5.
- Niederee MJ, Knudtson JL, Byrnes MC, Helmer SD, Smith RS. A survey of residents and faculty regarding work hour limitations in surgical training programs. *Arch Surg*. 2003;138(6):663–669. doi:10.1001/archsurg.138.6.663.
- Raj KS. Well-Being in Residency: A Systematic Review. *J Grad Med Educ*. 2016;8(5):674–684. doi:10.4300/JGME-D-15-00764.1
- Fahrenkopf AM, Sectish TC, Barger LK, Sharek PJ, Lewin D, Chiang VW, et al. Rates of medication errors among depressed and burnt out residents: prospective cohort study. *BMJ*. 2008;336(7642):488–491. doi:10.1136/bmj.39469.763218.BE.
- Shanafelt TD, Bradley KA, Wipf JE, Back AL. Burnout and self-reported patient care in an internal medicine residency program. *Ann Intern Med*. 2002;136(5):358–367. doi:10.7326/0003-4819-136-5-200203050-00008.
- Durning SJ, Costanzo M, Artino AR, Dyrbye LN, Beckman TJ, Schuwirth L, et al. Functional neuroimaging correlates of burnout among internal medicine residents and faculty members. *Front Psychiatry*. 2013;4:131. doi:10.3389/fpsyt.2013.00131.
- West CP, Dyrbye LN, Erwin PJ, Shanafelt TD. Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. *Lancet Lond Engl*. 2016;388(10057):2272–2281. doi:10.1016/S0140-6736(16)31279-X.
- Stansfield RB, Giang D, Markova T. Development of the resident wellness scale for measuring resident wellness. *J Patient-Centered Res Rev*. 2019;6(1):17–27. doi:10.17294/2330-0698.1653.
- Wayne State University. Providing the Resident Wellness Scale for Broad, Open-Source Use. <http://www.gme.wayne.edu/wellness/RWSPProtocol.html>. Accessed August 5, 2019.
- Sullivan GM, Feinn R. Using effect size—or why the P value is not enough. *J Grad Med Educ*.

- 2012;4(3):279–282. doi:10.4300/JGME-D-12-00156.1.
17. Drost E. Validity and reliability in social science research. *Educ Res Perspect*. 2011;38:105–124.
 18. Pantaleoni JL, Augustine EM, Sourkes BM, Bachrach LK. Burnout in pediatric residents over a 2-year period: a longitudinal study. *Acad Pediatr*. 2014;14(2):167–172. doi:10.1016/j.acap.2013.12.001.
 19. Davern M. Nonresponse rates are a problematic indicator of nonresponse bias in survey research. *Health Serv Res*. 2013;48(3):905–912. doi:10.1111/1475-6773.12070.
 20. Halbesleben JRB, Whitman MV. Evaluating survey quality in health services research: a decision framework for assessing nonresponse bias. *Health Serv Res*. 2013;48(3):913–930. doi:10.1111/1475-6773.12002.
 21. Groves RM. Nonresponse rates and nonresponse bias in household surveys. *Public Opin Q*. 2006;70(5):646–675. doi:10.1093/poq/nfl033.
 22. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*. 2008;72(2):167–189. doi:10.1093/poq/nfn011.
 23. Krumpal I. Determinants of social desirability bias in sensitive surveys: a literature review. *Qual Quant*. 2013;47(4):2025–2047. doi:10.1007/s11135-011-9640-9.
 24. Phillips DL, Clancy KJ. Some effects of “social desirability” in survey studies. *Am J Sociol*. 1972;77(5):921–940.
 25. Adams M, Willett LL, Wahi-Gururaj S, Halvorsen AJ, Angus SV. Usefulness of the ACGME Resident Survey: a view from internal medicine program directors. *Am J Med*. 2014;127(4):351–355. doi:10.1016/j.amjmed.2013.12.010.
 26. Schwarz N. Self-reports: how the questions shape the answers. *Am Psychol*. 1999;54(2):93–105. doi:10.1037/0003-066X.54.2.93.
 27. Krosnick JA. Survey Research. *Annu Rev Psychol*. 1999;50(1):537–567. doi:10.1146/annurev.psych.50.1.537.
 28. Krosnick JA, Berent MK. Comparisons of party identification and policy preferences: the impact of survey question format. *Am J Polit Sci*. 1993;37(3):941–964. doi:10.2307/2111580.
 29. Krosnick JA. The causes of no-opinion responses to attitude measures in surveys: they rarely are what they appear to be. In: RM Groves, DA Dillman, JL Eltinge, RJA Little, eds. *Survey Nonresponse*. New York, NY: Wiley; 2002:88–100.
 30. Holmbeck GN, Li ST, Schurman JV, Friedman D, Coakley RM. Collecting and managing multisource and multimethod data in studies of pediatric populations. *J Pediatr Psychol*. 2002;27(1):5–18. doi:10.1093/jpepsy/27.1.5.
 31. Appelbaum NP, Santen SA, Aboff BM, Vega R, Munoz JL, Hemphill RR. Psychological safety and support: assessing resident perceptions of the clinical learning environment. *J Grad Med Educ*. 2018;10(6):651–656. doi:10.4300/JGME-D-18-00286.1.
 32. Adams PS, Gordon EKB, Berkeley A, Monroe B, Eckert JM, Maldonado Y, et al. Academic faculty demonstrate higher well-being than residents: Pennsylvania anesthesiology programs’ results of the 2017–2018 ACGME well-being survey. *J Clin Anesth*. 2019;56:60–64. doi:10.1016/j.jclinane.2019.01.037.
 33. Willis GB, Artino AR. 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. doi:10.4300/JGME-D-13-00154.1.
 34. 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. doi:10.4300/JGME-D-12-00239.1.



At the time of writing, **Nital P. Appelbaum, PhD**, was Assistant Professor, Virginia Commonwealth University School of Medicine, and is now Instructor, Baylor College of Medicine; **Sally A. Santen, MD, PhD**, is Senior Associate Dean, Assessment, Evaluation, and Scholarship, and Professor, Department of Emergency Medicine, Virginia Commonwealth University School of Medicine; **Scott Vota, DO**, is Residency Program Director, Vice Chair, Department of Education, and Associate Professor, Department of Neurology, Virginia Commonwealth University School of Medicine; **Lauren Wingfield, MD**, is Chief Resident, Emergency Medicine, Virginia Commonwealth University Health System; **Roy Sabo, PhD**, is Associate Professor, Department of Biostatistics, Virginia Commonwealth University; and **Nicholas Yaghmour, MPP**, is Associate Director, Well-Being and Milestones Research, Accreditation Council for Graduate Medical Education.

Funding: The authors report no external funding source for this study.

Conflict of interest: Dr. Santen receives funding for evaluating Accelerating Change in Medicine Education from the American Medical Association. Mr. Yaghmour is a paid employee of the Accreditation Council for Graduate Medical Education.

The authors would like to thank the Virginia Commonwealth University Health’s residency coordinators, program directors, and residents for participating and helping with our internal survey efforts.

This work was previously presented at the ACGME Educational Conference, Orlando, Florida, March 7–10, 2019.

Corresponding author: Nital P. Appelbaum, PhD, Baylor College of Medicine, M220.20, One Baylor Plaza, Houston, TX 77030, 713.798.6806, nital.appelbaum@bcm.edu

Received March 28, 2019; revisions received May 9, 2019, and July 29, 2019; accepted July 30, 2019.