

Assessing Resident Well-Being After the ABSITE: A Bad Time to Ask?

Elaine O. Cheung, PhD,* Yue-Yung Hu, MD, MPH,†‡ Andrew Jones, PhD,§ Meixi Ma, MD, MS,†||¶
Cary Jo R. Schlick, MD, MS,† Judith T. Moskowitz, PhD, MPH,* Gaurava Agarwal, MD,#
and Karl Y. Bilimoria, MD, MS†||

Objectives: Assess the association of residents' exam performance and transient emotions with their reports of burnout, suicidality, and mistreatment.

Background: An annual survey evaluating surgical resident well-being is administered following the American Board of Surgery In-Training Examination (ABSITE). One concern about administering a survey after the ABSITE is that stress from the exam may influence their responses.

Methods: A survey was administered to all general surgery residents following the 2018 ABSITE assessing positive and negative emotions (scales range from 0 to 12), as well as burnout, suicidality over the past 12 months, and mistreatment (discrimination, sexual harassment verbal/emotional or physical abuse) in the past academic year. Multivariable hierarchical regressions assessed the associations of exam performance and emotions with burnout, suicidality, and mistreatment.

Results: Residents from 262 programs provided complete responses (N = 6987, 93.6% response rate). Residents reported high mean positive emotion (M = 7.54, SD = 2.35) and low mean negative emotion (M = 5.33, SD = 2.43). While residents in the bottom ABSITE score quartile reported lower positive and higher negative emotion than residents in the top 2 and 3 quartiles, respectively (P < 0.005), exam performance was not associated with the reported likelihood of burnout, suicidality, or mistreatment.

Conclusions: Residents' emotions after the ABSITE are largely positive. Although poor exam performance may be associated with lower positive and higher negative emotion, it does not seem to be associated with the likelihood of reporting burnout, suicidality, or mistreatment. After adjusting for exam performance and emotions, mistreatment remained independently associated with burnout and suicidality. These findings support existing evidence demonstrating that burnout and suicidality are stable constructs that are robust to transient stress and/or emotions.

Keywords: burnout, emotion, exam performance, residency, wellbeing

From the *Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL; †Surgical Outcomes and Quality Improvement Center (SOQIC), Department of Surgery and Center for Health Services and Outcomes Research, Feinberg School of Medicine, Northwestern University, Chicago, IL; ‡Division of Pediatric Surgery, Ann & Robert H. Lurie Children's Hospital, Feinberg School of Medicine, Northwestern University, Chicago, IL; §American Board of Surgery, Philadelphia, PA; ||American College of Surgeons, Chicago, IL; ¶Department of Surgery, University of Alabama at Birmingham Medical Center, Birmingham, AL; and #Department of Psychiatry, Feinberg School of Medicine, Northwestern University, Chicago, IL.

Disclosure: The authors declare that they have nothing to disclose. This study is supported by funding from the American Board of Surgery (ABS), American College of Surgeons (ACS), and Accreditation Council for Graduate Medical Education (ACGME). E.O.C. was supported by a postdoctoral research fellowship (National Science Foundation [NSF Number 1714952]).

SDC Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Reprints: Elaine O. Cheung, PhD, Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, 625 N Michigan Ave, #2722, Chicago, IL 60611. Email: elainecheung333@gmail.com.

Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Annals of Surgery (2022) 4:e209

Received: 1 December 2021; Accepted: 14 August 2022

Published online 28 September 2022

DOI: 10.1097/AS9.000000000000209

INTRODUCTION

There has been considerable recent interest in characterizing burnout and well-being among physicians and physicians-in-training.¹⁻³ Much of the research on physician burnout and well-being has been limited by low response rates, small sample sizes, and single institution studies, which can threaten the generalizability and validity of findings.^{4,5} One approach that has demonstrated efficacy in achieving high response rates is to administer surveys following mandatory, scheduled functions, such as exams.^{2,4,6,7} One of the most comprehensive surveys assessing burnout and well-being is administered annually after the American Board of Surgery In-Training Examination (ABSITE).^{2,6,7} The 2018 ABSITE survey achieved a near-complete response rate (99.3%); documented high rates of burnout (38.5%), suicidal thoughts (4.5%), and mistreatment (>50% reporting some form of mistreatment) among general surgery residents; and found that exposure to mistreatment (discrimination, sexual harassment verbal/emotional or physical abuse) in residency was associated with burnout and suicidal thoughts.²

However, one potential limitation of this approach is that residents' survey responses may be influenced by their emotional state, having just completed an intensive, up-to-5-hour exam with the potential to impact their standing within their program and their future fellowship opportunities. Previous research has documented an association between undergraduate students' exam performance and their post-exam emotions, with those who perform poorly reporting higher levels of negative emotions and lower levels of positive emotions following the exam.^{8,9} Additionally, previous experimental studies with undergraduate student samples have found that respondents

who receive a negative mood induction (eg, recalling a negative life event, responding on a rainy day) prior to completing a survey are more likely to evaluate their subjective well-being (eg, judgments of happiness and satisfaction with life) unfavorably relative to those who receive a positive mood induction (eg, recalling a positive life event, responding on a sunny day) or no mood induction.^{10–13} Thus, it is possible that completing a stressful exam may similarly put residents in a negative emotional state prior to completing the survey, which may influence their responses to questions assessing their burnout and well-being.

The current study used data from the 2018 ABSITE survey to evaluate the possibility that participants' survey responses were influenced by factors such as their exam performance and emotions. Specifically, the current study assessed the positive and negative emotions that residents reported following the 2018 ABSITE and examined the associations of residents' exam performance and emotions with the likelihood of reporting burnout, suicidal thoughts, and mistreatment (discrimination, sexual harassment verbal/emotional or physical abuse). In addition, we examined the associations between reported mistreatment and reported burnout and/or suicidal thoughts, after adjusting for ABSITE performance and emotions.

METHODS

Residents from 262 general surgery residency programs were administered a voluntary, electronic survey immediately following the January 2018 ABSITE.² The survey was preceded by a statement explaining that the purpose of the survey was research, participation in the survey was voluntary, and data would be deidentified before analysis.^{2,6} The Northwestern University Institutional Review Board reviewed the study and determined that it was exempt from human subjects review.

The survey was developed using previously published instruments wherever possible.^{2,14–16} Pretesting and iterative refinement via cognitive interviews and pilot tests were undertaken with general surgery residents from multiple institutions prior to the final survey administration.^{2,6}

Measures

Exam Performance

Exam performance was measured using respondents' standard scores on the 2018 ABSITE and categorized in quartiles (100–439, 440–517, 518–574, 575–750).

Emotions

Emotions were measured using a previously validated 6-item emotion scale,¹⁵ which examined the frequency that respondents reported experiencing 3 positive (happy, excited, content) and 3 negative emotions (worried, irritable/angry, sad) in the past week, using a scale from 0 = never to 4 = always. Continuous sum scores for positive and negative emotions (each scale ranging from 0 to 12) were calculated.

Burnout

Burnout was assessed using a modified, 6-item abbreviated version of the Maslach Burnout Inventory,^{2,14} measuring symptoms of emotional exhaustion and depersonalization. Prior studies on physician burnout have primarily conceptualized burnout using the emotional exhaustion and depersonalization subscales alone,^{2,17–19} thus we did not utilize the 3-item personal accomplishment subscale. In addition, we modified the response scale from the original 7-point scale to a 5-point scale (never, a few times a year, a few times a month, a few times a week, every day). Respondents were classified as burned out if they reported

symptoms of emotional exhaustion or depersonalization a few times a week or more.^{2,17–19}

Suicidal Thoughts

Suicidal thoughts were assessed using a single-item question asking respondents, "During the past 12 months, have you had thoughts of taking your own life?"¹⁶

Mistreatment

Residents reported how frequently they experienced gender discrimination, racial discrimination, discrimination based on past/present/expected pregnancy and/or childcare needs, sexual harassment, physical abuse, and verbal or emotional abuse since the beginning of residency, with response options of never, a few times a year, a few times a month, a few times a week, every day. Responses were categorized as never, a few times a year, and more than a few times a year.² Consistent with prior research,² we calculated a composite mistreatment variable representing the maximum reported frequency of any mistreatment exposure (discrimination based on gender, race, or pregnancy and/or childcare; sexual harassment; physical, verbal, or emotional abuse).

Resident and Program Characteristics

Resident characteristics included gender, clinical postgraduate year level (categorized as postgraduate year 1, 2/3, and 4/5), and relationship status (categorized as married/in a relationship, not in a relationship, or divorced/widowed).² Program characteristics included program size (categorized in quartiles: <26, 26–37, 38–51, >51 residents), program type (university-based, independent, or military), and geographic location of program (Northeast, Southeast, Midwest, Southwest, West).²

Residents were also asked to answer questions the number of months in which they violated the 80-hour per week duty-hour limit over the past 6 months, categorized as 0–2 and 3+ months.

Statistical Analysis

Descriptive summary statistics were used to characterize resident demographic and program characteristics, exam performance, and positive and negative emotions.

The associations of exam performance with positive and negative emotions were assessed using multivariable linear mixed-effects models, accounting for resident clustering within programs. We regressed positive and negative emotions, in separate models, on exam performance and resident and program characteristics.

To examine the associations of exam performance with other resident well-being outcomes (burnout, suicidal thoughts, and mistreatment), we conducted multivariable hierarchical logistic regression models, regressing each outcome on exam performance and resident and program characteristics.

All statistical analyses were performed using IBM SPSS version 26.0 (IBM Corp., Armonk, NY).

RESULTS

Of 7464 clinically active residents, 6972 residents (93.4%) had complete survey responses. Descriptive statistics summarizing the sample characteristics are presented in Table 1. Residents' ABSITE scores (standard score) ranged from 100 to 746 (mean and standard deviation: $M = 499.32$, $SD = 100.74$). Residents' positive emotion scores ranged from 0 to 12 ($M = 7.54$, $SD = 2.35$). Residents' negative emotion scores ranged from 0 to 12 ($M = 5.33$, $SD = 2.43$).

TABLE 1.
Sample Characteristics

Variable	n (%) (N = 6972)
Gender	
Male	4178 (59.9)
Female	2794 (40.1)
Clinical PGY	
1	1968 (28.2)
2/3	2697 (38.7)
4/5	2307 (33.1)
Relationship status	
Married/relationship	5149 (73.9)
No relationship	1704 (24.4)
Divorced/widowed	119 (1.7)
Program size (total number of residents)	
Quartile 1 (<26)	1928 (27.7)
Quartile 2 (26–37)	1718 (24.6)
Quartile 3 (38–51)	1629 (23.4)
Quartile 4 (>51)	1697 (24.3)
Program type	
Academic	4361 (62.6)
Community	2395 (34.4)
Military	216 (3.1)
Program location	
Northeast	2349 (33.7)
Southeast	1350 (19.4)
Midwest	1519 (21.8)
Southwest	801 (11.5)
West	953 (13.7)
80-hour violations	
0–2 months	956 (13.7)
3+ months	6016 (86.3)
ABSITE exam performance	
Quartile 1 (100–439)	1837 (26.3)
Quartile 2 (440–517)	1627 (23.3)
Quartile 3 (518–574)	1723 (24.7)
Quartile 4 (575–750)	1785 (25.6)

PGY indicates postgraduate year.

Association of Exam Performance With Emotions

Positive emotion scores were lower among residents scoring in the bottom ABSITE quartile (quartile 1 model-predicted mean estimate and standard error: $M = 6.68, SE = 0.12$) relative to residents scoring in the top 2 quartiles (quartile 3: $M = 7.02, SE = 0.12$; quartile 4: $M = 7.24, SE = 0.12$; Table 2). Moreover, negative emotion scores were higher among residents scoring in the bottom ABSITE quartile (quartile 1: $M = 6.22, SE = 0.12$) relative to residents in the top 3 ABSITE quartiles (quartile 2: $M = 5.97, SE = 0.12$; quartile 3: $M = 5.85, SE = 0.12$; quartile 4: $M = 5.56, SE = 0.12$; Table 2).

Association of Exam Performance With Resident Well-Being Outcomes

There were no associations of exam performance with well-being outcomes (Table 3). Residents scoring in the bottom ABSITE quartile did not differ in their likelihood of reporting burnout relative to residents who scored in the top 3 quartiles (quartile 2 odds ratio [OR] and 95% confidence interval: OR = 0.98, 0.85–1.12; quartile 3: OR = 0.84, 0.78–1.12; quartile 4: OR = 0.83, 0.69–1.02). In addition, residents who scored in the bottom ABSITE quartile did not differ in the likelihood of reporting suicidal thoughts relative to residents who scored in the top 3 quartiles (quartile 2: OR = 0.89, 0.61–1.30; quartile 3: OR = 0.85, 0.56–1.31; quartile 4: OR = 0.72, 0.47–1.12). Finally, residents who scored in the bottom ABSITE quartile did not differ in the likelihood of reporting mistreatment (discrimination, sexual harassment verbal/emotional or physical abuse) relative to residents who scored in the top 3 quartiles (quartile 2: OR = 0.97,

TABLE 2.
Associations of ABSITE Performance With Positive and Negative Emotions

Variable	M (SE)	b (95% CI)	P
Positive emotion			
ABSITE performance			
Quartile 1 (100–439)	6.68 (0.12)	REF	REF
(lowest scores)			
Quartile 2 (440–517)	6.83 (0.12)	0.15 (–0.02 to 0.32)	0.09
Quartile 3 (518–574)	7.02 (0.12)	0.34 (0.15–0.53)	<0.001
Quartile 4 (575–750)	7.24 (0.12)	0.56 (0.35–0.76)	<0.001
(highest scores)			
Negative emotion			
ABSITE performance			
Quartile 1 (100–439)	6.22 (0.12)	REF	REF
(lowest scores)			
Quartile 2 (440–517)	5.97 (0.12)	–0.25 (–0.42 to –0.07)	0.005
Quartile 3 (518–574)	5.85 (0.12)	–0.36 (–0.55 to –0.17)	<0.001
Quartile 4 (575–750)	5.56 (0.12)	–0.65 (–0.86 to –0.45)	<0.001
(highest scores)			

Models adjusted for gender, PGY level, relationship status, program size, program type, geographic location, and duty hour violations.

b indicates unstandardized beta coefficient; CI, confidence interval; M, model-predicted estimated marginal mean; PGY, postgraduate year; SE, standard error of the mean.

TABLE 3.
Associations of ABSITE Performance With the Reported Likelihood of Burnout, Suicidal Thoughts, and Mistreatment

Variable	n (%)	Odds Ratio (95% CI)	P
Burnout			
ABSITE performance			
Quartile 1 (100–439) (lowest scores)	759 (41.3)	—	—
Quartile 2 (440–517)	655 (40.3)	0.98 (0.85–1.12)	0.75
Quartile 3 (518–574)	660 (38.3)	0.94 (0.78–1.12)	0.49
Quartile 4 (575–750) (highest scores)	629 (35.2)	0.83 (0.69–1.02)	0.07
Suicidal thoughts			
ABSITE performance			
Quartile 1 (100–439) (lowest scores)	96 (5.2)	—	—
Quartile 2 (440–517)	75 (4.6)	0.89 (0.61–1.30)	0.54
Quartile 3 (518–574)	76 (4.4)	0.85 (0.56–1.31)	0.46
Quartile 4 (575–750) (highest scores)	65 (3.6)	0.72 (0.47–1.12)	0.15
Mistreatment (discrimination, sexual harassment verbal/emotional or physical abuse)			
ABSITE performance			
Quartile 1 (100–439) (lowest scores)	900 (49.0)	—	—
Quartile 2 (440–517)	850 (52.2)	0.97 (0.82–1.15)	0.74
Quartile 3 (518–574)	919 (53.3)	1.00 (0.83–1.20)	0.97
Quartile 4 (575–750) (highest scores)	900 (50.4)	0.92 (0.75–1.12)	0.38

Models adjusted for gender, PGY level, relationship status, program size, program type, geographic location, and duty hour violations. Adjustment for emotion is not included.

CI indicates confidence interval; PGY, postgraduate year.

0.82–1.15; quartile 3: OR = 1.00, 0.83–1.20; quartile 4: OR = 0.92, 0.75–1.12).

DISCUSSION

The current study used data from the 2018 post-ABSITE survey to examine the possibility that residents’ survey responses were influenced by their exam performance and/or their transient emotional state. Despite concerns about exam-related stress, residents generally reported high mean positive emotion and low mean negative emotion following the ABSITE. This finding, suggesting higher levels of happiness than distress, is consistent with previous research in undergraduate samples that has found that students commonly report experiencing positive emotions (eg, happiness, relief) following an exam.^{8,9}

Residents who scored in the bottom quartile reported lower mean positive emotion scores than residents who scored in the

top 2 quartiles and higher mean negative emotion scores than residents who scored in the top 3 quartiles. Nonetheless, the current study found no associations between residents' exam performance and the likelihood of reporting burnout, suicidal thoughts, or mistreatment, supporting existing data demonstrating that these constructs are stable to situational context and transient emotions. These findings are consistent with previous research demonstrating that measures of burnout, suicidal thoughts, and well-being are stable and enduring,^{20–27} demonstrating longitudinal invariance over time. Moreover, the association between short term fluctuations in emotions and well-being in daily life is small and relatively inconsequential, implying a lack of susceptibility to momentary, situational factors.^{28–30} These findings suggest that measures of burnout and well-being capture stable aspects of respondents' quality of life and should be robust against the situational context in which these measures are administered.

Previous research using survey data from 2018 ABSITE found that exposure to mistreatment (discrimination, sexual harassment verbal/emotional or physical abuse) in residency was associated with burnout and suicidal thoughts.² Notably, these additional supplementary analyses adjusting for residents' ABSITE performance and emotions demonstrated that mistreatment exposure (discrimination, sexual harassment verbal/emotional or physical abuse) remained independently associated with greater reported likelihood of burnout and suicidal thoughts. This finding highlights the effects of mistreatment on well-being, above and beyond residents' emotions or their exam performance (Supplemental Table 1, <http://links.lww.com/AOSO/A170>).

The study should be interpreted within the context of its limitations: (1) It is possible that there are other potential variables that may have influenced the association between exam performance and/or post-exam emotion and long-term well-being that were not assessed here. For example, residents' perceptions of the stakes attached their ABSITE scores, which may have moderated the association between performance or emotion and well-being, were not assessed in the current study. (2) Additionally, we relied on residents' exam performance as a proxy for how the exam context may have influenced residents' evaluations of their well-being; we did not assess residents' "perceptions" of their performance, which would have provided a more direct measure of residents' mindsets following the exam. Given that we found that residents in the bottom ABSITE score quartile reported higher levels of negative emotion and lower levels of positive emotion following the exam, we believe that residents' exam performance should serve as a reasonable proxy for their perceptions about their performance. (3) The emotion measure in the study assesses respondents' emotions experienced in the "past week," which may not have reflected their current, momentary emotions following the ABSITE. However, previous research on the "peak-end rule" has demonstrated that when people are asked to retrospectively evaluate their emotions over a short time period (ie, a few weeks or less), they tend to overweigh the emotion experienced at the "peak" emotion intensity and the emotion experience at the "end" of the time period.^{31,32} As such, it is likely that the emotion measure used may indeed have reflected residents' post-exam emotional state, as the exam is likely both a "peak" intensity experience and occurs at the "end" of the time period preceding the survey. (4) We modified the response scale of the burnout measure (abbreviated Maslach Burnout Inventory) from a 7-point scale to a 5-point scale, which may limit the comparability of the current findings with previous research. Nevertheless, psychometric research suggests equivalency of data characteristics when changing between 5- and 7-point response scale formats.^{33,34} (5) Finally, the current study relies on a cross-sectional design, which precludes causal inferences and makes it difficult to determine the directionality of effects. Although the current study assumes that momentary situational factors, such as residents' exam performance and

emotions, influenced their responses when evaluating burnout and well-being, it is also possible that residents' burnout and well-being may have influenced their exam performance and emotions. Without collecting repeated assessments of resident burnout and well-being across multiple time points for comparison, it is difficult to fully rule out the possibility that situational factors in the exam context may influence responses to the survey. Further study is warranted to draw more definitive conclusions regarding the associations among exam performance, post-exam emotion, and well-being.

The current findings provide preliminary evidence suggesting that surveys evaluating resident well-being may be reliably administered after an intensive, high-stakes exam, such as the ABSITE. The Surgical Education Culture Optimization through targeted interventions based on National Comparative Data Trial,³⁵ a prospective, pragmatic, cluster-randomized trial, will use this survey administration approach to collect data on residents' perceptions of the learning environment and their well-being. Enrolled programs will receive aggregated deidentified reports of their residency program's performance on various resident well-being metrics. The efficacy of this intervention will be evaluated by assessing changes in program-level burnout and well-being. Although findings from the current study provide preliminary support for the validity of using the post-ABSITE survey to assess resident well-being in the Surgical Education Culture Optimization through targeted interventions based on National Comparative Data Trial, further study is needed to make more definitive conclusions.

CONCLUSIONS

It is important for residency programs to be able to accurately measure well-being in their residents. The current study offers initial support that previously validated measures of burnout and suicidality may indeed capture stable, enduring aspects of residents' quality-of-life that are robust to transient stress and/or emotion related to concurrent exam administration.

REFERENCES

1. National Academies of Sciences, Engineering, and Medicine; National Academy of Medicine; Committee on Systems Approaches to Improve Patient Care by Supporting Clinician Well-Being. Taking Action Against Clinician Burnout: A Systems Approach to Professional Well-Being. National Academies Press; 2020.
2. Hu YY, Ellis RJ, Hewitt DB, et al. Discrimination, abuse, harassment, and burnout in surgical residency training. *N Engl J Med*. 2019;381:1741–1752.
3. Shanafelt TD, Bradley KA, Wipf JE, et al. Burnout and self-reported patient care in an internal medicine residency program. *Ann Intern Med*. 2002;136:358–367.
4. Yarger JB, James TA, Ashikaga T, et al. Characteristics in response rates for surveys administered to surgery residents. *Surgery*. 2013;154:38–45.
5. VanGeest JB, Johnson TP, Welch VL. Methodologies for improving response rates in surveys of physicians: a systematic review. *Eval Health Prof*. 2007;30:303–321.
6. Bilimoria KY, Chung JW, Hedges LV, et al. National cluster-randomized trial of duty-hour flexibility in surgical training. *N Engl J Med*. 2016;374:713–727.
7. Zhang LM, Ellis RJ, Ma M, et al. Prevalence, types, and sources of bullying reported by US general surgery residents in 2019. *JAMA*. 2020;323:2093–2095.
8. Smith CA, Ellsworth PC. Patterns of appraisal and emotion related to taking an exam. *J Pers Soc Psychol*. 1987;52:475–488.
9. Folkman S, Lazarus RS. If it changes it must be a process: study of emotion and coping during three stages of a college examination. *J Pers Soc Psychol*. 1985;48:150–170.
10. Yardley JK, Rice RW. The relationship between mood and subjective well-being. *Soc Indic Res*. 1991;24:101–111.
11. Schwarz N, Clore GL. Mood as information: 20 years later. *Psychol Inq*. 2003;14:296–303.

12. Schwarz N, Clore GL. Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *J Pers Soc Psychol.* 1983;45:513.
13. Levine LJ, Safer MA. Sources of bias in memory for emotions. *Curr Dir Psychol Sci.* 2002;11:169–173.
14. McManus IC, Winder BC, Gordon D. The causal links between stress and burnout in a longitudinal study of UK doctors. *Lancet.* 2002;359:2089–2090.
15. Steptoe A, Wardle J. Positive affect measured using ecological momentary assessment and survival in older men and women. *Proc Natl Acad Sci U S A.* 2011;108:18244–18248.
16. Shanafelt TD, Balch CM, Dyrbye L, et al. Special report: suicidal ideation among American surgeons. *Arch Surg.* 2011;146:54–62.
17. West CP, Dyrbye LN, Sloan JA, et al. Single item measures of emotional exhaustion and depersonalization are useful for assessing burnout in medical professionals. *J Gen Intern Med.* 2009;24:1318–1321.
18. West CP, Dyrbye LN, Satele DV, et al. Concurrent validity of single-item measures of emotional exhaustion and depersonalization in burnout assessment. *J Gen Intern Med.* 2012;27:1445–1452.
19. Shanafelt TD, Boone S, Tan L, et al. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. *Arch Intern Med.* 2012;172:1377–1385.
20. Mäkikangas A, Hättinen M, Kinnunen U, et al. Longitudinal factorial invariance of the Maslach Burnout Inventory-General Survey among employees with job-related psychological health problems. *Stress Health.* 2011;27:347–352.
21. de Beurs DP, Fokkema M, de Groot MH, et al. Longitudinal measurement invariance of the Beck Scale for Suicide Ideation. *Psychiatry Res.* 2015;225:368–373.
22. Kim H, Ji J. Factor structure and longitudinal invariance of the Maslach Burnout Inventory. *Res Soc Work Pract.* 2009;19:325–339.
23. Boersma K, Lindblom K. Stability and change in burnout profiles over time: a prospective study in the working population. *Work Stress.* 2009;23:264–283.
24. Tyssen R, Vaglum P, Grønvold NT, et al. Suicidal ideation among medical students and young physicians: a nationwide and prospective study of prevalence and predictors. *J Affect Disord.* 2001;64:69–79.
25. Fujita F, Diener E. Life satisfaction set point: stability and change. *J Pers Soc Psychol.* 2005;88:158–164.
26. Sonnenschein M, Mommersteeg PM, Houtveen JH, et al. Exhaustion and endocrine functioning in clinical burnout: an in-depth study using the experience sampling method. *Biol Psychol.* 2007;75:176–184.
27. Schuler KR, Smith PN, Rufino KA, et al. Examining the temporal stability of suicide capability among undergraduates: a latent growth analysis. *J Affect Disord.* 2021;282:587–593.
28. Eid M, Diener E. Global judgments of subjective well-being: situational variability and long-term stability. *Soc Indic Res.* 2004;65:245–277.
29. Jayawickreme E, Tsukayama E, Kashdan TB. Examining the effect of affect on life satisfaction judgments: a within-person perspective. *J Res Pers.* 2017;68:32–37.
30. Yap SCY, Wortman J, Anusic I, et al. The effect of mood on judgments of subjective well-being: nine tests of the judgment model. *J Pers Soc Psychol.* 2017;113:939–961.
31. Fredrickson BL. Extracting meaning from past affective experiences: the importance of peaks, ends, and specific emotions. *Cogn Emot.* 2000;14:577–606.
32. Geng X, Chen Z, Lam W, Zheng Q. Hedonic evaluation over short and long retention intervals: the mechanism of the peak–end rule. *J Behav Decis Mak.* 2013;26:225–236.
33. Dawes J. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *Int J Mark Res.* 2008;50:61–104.
34. Colman AM, Norris CE, Preston CC. Comparing rating scales of different lengths: equivalence of scores from 5-point and 7-point scales. *Psychol Rep.* 1997;80:355–362.
35. Hu YY, Bilimoria, KB. The Surgical Education Culture Optimization Through Targeted Interventions Based on National Comparative Data (SECOND) Trial. November 14, 2018. Accessed July 20, 2020. <http://www.thesecondtrial.org>