Pursuing Excellence

The Power of Selection Science to Provide Meaningful Data and Enhance Efficiency in Selecting Surgical Trainees

Aimee K. Gardner, PhD† and Brian J. Dunkin, MD†‡*

Introduction: As current screening methods for selecting surgical trainees are receiving increasing scrutiny, development of a more efficient and effective selection system is needed. We describe the process of creating an evidence-based selection system and examine its impact on screening efficiency, faculty perceptions, and improving representation of underrepresented minorities.

Methods: The program partnered with an expert in organizational science to identify fellowship position requirements and associated competencies. Situational judgment tests, personality profiles, structured interviews, and technical skills assessments were used to measure these competencies. The situational judgment test and personality profiles were administered online and used to identify candidates to invite for on-site structured interviews and skills testing. A final rank list was created based on all data points and their respective importance. All faculty completed follow-up surveys regarding their perceptions of the process. Candidate demographic and experience data were pulled from the application website.

Results: Fifty-five of 72 applicants met eligibility requirements and were invited to take the online assessment, with 50 (91%) completing it. Average time to complete was 42 ± 12 minutes. Eighteen applicants (35%) were invited for on-site structured interviews and skills testing—a greater than 50% reduction in number of invites compared to prior years. Time estimates reveal that the process will result in a time savings of 68% for future iterations, compared to traditional methodologies. Fellowship faculty ($N = 5$) agreed on the value and efficiency of the process. Underrepresented minority candidates increased from an initial 70% to 92% being invited for an interview and ranked using the new screening tools.

Discussion: Applying selection science to the process of choosing surgical trainees is feasible, efficient, and well-received by faculty for making selection decisions.

Keywords: applicant, diversity, fellowship, selection, situational judgment test, underrepresented minority

(*Ann Surg* 2019;270:188–192)

Recent estimates suggest that approximately 80% of general surgery residents choose to pursue advanced surgical training.[1] These fellowships require trainees to hone advanced surgical techniques, master new technologies, and develop the ability to work independently, all within a short timeframe of 1 to 2 years. Given the abbreviated amount of time to ensure trainees reach uniform competency, programs must have effective screening processes in place to identify which fellowship applicants will thrive in their program.

Unfortunately, the current selection system for identifying high-potential fellowship candidates is burdensome for all parties involved. Estimates suggest that applicants spend from $4000 to $10,000 on travel for fellowship interviews,[2–5] with a third of all applicants borrowing money to pay for these trips.[5] In addition, the institution in which these applicants are currently being trained must manage 1 to 2 weeks of resident absences related to time away for interviews.[5] Finally, traditional screening methodologies, which entail review of applicant examination scores, letters of recommendation, and personal statements to determine who is invited for an on-site interview, are increasingly inefficient,[6] can cost programs up to $1000 per interview,[7,8] and have been criticized for their potential of bias and inequality.[9,10] These aforementioned concerns, combined with the lack of validity evidence for current screening tools to predict performance in surgical training,[6,11–13] have prompted both surgical educators[6,9,14–16] and applicants[5] to favor changes in the current selection process.

The goal of this study is to explore how an evidence-based selection system can provide objective data for decision makers. We describe the impact of this process on applicant participation, screening efficiency, and faculty perceptions. We also explore how adoption of best practices in selection can promote more equitable consideration and provision of surgery training positions to those applying from underrepresented demographic groups.

METHODS

This study was conducted within a fellowship program seeking to fill 2 one-year positions concentrating on minimally invasive surgery and bariatric surgery. The program partnered with an expert in organizational science to conduct a job analysis, in which interviews and focus groups are conducted with key stakeholders to identify position requirements and identify requisite knowledge, skills, and abilities to fulfill those requirements.

Based on the job analysis, assessment tools were chosen that would measure the desired and required competencies in the most efficient, effective, and objective manner. These included situational judgment tests (SJTs), personality profiles, interviews, and technical skills assessments. A multiple hurdle approach, in which applicants complete assessment tools in consecutive rounds, was chosen to maximize the potential of the assessments and increase administrative efficiency. Specifically, after applications were reviewed for eligibility for fellowship (eligibility for state licensure and institutional credentialing, visa status, etc), all eligible applicants were directed to an online assessment platform in which they completed a 26-item SJT and 108-item self-report personality profile. Applicants

From the *School of Health Professions, Department of Surgery, Baylor College of Medicine, Houston, TX; †SurgWise Consulting, Houston, TX; and ‡Department of Surgery, Houston Methodist Hospital, Houston, TX.

A.G. and B.D. provide advice on assessment and selection methodology through SurgWise Consulting.

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: Aimee K. Gardner, PhD, Assistant Dean of Evaluation and Research, Associate Professor of Health Professions, Surgery, Surgery, Baylor College of Medicine, Baylor College of Medicine, MS: BCM115, DeBakey Bldg, M108K, One Baylor Plaza, Houston, TX 77030.
E-mail: aimee.gardner@bcm.edu.

Copyright © 2018 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0003-4932/18/27001-0188
DOI: 10.1097/SLA.0000000000002806

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

were instructed to complete the assessment in one sitting and that there would be a 60-minute time limit.

The SJT provided applicants with vignettes of scenarios likely to be encountered on the job and asked how they would manage these problems or situations by providing ratings indicating the extent to which they deemed each response item effective on a 1 (not effective at all) to 5 (very effective) scale (see example in Appendix; http://links.lww.com/SLA/B421). Development of the SJT and scoring key followed best practices described elsewhere.[12,17] Results from this assessment were compared to the faculty-derived scoring key, such that assessment scores reflect the extent to which applicant decisions match those recommended by faculty members.

The personality profile used was a proprietary questionnaire[18] with extensive validity evidence in which applicants are provided with statements and asked to indicate the extent to which each item characterized them on a 1 (strongly disagree) to 5 (strongly agree) Likert scale. Results from the assessment tool provide information about applicants' conscientiousness, openness to new experiences, extraversion, agreeableness, and neuroticism along a spectrum from low to high.

In an effort to compare the efficiency of the new selection system to traditional screening processes, we also asked leadership to review the applications as they normally would. To achieve this aim, before receiving the results of the online assessments, the fellowship Program Director and Associate Program Director independently reviewed all applications as they traditionally have done in past years for interview selection. Specifically, they examined the data available to them in the standardized Fellowship Council online application form, which includes letters of recommendation, examination scores, case logs, and personal statements and created an ''invite / do-not-invite'' list.

The results of the SJT and personality profile were reviewed by the organizational scientist and applicants were grouped into green, yellow, and red categories, indicating high, medium, and low potential for success in the program, respectively. These recommendations were shared with the fellowship faculty. Candidates who fell into the green category were invited for an on-site interview, in which the final two assessments—structured interviews and technical skills testing—took place.

Before conducting the interviews, all fellowship faculty completed a half-day structured interview course described elsewhere[14] to discuss common biases during interviews and ensure reliable assessment of position-related competencies during the on-site interviews.

On interview day, candidates rotated through three 30-minute interview rooms and a 30-minute skills testing room. Skills testing was done in a laparoscopic task trainer box modified to allow the user to work with the camera located off-axis to the target task. Applicants were asked to perform 2 tasks adapted from the Fundamentals of Laparoscopic Surgery (FLS Task 1—Peg Transfer and Task 5—Intracorporeal Suturing).[19] Each participant was oriented to the tasks and then given a maximum of 10 minutes to complete each. Data from each element of the application process were summarized and a rank list was created.

After a rank list was submitted to the Fellowship Council, but before results of the match were available, faculty were asked to complete a follow-up survey regarding their opinions of the utility of the new selection process and experiences with the on-site interviews. They were also given the opportunity to provide open-ended comments about the overall process.

After all processes were completed, candidate demographic data, including race and sex, were obtained from the fellowship application Web site. Applicants were categorized as underrepresented minority (URM) candidates if they were non-Caucasian and/or women. We also obtained available historical data on applicants and number of interviews conducted from the program, which spanned the prior 3 years.

Basic descriptive statistics (means, standard deviation, frequencies) were used to examine scores from the SJT, personality profile, on-site interview, skills examination, and follow-up survey. Analysis of variance was used to explore differences in data from the application system (case logs, ABSITE, USMLE, etc) across recommendation categories (green, yellow, red). SPSS version 24.0 (IBM, Chicago, IL) was used to conduct these analyses.

RESULTS

Demographics

Seventy-two applicants applied to the fellowship program. Fifty-five (70% URM) were deemed able to fulfill the logistical requirements of the position including eligibility to obtain a state medical license and to be credentialed to practice in the host hospital. All fifty-five applicants were invited to take the online SJT and personality profile assessment with 50 (91%) completing it, resulting in a total representation of 76% URMs after the online screening phase.

Based on scores from the SJT and personality profile, 18 applicants (35%) received a green, 17 received a yellow (33%), and 15 (30%) received a red recommendation for further consideration. All 18 ''green'' candidates (83% URM) were invited for an on-site interview with 12 attending, resulting in a final pool of 92% URMs for on-site interviews. If the traditional screening method were used (ie, making decisions on application data only), only 71% of individuals invited for an onsite interview would have been from an URM group.

After interviews, a final rank list of the 12 applicants (92% URM) was created using all data points (ie, SJT, personality profile, interview performance, and technical skills) and their respective importance. The Fellowship Council match process resulted in filling both program positions with candidates from the top 4 positions, both of which represented individuals from URM groups. This process is illustrated in Figure 1.

Efficiency

Comparisons between the traditional screening process and the new methodology were conducted. If no on-line prescreening assessments were used and decisions were based solely on application information, the program would have invited 38 (69%) applicants for the on-site visit—over twice as many as invited with the new screening approach. In addition, historical data from the past 3 years indicate that an average of 44% of eligible applicants was interviewed, a number twice that of the number of eligible applicants interviewed using the new system (12/55; 22%).

There was no correlation between the traditional applicant review process and the on-line prescreening process when comparing interview invite decisions. An examination of the specific objective elements contained within the Fellowship Council standardized application also revealed that there was no correlation between test performance (ABSITE or USMLE) and laparoscopic clinical experience and the green, yellow, and red recommendations for candidates (Table 1).

Table 2 shows the faculty time commitment required for the traditional selection system and the new one. The total faculty time commitment required to create and implement the selection process for this cohort of 55 applicants approximated 53 hours. This effort was divided into 4 steps: (1) participate in the job analysis—2 hours/faculty; (2) provide input and consensus for the SJT—1 hour/faculty; (3) train to conduct structured interviews—4 hours/faculty; and (4)

© 2018 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

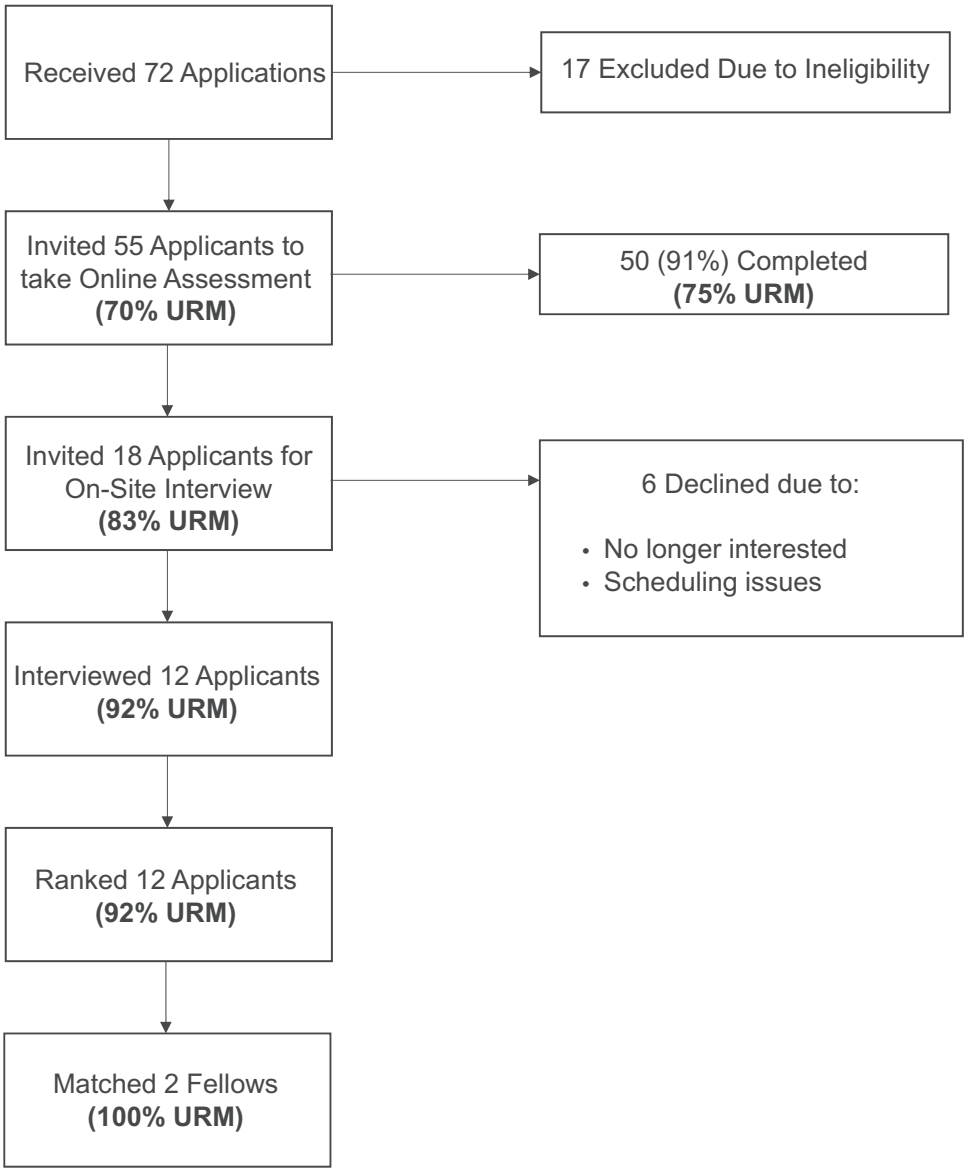

FIGURE 1. Overview of selection process.

conduct on-site interviews—18 hours (12 applicants × 3 interviews × 0.5 hours). As the structured interviews provided greater consistency among interviewers in both conducting the sessions and making ratings,[14] only 3 of the 5 faculty were required to participate in each interview day. This compares to approximately 56 hours of faculty time using the traditional method: (1) review applicant files—14 hours (55 × 15 min per file); (2) interview—42 hours (25 × 0.33 hours × 5 interviews) assuming the same percentage of interview invites to on-site visits as seen in the new selection system. The traditional method requires the same application review and interview activities year-to-year. However, because steps 1 to 3 of the new selection system represent up-front time commitments for

TABLE 1. Metrics From Fellowship Council Application by Recommendation Group

	ABSITE			USMLE*			Case Logs			
	PGY1	PGY2	PGY3	1	2	3	Lap Cases	Open Cases	Endo Cases	% of Lap That Are Chole/Appy
Green (N = 18)	57	49	49	228	238	218	259	176	94	61
Yellow (N = 17)	57	45	46	224	239	213	246	178	86	59
Red (N = 15)	50	42	48	218	234	210	201	146	100	65

*USMLE scores unavailable for 11 applicants (22%). No significant differences among groups on any metric.

© 2018 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

TABLE 2. Faculty Time Commitment Comparing Traditional and New Selection Methods

Traditional Selection System		New Selection System	
Step	**Total Faculty Hours**	**Step**	**Total Faculty Hours**
1. Review applications	14	1. Participate in job analysis	10
2. Conduct interviews	42	2. Provide input and consensus on assessment tools and scoring keys	5
		3. Participate in structured interview training	20
		4. Conduct interviews	18
Year 1 total	56	Year 1 total	53
Year 2 total	56	Year 2 total	18

development, the new process will result in a 68% savings in faculty time after the first year (18 vs 56 hours). These estimates account for neither time spent by staff screening applicant files and coordinating the interview days, nor for faculty time creating rank lists using less scientific methods. In addition, average time to complete the assessments by candidates was 42 ± 12 minutes, with a range of 23 to 71 minutes.

Faculty Perceptions

Follow-up surveys were completed by all 5 fellowship faculty members. When asked about their opinions regarding the process of participating in the job analysis, creation and validation of the SJTs, administration of pre-screening assessments, and creation of applicant reports, all faculty ($N = 5$) indicated that they ''very much'' agreed that they saw value in going through the process. In addition, all faculty agreed that the process helped them understand what attributes are important for a fellow to be successful in the program, resulted in greater confidence in identifying which candidates would be a good fit compared to prior years, and that the process helped them better understand the requirements of their fellowship. Faculty also agreed that the process was not overly burdensome for them and they believed it gave applicants better insight into what is expected of a fellow in the program. When asked about whether or not they implemented tips and techniques from the structured interview training program, all faculty indicated ''yes.'' One hundred percent of faculty also indicated that they were able to obtain more meaningful information from the candidate during the structured interview process, compared to prior years. Finally, when asked for open-ended comments about the overall selection process in general, faculty indicated that the process was ''enjoyable,'' ''helpful for parsing out candidates that all look good on paper,'' provided ''more and better information about our applicants than I ever had in the past,'' and that they ''felt empowered to separate one candidate from another based on real data.''

DISCUSSION

This study demonstrates that incorporating an evidence-based selection process into a fellowship program is both feasible and efficient. Faculty spent no more than 7 hours individually helping to construct and provide input toward the assessment tools and learning to conduct structured interviews, with these hours representing an ''up-front'' investment that won't be required again for 3 to 5 years. In addition, candidates only spent an average of 42 minutes completing the on-line assessment—a small time investment to gain deep insight into what is required for the job for which they are interviewing. The process also resulted in less time reviewing candidate files by both the program directors and coordinator, inviting a relatively small cohort of applicants to interview, conducting fewer on-site interviews than in the past, and requiring less faculty for the interviews (3 of 5 faculty were required to interview vs 5 historically) while gathering more relevant and reliable information. Finally, and of most importance to those concerned that a ''non-standard'' selection process may prevent top candidates from engaging; we found that 91% of applicants invited to participate did so. As research has shown that applicants who highly desire a position are unlikely to drop out of even the most rigorous screening processes,[20] the 5 applicants who chose not to complete the online assessment battery were likely not driven to join the program.

Importantly, this screening process can also ease the burden of time and travel dedicated to fellowship interviews. Scholars have noted that fellowship applicants spend anywhere from $4000 to $10,000 and up to 2 weeks of time to attend on-site interviews.[2–5] Our approach allowed applicants to participate in a minimally burdensome process (ie, <1 hour) wherein they could learn more about the program and requirements of the position early on, whereas the program simultaneously obtains information from applicants about key competencies required to be a fellow in the program. If invited for an on-site interview, applicants could then decide for themselves if they want to invest the time and resources to travel for the interview. As a result, the application process was transformed into a two-way experience, in which both applicants and program leaders are determining their fit with one another.

This study may also provide preliminary evidence that implementing an evidence-based selection process into a fellowship program can level the playing field for URM candidates. The percentage of URM candidates increased throughout each stage of screening process, moving from an initial eligible pool of 70% URMs, to 83% after the online screening process, to 92% interviewed, and ultimately 100% matched. Importantly, we also determined that only 71% of those invited for interviews would have been URMs, if decisions were based solely on data from the application. These data represent the power of this selection methodology. As noted in the follow-up surveys, program leaders felt empowered to make decisions about inviting applicants to interview based on reliable and specific data rather than test scores, clinical experience, and unstructured letters of recommendation that have no correlation to job fit and can serve as barriers to underrepresented minorities.[21–23]

In addition, interviews were conducted with specific goals and measurement metrics that provided insight into the candidates' competencies and steered interviewers away from common biases and pitfalls.[14,24] Although the term ''unconscious bias training'' was never formally used for the interviewer training program, we did discuss the array of biases we hold and how these can inadvertently affect our actions and judgments. Faculty practiced identifying various biases, discussed strategies to overcome them, and worked on retrieving the same objective information from all candidates. In addition, the use of, and training on, our structured interview evaluation forms may have served as an important ''inclusion nudge'' (those processes and tools that ''nudge'' people toward making unbiased decisions at work)[25] for our faculty. Inclusion nudges differ from diversity training because they influence behavior as it occurs,

© 2018 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

and remove opportunities for bias and unfair treatment. Thus, the combination of our training program, and the use of our structured processes and evaluation tools, may have combated unconscious biases, thereby allowed for wider inclusion, and ultimately selection, of diverse candidates.

As with any study, there are some limitations to our findings. First, these data are from a single fellowship in a single institution, making the generalizability of these findings to other surgery programs and specialties unknown. It is, however, likely that training programs selecting more trainees and with more faculty involved in the process would demonstrate even more drastic improvements in efficiency as a result of new screening processes. In addition, although we were able to compare URM representation between traditional screening practices and our new process using one year of data, we have no historical data on applicant demographics to demonstrate statistically significant improvement in URMs compared to years past as a result of this process. Given the high prevalence (70%) of URMs within the initial group of eligible applicants at this institution, these findings may be even more pronounced for programs with a less diverse applicant pool. In addition, it is worth noting that enhancing diversity was not a primary aim of adopting the new selection process, and thus limiting the likelihood that faculty subconsciously rated or ranked interviewed applicants higher to achieve this goal. Finally, the true success of any selection system is of course the performance and successful completion of those hired. Future work will continue to follow the selected applicants and measure their performance and ''fit'' within the program.

CONCLUSIONS

Providing candidates with an equal opportunity to demonstrate critical knowledge, skills, and abilities required for a specific training position, whereas de-emphasizing traditional practices that allow for common biases and discriminatory practices to emerge, can empower surgery programs to simultaneously make selection decisions based on objective data and potentially enhance the diversity of the surgeon workforce. Our findings suggest that integrating selection science into the methodology used by surgical training programs may be a ''triple win'' from an effectiveness, efficiency, and diversity standpoint.

REFERENCES

1. Bell RH Jr, Banker MB, Rhodes RS, et al. Graduate medical education in surgery in the United States. *Surg Clin North Am*. 2007;87:811–823.
2. Wood JS, David LR. Outcome analysis of factors impacting the plastic surgery match. *Ann Plast Surg*. 2010;64:770–774.
3. Beres A, Baird R, Pugligandla PS. Success in the pediatric surgery match: a national survey of the 2010 applicant pool. *J Pediatr Surg*. 2011;46:957–961.
4. Oladeji LO, Pehler SF, Raley JA, et al. Is the orthopedic fellowship interview process broken? A survey of program directors and residents. *Am J Orthop (Belle Mead NJ)*. 2015;44:e444–e453.
5. Watson SL, Hollis RH, Oladeji L, et al. The burden of the fellowship interview process on general surgery residents and programs. *J Surg Educ*. 2017;74:167–172.
6. Gardner AK, Grantcharov T, Dunkin BJ. The science of selection: using best practices from industry to improve success in surgery training. *J Surg Educ*. 2018;75:278–285.
7. Brummond A, Sefcik S, Halvorsen AJ, et al. Resident recruitment costs: a national survey of internal medicine program directors. *Am J Med*. 2013;126:646–653.
8. Gardner AK, Smink DS, Scott BG, et al. Money talks: How much are we spending on resident selection? Podium presentation at the Association of Program Directors in Surgery (APDS) Annual Meeting. Austin, Texas. May 4, 2018.
9. Gardner AK. How can best practices in recruitment and selection improve diversity in surgery? *Ann Surg*. 2018;267:e1–e2.
10. Higgins RSD. Inside baseball—leveling the playing field in the surgical residency selection process. *Ann Surg*. 2018;267:e3.
11. Salgado J, Grantcharov TP, Papasavas PK, et al. Technical skill assessment as part of the selection process for a fellowship in minimally invasive surgery. *Surg Endosc*. 2009;23:641–644.
12. Gardner AK, Dunkin BJ. Evaluation of validity evidence for personality, emotional intelligence, and situational judgment tests to identify successful residents. *JAMA Surg*. 2017. in press. DOI: 10.1001; jamasurg.2017.5013.
13. Stephenson-Famy A, Houmard BS, Oberoi S, et al. Use of the interview in resident candidate selection: a review of the literature. *J Grad Med Educ*. 2015;7:539–548.
14. Gardner AK, D'Onofrio BC, Dunkin BJ. Can we get faculty interviewers on the same page? An examination of a structured interview course for surgeons. *J Surg Educ*. 2018;75:72–77.
15. Gardner AK, Ritter EM, Paige JT, et al. Simulation-based selection of medical trainees: considerations, challenges, and opportunities. *J Am Coll Surg*. 2016;223:530–536.
16. Gardner AK, Steffes CP, Nepomnayshy D, et al. Selection bias: examining the feasibility, utility, and participant receptivity to incorporating simulation into the general surgery residency selection process. *Am J Surg*. 2017;213:1171–1177.
17. Lievens F. Adjusting medical school admission: assessing interpersonal skills using situational judgment tests. *Med Educ*. 2013;47:182–189.
18. Jackson DN, Paunonen SV, Tremblay PF. *Six Factor Personality Questionnaire.* Port Huron, MI: Sigma Assessments Systems; 2000.
19. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. *Ann Surg*. 2004;240:518–528.
20. Truxillo DM, Bauer TN, Campion MA, et al. Selection fairness information and applicant reactions: a longitudinal field study. *J Appl Psychol*. 2002;87:1020–1031.
21. Outtz JL. The role of cognitive ability tests in employment selection. *Human Performance*. 2002;15:161–171.
22. Koenig JA, Sireci SG, Wiley A. Evaluating the predictive validity of MCAT scores across diverse applicant groups. *Acad Med*. 1998;73:1095–1106.
23. Huff KL, Koenig JA, Treptau MM, et al. Validity of MCAT scores for predicting clerkship performance of medical students grouped by sex and ethnicity. *Acad Med*. 1999;7:S41–S44.
24. Hern HG Jr, Alter HJ, Wills CP, et al. How prevalent are potentially illegal questions during residency interviews? *Acad Med*. 2013;88:1116–1121.
25. Nielsen TC, Kepinski L. *Inclusion Nudges Guidebook: Practical Techniques for Changing Behavior, Culture, and Systems to Mitigate Unconscious Bias and Create Inclusive Organizations.* 2nd ed, Charleston, SC: CreateSpace Independent Publishing; 2016.

© 2018 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.