

Evaluation of Validity Evidence for Personality, Emotional Intelligence, and Situational Judgment Tests to Identify Successful Residents

Aimee K. Gardner, PhD; Brian J. Dunkin, MD

IMPORTANCE The ability to identify candidates who will thrive and successfully complete their residency is especially critical for general surgery programs.

OBJECTIVE To assess the extent to which 3 screening tools used extensively in industrial selection settings—emotional intelligence (EQ), personality profiles, and situational judgment tests (SJTs)—could identify successful surgery residents.

DESIGN, SETTING, AND PARTICIPANTS In this analysis, personality profiles, EQ assessments, and SJTs were administered from July through August 2015 to 51 postgraduate year 1 through 5 general surgery residents in a large general surgery residency program. Associations between these variables and residency performance were investigated through correlation and hierarchical regression analyses.

INTERVENTIONS Completion of EQ, personality profiles, and SJT assessments.

MAIN OUTCOMES AND MEASURES Performance in residency as measured by a comprehensive performance metric. A score of zero represented a resident whose performance was consistent with that of their respective cohort's performance; below zero, worse performance; and greater than zero, better performance.

RESULTS Of the 61 eligible residents, 51 (84%) chose to participate and 22 (43%) were women. US Medical Licensing Examination Step 1 (USMLE1), but not USMLE2, emerged as a significant factor ($t_{2,49} = 1.98$; $\beta = 0.30$; $P = .03$) associated with overall performance. Neither EQ facets nor overall EQ offered significant incremental validity over USMLE1 scores. Inclusion of the personality factors did not significantly alter the test statistic and did not explain any additional portion of the variance. By contrast, inclusion of SJT scores accounted for 15% more of the variance than USMLE1 scores alone, resulting in a total of 25% of the variance explained by both USMLE1 and SJT scores ($F_{2,57} = 7.47$; $P = .002$). Both USMLE1 ($t = 2.21$; $P = .03$) and SJT scores ($t = 2.97$; $P = .005$) were significantly associated with overall resident performance.

CONCLUSIONS AND RELEVANCE This study found little support for the use of EQ assessment and only weak support for some distinct personality factors (ie, agreeableness, extraversion, and independence) in surgery resident selection. Performance on the SJT was associated with overall resident performance more than traditional cognitive measures (ie, USMLE scores). These data support further exploration of these 2 screening assessments on a larger scale across specialties and institutions.

JAMA Surg. doi:10.1001/jamasurg.2017.5013
Published online December 27, 2017.

 [Invited Commentary](#)

 [Supplemental content](#)

Author Affiliations: Department of Surgery, School of Allied Health Sciences, Baylor College of Medicine, Houston, Texas (Gardner); Department of Surgery, Houston Methodist Hospital, Houston, Texas (Dunkin).

Corresponding Author: Aimee K. Gardner, PhD, Department of Surgery, School of Allied Health Sciences, Baylor College of Medicine, MS BCM115, DeBakey Bldg, M108K, One Baylor Plaza, Houston, TX 77030 (aimee.gardner@bcm.edu).

Medical educators are increasingly investigating improved methods for screening and selecting applicants for medical training programs.¹⁻³ Screening assessments to determine applicant fit with a residency often include US Medical Licensing Exam (USMLE) scores, medical student performance evaluations, letters of recommendation, personal statements, and in-person interviews.^{4,5} However, scholars have observed wide variability not only in the way each of these data points are used but also in their ability to estimate later performance in residency.^{6,7}

The ability to select candidates who will thrive and successfully complete a residency is especially critical for general surgery programs. General surgery residency typically spans 5 to 7 years of intense training, most often followed by an additional 1 to 2 years of specialty training.⁸ These factors require program directors to identify candidates who not only demonstrate the competencies and aptitude required to be a surgeon but also can manage the extended length of training in a high-stress environment. However, literature reviews have shown that up to 30% of residents in surgery programs require at least 1 remediation intervention for performance issues,⁹ most of which involve nontechnical competencies, such as interpersonal skills and professionalism.¹⁰⁻¹² In addition, approximately a quarter of those who enter surgery training programs do not stay, resulting in one of the highest attrition rates across medical specialties.¹³

There are undoubtedly a number of factors leading to these high attrition rates and thus multiple potential solutions (eg, providing more realistic previews of surgical careers to students, enhancing the quality of training programs, and incorporating methods to identify residents at risk for remediation or attrition). However, given the resources involved with current selection practices, remediation programs, and costs of attrition,³ it is critical that program directors are able to effectively and efficiently identify candidates who will be successful in their particular training programs.

We investigated whether candidate assessment practices commonly used in industry could be applied to the resident screening process to maximize applicant-organization fit. Specifically, we used correlation and hierarchical linear regression analyses to assess the extent to which emotional intelligence (EQ), personality profiles, and situational judgment tests (SJTs)—3 screening tools that have received extensive attention for their use in candidate selection in industrial settings—were associated with resident performance 1 year after administration in a large general surgery residency program.

Methods

The 3 screening tools—EQ, personality profile, and situational judgment tests—were administered from July through August 2015 to general surgery residents who were in a large residency training program, and the test results were correlated 1 year later with a multidimensional performance metric. The screening tools and resident performance metric were created and administered as described below. The institutional review board at the University of Texas Southwestern

Key Points

Question Are personality profiles, emotional intelligence, and situational judgment tests useful applicant screening tools for identifying successful residents?

Findings This analysis of 3 screening tool results among 51 postgraduate year 1 through 5 general surgery residents found that, although emotional intelligence and personality factors were significantly correlated with various performance dimensions, only US Medical Licensing Examination Step 1 (accounting for 12% of performance variance) and situational judgment test scores were associated with overall performance 1 year later. Both tools together accounted for 25% of overall resident performance variance.

Meaning Inclusion of situational judgment test assessments in the resident selection process may be warranted.

Medical Center, Dallas, waived the need for review and documentation of participant consent.

Emotional Intelligence

Emotional intelligence was assessed with the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT), version 2.0 (Multi-Health Systems, Inc). This widely investigated tool consists of 141 items and measures each of the 4 EQ branches (eAppendix 1 in the [Supplement](#)). Moderate but significant correlations have been found between MSCEIT scores and measures of cognitive ability and the 5 basic dimensions of personality, termed the “Big 5” personality traits (ie, openness, conscientiousness, extraversion, agreeableness, and neuroticism), suggesting that EQ is associated with but distinguishable from intelligence and personality.¹⁴⁻¹⁶ In addition, positive correlations have been reported with academic achievement¹⁷ and psychological well-being.¹⁵ Scores are calculated similar to intelligence quotient assessments in that the mean (SD) score is 100 (15).

Personality

The Six Factor Personality Questionnaire (SFPQ)¹⁸ was used to assess personality. The SFPQ contains 108 items that assess the traditional Big 5 personality traits but bifurcates conscientiousness into methodicalness and industriousness facets. Each factor has 3 narrow facet scales that are assessed by 6 items each (eAppendix 2 in the [Supplement](#)). The factor scale scores range from 18 to 90. Participants responded to SFPQ items using a 5-point Likert scale, with 1 indicating strongly disagree and 5 indicating strongly agree.

Situational Judgment Test

During an SJT, participants are presented with hypothetical but realistic job-relevant scenarios, and they must determine the most and least effective options from a list of potential responses. These items do not measure medical knowledge but measure judgment as well as decision-making and problem-solving skills across a wide array of situations. An example item is provided in eAppendix 3 in the [Supplement](#).

The 50-item SJT and scoring key were created in accordance with other studies.¹⁹ The Kendall coefficient of concordance computed for each ranking item showed 0.68 concordance, indicat-

Table 1. Performance Measures and Their Contributions to Overall Performance

| Description | Administration Rate | Rater | Weight Given by CCC, % |
|--|---------------------|---|------------------------|
| Faculty evaluation | | | |
| 21-Item tool measuring display of ACGME competencies | Monthly | Clinical supervising faculty | 37 |
| Professionalism | | | |
| Completion of administrative tasks (eg, case logs completed on time), conference attendance, duty hour compliance, etc | Monthly | Administrative staff | 18 |
| Case logs | | | |
| Procedural case activity from ACGME case log data | NA | NA | 17 |
| ABSITE | | | |
| In-training examination | Yearly | NA | 14 |
| Scholarly activity | | | |
| Presentations (local/national) and peer-reviewed presentations | NA | NA | 8 |
| Medical student evaluations | | | |
| 15-Item tool measuring teaching, interpersonal skills, and professionalism | Monthly | Medical students rotating with resident | 6 |

Abbreviations: ABSITE, American Board of Surgery In-Training Examination; ACGME, Accreditation Council for Graduate Medical Education; CCC, Clinical Competency Committee; NA, not applicable.

ing adequate interrater agreement. Thus, 34 of 50 items (68%) were of sufficient psychometric quality to be included in the final assessment. The maximum score on the SJT assessment was 77 points. A knowledge-based response instruction format (ie, “What should you do?” vs “What would you do?”) was used because this format is less prone to insincere responses.^{20,21}

Performance

Resident performance measures consisted of data from monthly faculty evaluations, faculty- and staff-rated professionalism metrics (ie, interpersonal and communication skills, completion of administrative tasks, conference attendance, duty hour compliance, and so forth that align with Accreditation Council for Graduate Medical Education [ACGME] milestones) that are completed monthly, procedural activity from ACGME case log data, and scholarly activity (ie, raw number of presentations and peer-reviewed publications). An overview of these performance measures is provided in Table 1. Because all programs have unique cultures and because the expectations of residents and the existing evaluation metrics (ie, milestones) at the time of this study were based primarily on monthly faculty evaluations, an institution-specific overall performance metric using the aforementioned measures was created. Nine faculty members who had been tasked with monitoring and evaluating resident performance through the department Clinical Competency Committee were asked to provide weights for each of these factors. The following is the resulting overall performance equation for the program:

$$\text{Resident Performance} = (0.37 \times \text{Faculty Evaluation Score}) + (0.18 \times \text{Professionalism Score}) + (0.17 \times \text{Case Log Score}) + (0.14 \times \text{American Board of Surgery In-Training Examination score}) + (0.08 \times \text{Scholarly Score}) + (0.06 \times \text{Medical Student Evaluation Score}).$$

Because senior residents are likely to have higher values on some variables (eg, scholarly publications, case logs, and professionalism) owing to their having been in the program lon-

ger, z scores were created based on the postgraduate year (PGY) mean, enabling aggregation of performance data from the 5 PGY cohorts. Overall performance was calculated according to the equation above and multiplied by 100. Thus, a score of zero represented a resident whose performance was consistent with that of their respective cohort; below zero, worse performance; and greater than zero, better performance. Given the variable weighting of each performance measure, as described above, a resident who far exceeded the cohort mean on faculty evaluations would have a substantially larger overall performance score than a resident who far exceeded the cohort mean on medical student evaluations.

Assessment Administration

Surgery residents at a single institution were invited to participate in this study at the beginning of the academic year. Trainees had 2 hours to complete the assessments. The USMLE scores and performance measures obtained 1 year later were collected for those residents who chose to participate.

Statistical Analysis

Descriptive statistics were computed for the MSCEIT, SFPQ, SJT, and performance measures. Pearson correlations were used to examine associations between variables. Independent paired-samples 2-tailed *t* tests were used to examine performance differences between residents who chose to complete the assessments and those who did not. Linear regression analysis was used to identify factors independently associated with the performance measures. All statistical tests were 2-sided, and *P* < .05 was considered statistically significant. All data were analyzed using SPSS software, version 24.0 (IBM).

Results

Descriptive Statistics

Of the 61 eligible residents, 51 (84%) chose to participate in this study (PGY1, 13 of 13; PGY2, 10 of 13; PGY3, 12 of 13; PGY4, 9 of

Table 2. Means and SDs of Evaluation Variables by Resident PGY

| PGY | Mean (SD) | | | | | | | | |
|----------------|------------------------|-----------------------|---------------|--------------|-------------|----------------|-----------------|----------------|----------------|
| | Emotional Intelligence | Personality Dimension | | | | | USMLE | | |
| | | Extraversion | Agreeableness | Independence | Openness | Methodicalness | Industriousness | USMLE1 | USMLE2 |
| PGY1 (n = 13) | 97.46 (13.70) | 3.81 (0.31) | 3.15 (0.25) | 2.58 (0.38) | 3.44 (0.49) | 3.60 (0.64) | 3.71 (0.37) | 244.42 (9.07) | 257.27 (8.36) |
| PGY2 (n = 10) | 104.20 (13.51) | 3.71 (0.52) | 3.03 (0.48) | 2.80 (0.54) | 3.15 (0.53) | 3.27 (0.55) | 3.54 (0.22) | 241.21 (7.00) | 250.46 (12.74) |
| PGY3 (n = 12) | 94.08 (22.45) | 3.42 (0.39) | 2.94 (0.29) | 2.69 (0.38) | 3.38 (0.34) | 3.51 (0.42) | 3.51 (0.42) | 247.31 (13.19) | 260.25 (12.12) |
| PGY4 (n = 9) | 96.11 (13.91) | 3.42 (0.47) | 2.60 (0.25) | 2.78 (0.37) | 3.20 (0.50) | 3.77 (0.54) | 3.81 (0.49) | 235.40 (11.96) | 254.00 (14.70) |
| PGY5 (n = 7) | 92.57 (12.34) | 3.65 (0.65) | 3.04 (0.49) | 2.88 (0.62) | 3.84 (0.40) | 3.55 (0.29) | 3.40 (0.31) | 243.89 (12.27) | 244.67 (8.98) |
| Total (N = 51) | 97.08 (15.92) | 3.60 (0.46) | 2.96 (0.38) | 2.72 (0.44) | 3.38 (0.48) | 3.54 (0.53) | 3.61 (0.39) | 242.89 (11.25) | 254.10 (12.55) |

Abbreviations: PGY, postgraduate year; USMLE1, US Medical Licensing Examination Step 1; USMLE2, USMLE Step 2.

Table 3. Means and SDs of Standardized Criterion Variables by Resident PGY

| PGY | Faculty Evaluation, % | ABSITE, % | Medical Student Evaluation, % | Case Logs ^a | Professionalism, % | Scholarly Works, No. | Overall Performance ^b |
|----------------|-----------------------|---------------|-------------------------------|------------------------|--------------------|----------------------|----------------------------------|
| PGY1 (n = 13) | | | | | | | |
| Mean (SD) | 86.63 (4.08) | 71.36 (26.48) | 88.37 (5.1) | 106.46 (23.67) | 53.41 (6.67) | 0 | 2.72 (27.84) |
| Range | 80.70 to 93.90 | 24 to 97 | 84.0 to 96.9 | 77 to 156 | 38 to 63 | 0 | -34.77 to 51.54 |
| PGY2 (n = 10) | | | | | | | |
| Mean (SD) | 79.69 (8.66) | 64.36 (27.11) | 87.72 (6.51) | 275.28 (24.55) | 49.4 (5.63) | 0.07 (0.27) | -8.78 (54.69) |
| Range | 51.1 to 85.67 | 23 to 97 | 72.7 to 94.5 | 240 to 321 | 42 to 63 | 0 to 1 | -166.45 to 61.77 |
| PGY3 (n = 12) | | | | | | | |
| Mean (SD) | 85.13 (2.79) | 72 (22.72) | 88.46 (4.36) | 534.00 (60.22) | 72.32 (8.74) | 0.36 (0.72) | 1.34 (57.40) |
| Range | 78.57 to 90.02 | 25 to 97 | 81.8 to 96.2 | 451 to 632 | 58 to 88 | 0 to 2 | -125.46 to 124.67 |
| PGY4 (n = 9) | | | | | | | |
| Mean (SD) | 91.22 (2.66) | 65.23 (26.1) | 85.32 (5.91) | 777.25 (42.33) | 61.76 (10.96) | 1.53 (2.93) | -0.47 (49.67) |
| Range | 85.21 to 94.08 | 14 to 96 | 74.7 to 94.4 | 753 to 861 | 63 to 96 | 0 to 9 | -108.202 to 67.73 |
| PGY5 (n = 7) | | | | | | | |
| Mean (SD) | 93.26 (2.22) | 64.44 (25.11) | 88.24 (4.66) | 1182.32 (75.2) | 92.59 (5.01) | 2.67 (1.94) | 0.03 (58.82) |
| Range | 89.54 to 97.13 | 21 to 99 | 81.1 to 94.3 | 952 to 1200 | 83 to 100 | 0 to 6 | -98.65 to 87.40 |
| Total (N = 51) | | | | | | | |
| Mean (SD) | 86.18 (7.28) | 67.87 (24.88) | 87.77 (5.33) | 575.06 (424.50) | 68.44 (17.55) | 0.95 (2.39) | -1.25 (49.91) |

Abbreviations: ABSITE, American Board of Surgery In-Training Examination; PGY, postgraduate year.

^b Scaled score indicating overall performance.

^a Absolute value compared with cohort.

13; and PGY5, 7 of 9) and 22 (43%) were women. The results of paired-samples *t* tests revealed that the performance measures of those who chose to participate did not significantly differ from those who chose not to participate. **Table 2** presents the means and SDs of the evaluation variables (EQ, personality profile, and SJT results) by PGY. **Table 3** provides the standardized means and SDs of performance criterion variables by PGY.

Correlations

Table 4 provides the Pearson correlation coefficients between all evaluation and criterion variables. Overall EQ was associated with the personality facet of industriousness ($r = 0.31; P = .03$). Within the personality factors, extraversion was significantly associated with both independence ($r = -0.34; P = .01$) and openness ($r = 0.42; P = .002$). Methodical-

ness and industriousness were significantly associated with each other ($r = 0.31; P = .02$). Both USMLE1 ($r = 0.48; P < .001$) and USMLE2 ($r = 0.37; P = .004$) were significantly associated with American Board of Surgery In-Training Examination (ABSITE) scores. USMLE1 was also significantly associated with overall performance ($r = 0.30; P = .02$), USMLE2 ($r = 0.56; P < .001$), and agreeableness ($r = 0.29; P = .04$). Performance on the SJT was significantly associated with faculty evaluations ($r = 0.31; P = .03$), medical student evaluations ($r = 0.38; P = .03$), overall performance ($r = 0.41; P = .006$), and overall EQ ($r = 0.35; P = .02$).

Associative Validity

Hierarchical regression analyses were conducted to further examine the associations among these variables. Specifically, we

Table 4. Pearson Correlation Coefficients Between All Evaluation and Criterion Variables

| Variable | ABSITE | MS Eval | Case Logs | Profes- sionalism | Scholarly Works | Overall Perfor- mance | Overall EQ | USMLE1 | USMLE2 | Extra- version | Agree- ableness | Indepen- dence | Openness | Methodi- calness | Industri- ousness | SJT |
|---------------------|--------|-------------------|-----------|----------------------|--------------------|-----------------------------|------------|-------------------|-------------------|-------------------|--------------------|--------------------|-------------------|---------------------|----------------------|-------------------|
| Faculty evaluation | 0.02 | 0.27 ^a | 0.23 | -0.03 | -0.13 | 0.81 ^b | 0.02 | 0.17 | -0.01 | 0.21 | 0.15 | 0.18 | 0.14 | -0.13 | -0.18 | 0.31 ^a |
| ABSITE | | 0.20 | 0.12 | 0.00 | 0.22 | 0.37 ^b | -0.13 | 0.48 ^b | 0.37 ^b | 0.19 | -0.01 | -0.22 | -0.02 | -0.10 | 0.06 | 0.05 |
| MS eval | | | 0.00 | 0.16 | 0.02 | 0.42 ^b | 0.17 | 0.25 | 0.10 | 0.31 ^a | 0.43 ^b | -0.06 | 0.08 | 0.16 | 0.09 | 0.38 ^b |
| Case logs | | | | -0.15 | -0.12 | 0.47 ^b | 0.13 | 0.05 | 0.24 | -0.06 | -0.02 | 0.31 ^a | 0.04 | -0.17 | 0.19 | 0.22 |
| Professionalism | | | | | -0.03 | 0.29 ^a | -0.16 | 0.00 | 0.06 | -0.21 | 0.03 | 0.16 | -0.08 | 0.12 | 0.01 | 0.14 |
| Scholarly | | | | | | 0.05 | -0.18 | 0.15 | 0.10 | 0.09 | -0.10 | 0.03 | 0.02 | 0.24 | 0.09 | 0.06 |
| Overall performance | | | | | | | -0.03 | 0.30 ^a | 0.23 | 0.16 | 0.15 | 0.24 | 0.10 | -0.09 | -0.02 | 0.41 ^b |
| Overall EQ | | | | | | | | -0.10 | -0.06 | 0.19 | 0.21 | -0.01 | 0.01 | 0.14 | 0.31 ^a | 0.35 ^a |
| USMLE1 | | | | | | | | | 0.56 ^b | 0.14 | 0.29 ^a | -0.07 | -0.04 | -0.02 | -0.11 | 0.05 |
| USMLE2 | | | | | | | | | | 0.04 | 0.09 | 0.01 | -0.16 | -0.05 | 0.08 | 0.04 |
| Extraversion | | | | | | | | | | | 0.11 | -0.34 ^a | 0.42 ^b | -0.03 | 0.10 | 0.23 |
| Agreeableness | | | | | | | | | | | | -0.05 | 0.12 | 0.16 | -0.04 | 0.08 |
| Independence | | | | | | | | | | | | | -0.02 | -0.17 | 0.04 | -0.02 |
| Openness | | | | | | | | | | | | | | -0.07 | 0.03 | 0.07 |
| Methodicalness | | | | | | | | | | | | | | | 0.31 ^a | -0.10 |
| Industriousness | | | | | | | | | | | | | | | | 0.04 |

Abbreviations: ABSITE, American Board of Surgery In-Training Examination; EQ, emotional intelligence; MS Eval, medical student evaluation; MS Eval, situational judgment test; USMLE1, US Medical Licensing Examination Step 1; USMLE2, USMLE Step 2.

^a P < .05.

^b P < .01.

wanted to know the extent to which scores on the evaluation variables (USMLE, EQ, personality factors, and SJT performance) were associated with the overall performance of residents during their residency. We began by including both USMLE scores (Steps 1 and 2) in a regression equation. We found that these scores accounted for 12% of the criterion variance ($F_{2,57} = 3.68$; $P = .03$). However, only results from USMLE1 emerged as a significantly associated factor ($t_{2,49} = 1.98$; $\beta = 0.30$; $P = .03$). We then entered USMLE1 in a first block of the regression equation and EQ facets in a second block. Neither EQ facets nor overall EQ offered significant incremental variance over the use of USMLE1 scores alone. We performed another set of regression analyses with USMLE1 entered in the first block and personality factors in the second block. Inclusion of personality factors did not significantly alter the test statistic and did not account for any additional portion of the variance. Finally, we conducted analyses with USMLE1 in the first block and SJT scores in the second block. This model accounted for 15% more of the variance than the USMLE1 scores alone, resulting in a total of 25% of the variance explained by USMLE1 and SJT scores together ($F_{2,57} = 7.47$; $P = .002$) and indicating that SJT scores had significant incremental validity over using USMLE1 scores alone. Both USMLE1 ($t = 2.21$; $P = .03$) and SJT scores ($t = 2.97$; $P = .005$) were significantly associated with overall resident performance.

Discussion

This study used correlation and hierarchical regression analyses to assess the extent to which EQ, personality, and SJT scores obtained from 51 residents were associated with their performance in a large general surgery residency program 1 year later. The results showed that the USMLE1 score accounted for a reasonable level of criterion variance and was significantly associated with resident performance, likely because of its strong association with the ABSITE score. The USMLE2 score, however, demonstrated no significant association with our criterion. These findings suggested that, although the USMLE was originally created to inform licensure decisions, the use of USMLE1 scores as 1 component in the resident selection decision can be supported at this institution.

Despite increasing interest in the construct of EQ in the surgery literature,^{22,23} our data do not support the use of EQ assessments as a screening tool for general surgery residency applicants. We were unable to find significant associations between any of the facets of EQ or overall EQ with any of our performance criteria. A recent review of studies assessing EQ in surgery by McKinley and Phitayakorn²⁴ concluded that no study found a significant link between surgical resident EQ and clinical performance. Even more recently, Hollis et al²⁵ were unable to correlate EQ with either ABSITE scores or faculty evaluations of clinical competency. Thus, despite the growing interest in EQ measures in the surgical community, no data currently exist to support their use as a selection tool.

Personality assessments are often used for applicant selection in industries outside of medicine because such assessments have been shown to have reasonable validity evidence

and result in less potential discrimination of protected groups.^{26,27} In fact, approximately two-thirds of medium to large organizations use some type of personality or aptitude test in applicant screening.²⁸ The present study did not find a direct association in a regression model between any personality factor examined and overall performance. Correlation analyses did, however, indicate a positive association between evaluations received from medical students and both extraversion and agreeableness, such that the more outgoing and kind residents were, the higher their evaluation scores were from the students. The data also revealed a positive association between independence and case log numbers, suggesting that residents who were less reliant on others were more likely to take advantage of opportunities to participate in surgical procedures. Thus, personality factors may contribute to important indicators of success in residency but may not play a sufficiently strong role to have a direct association with overall performance criteria that do not heavily weigh medical student evaluations and procedural activity. However, programs that place more importance on medical student evaluations and procedural activity may find that these personality factors are important factors associated with performance.

The SJT assessment in the present study consisted of written common clinical scenarios presenting residents with challenging situations likely to be encountered in residency. Residents had to make judgments regarding the potential responses under a degree of uncertainty, a concept that is receiving increasing attention in the medical education literature.^{3,29-33} Residents were scored against a predetermined key defined by 12 clinical faculty members entrenched in the surgical education milieu. The results indicated a positive association between performance on the SJT and overall resident performance. The findings showed that, in this institution, the SJT estimated performance significantly better than a traditional cognitive measure (ie, USMLE score) alone. This finding can considerably contribute to how residency programs screen applicants. By developing customized tools that ask applicants to respond to unique situations that are likely to be encountered in a particular residency program, decision makers may have an opportunity to not only identify who will be successful in that program but also display the organization's unique culture and values.

Organizational consultants have noted that the use of the SJT for screening applicants provides a realistic job preview, giving the applicant common scenarios in which they would be placed on a frequent basis. In specialties that experience high rates of attrition, such as surgery,^{9,13} implementing SJTs for candidate selection may be additionally valuable. In fact, the Association of American Medical Colleges (AAMC) is already undertaking preliminary work to incorporate SJTs into medical student selection.³³ The results of the present study and efforts such as those of the AAMC support the role of SJTs in medical trainee selection.

In addition, the powerful association between SJTs and performance observed in the present study aligns with efforts to enhance diversity in surgery.³⁴ For programs actively pursuing these efforts, inclusion of nondiscriminatory screening tools that can estimate later performance is needed. Traditional tests

of general mental ability and specific cognitive abilities (eg, numerical, verbal, or spatial ability) have elicited concerns regarding fairness because these tests can result in substantial racial differences in test performance that are not matched in job performance.³⁵ As such, the use and weight given to written examinations, such as the USMLE, during the screening process may not align with efforts to enhance diversity. Other screening tools, such as SJTs, have been shown to be equally as associated with performance as cognitive-based assessments but without the discriminatory potential.³⁶ Thus, as indicated by our results, SJTs not only may offer predictive value in estimating performance in a residency but also may play a key role in enhancing diversity in surgical training programs.

Once sufficient data are accumulated to support the use of SJTs and other innovative screening tools, programs have a number of options regarding how and when to use these tools during the screening process. One of the most efficient methods may be to screen all applicants for eligibility and then invite eligible applicants to participate in an online assessment tool that must be completed in a timed setting. This process can provide program directors with standardized and program-specific information that can then be used to identify which individuals should be invited for the next round of screening, whether that consists of another round of assessments, a telephone interview, or simply fewer applicants invited to an on-site interview. Ultimately, the goal is to enhance the quality and relevance of data available to program directors, enabling them to make more informed decisions during the application review and interview invitation process.

Limitations

There are some limitations to our findings. First, these data are from a single specialty in a single institution, making the generalizability of these findings to other surgery programs and specialties unknown. However, because this institution is one of the largest general surgery residency programs in the country, there is little opportunity to create a more robust evaluation within a single institution. Multi-institutional studies can be conducted to further investigate these associations, but the unique values,

and performance measures within each program would need to be thoughtfully considered. In addition, despite the rigor with which the resident assessments and processes were created and collected, these evaluations are subject to biases prevalent across medical educational settings.^{37,38} To our knowledge, no other study has examined resident performance in such a robust manner by creating an overall “performance equation” that consists of weighted values of faculty evaluations, medical student evaluations, departmental staff evaluations of professionalism and administrative responsibilities, in-training examinations, procedural activities, and scholarship. Finally, SJT development, assessment administration, and data analyses were resource intensive. Programs without access to individuals with knowledge and experience in these domains may be unable to adopt these processes, thus limiting distribution of this methodology. However, nonmedical industries have overcome this limitation by using expert consultants in the science of selection to help create the necessary infrastructure, reasoning that they gain a return on their investment through reduced employee attrition and remediation rates. Thus, residency programs without such resources that are interested in replicating or exploring this methodology may similarly benefit by seeking professional consultation. As noted by Sklar,³⁹ better information is not the complete solution; the right people with the right training who know what to do with the information that is collected are also required.

Conclusions

The goal of this study was to explore the extent to which 3 distinct assessments—EQ, personality profiles, and SJTs—offered enough evidence to support their use in resident selection. We found little support for the use of EQ and weak support for some distinct personality factors (ie, agreeableness, extraversion, and independence). However, performance on an SJT assessment better estimated overall performance of residents 1 year later than traditional cognitive measures (ie, USMLE scores) used alone. These data support further exploration of these screening assessments on a larger scale across specialties and institutions.

ARTICLE INFORMATION

Accepted for Publication: September 4, 2017.

Published Online: December 27, 2017.
doi:10.1001/jamasurg.2017.5013

Author Contributions: Dr Gardner had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Both authors.

Acquisition, analysis, or interpretation of data: Both authors.

Drafting of the manuscript: Both authors.

Critical revision of the manuscript for important intellectual content: Both authors.

Statistical analysis: Both authors.

Obtained funding: Both authors.

Administrative, technical, or material support: Gardner.

Study supervision: Gardner.

Conflict of Interest Disclosures: Drs Gardner and Dunkin reported providing advice on selection methodology and assessment through SurgWise Consulting, LLC, in which they have ownership interest.

REFERENCES

- Bandiera G, Abrahams C, Ruetalo M, Hanson MD, Nickell L, Spadafora S. Identifying and promoting best practices in residency application and selection in a complex academic health network. *Acad Med.* 2015;90(12):1594-1601.
- Bandiera G, Maniate J, Hanson MD, Woods N, Hodges B. Access and selection: Canadian perspectives on who will be good doctors and how to identify them. *Acad Med.* 2015;90(7):946-952.
- Gardner AK, Grantcharov T, Dunkin BJ. The science of selection: using best practices from

industry to improve success in surgery training. *J Surg Educ.* 2017;S1931-7204(17)30352-30355.

4. Makdasi G, Takeuchi T, Rodriguez J, Rucinski J, Wise L. How we select our residents—a survey of selection criteria in general surgery residents. *J Surg Educ.* 2011;68(1):67-72.

5. Wagoner NE, Suriano JR. Program directors' responses to a survey on variables used to select residents in a time of change. *Acad Med.* 1999;74(1):51-58.

6. Keck JW, Arnold L, Willoughby L, Calkins V. Efficacy of cognitive/noncognitive measures in predicting resident-physician performance. *J Med Educ.* 1979;54(10):759-765.

7. Selber JC, Tong W, Koshy J, Ibrahim A, Liu J, Butler C. Correlation between trainee candidate selection criteria and subsequent performance. *J Am Coll Surg.* 2014;219(5):951-957.

8. Kluger MD, Vigano L, Barroso R, Cherqui D. The learning curve in laparoscopic major liver resection. *J Hepatobiliary Pancreat Sci*. 2013;20(2):131-136.
9. Yaghoobian A, Galante J, Kaji A, et al. General surgery resident remediation and attrition: a multi-institutional study. *Arch Surg*. 2012;147(9):829-833.
10. Sanfey H, Williams R, Dunnington G. Recognizing residents with a deficiency in operative performance as a step closer to effective remediation. *J Am Coll Surg*. 2013;216(1):114-122.
11. Williams RG, Roberts NK, Schwind CJ, Dunnington GL. The nature of general surgery resident performance problems. *Surgery*. 2009;145(6):651-658.
12. Bergen PC, Littlefield JH, O'Keefe GE, et al. Identification of high-risk residents. *J Surg Res*. 2000;92(2):239-244.
13. Longo WE. Attrition: our biggest continuing challenge. *Am J Surg*. 2007;194(5):567-575.
14. Bastian VA, Burns NR, Nettelbeck T. Emotional intelligence predicts life skills, but not as well as personality and cognitive abilities. *Pers Individ Dif*. 2005;39(6):1135-1145. doi:10.1016/j.paid.2005.04.006
15. Brackett MA, Mayer JD. Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Pers Soc Psychol Bull*. 2003;29(9):1147-1158.
16. Brackett MA, Mayer JD, Warner RM. Emotional intelligence and its relation to everyday behavior. *Pers Individ Dif*. 2004;36:1387-1402. doi:10.1016/S0191-8869(03)00236-8
17. Lyons JB, Schneider TR. The influence of emotional intelligence on performance. *Pers Individ Dif*. 2005;39(4):693-703. doi:10.1016/j.paid.2005.02.018
18. Jackson DN, Paunonen SV, Tremblay PF. *Six Factor Personality Questionnaire*. Port Huron, MI: Sigma Assessments Systems; 2000.
19. Lievens F, Patterson F. The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *J Appl Psychol*. 2011;96(5):927-940.
20. Lievens F, Sackett PR, Buyse T. The effects of response instructions on situational judgment test performance and validity in a high-stakes context. *J Appl Psychol*. 2009;94(4):1095-1101.
21. McDaniel MA, Hartman NS, Whetzel DL, Grubb WL. Situational judgment tests, response instructions, and validity: a meta-analysis. *Pers Psychol*. 2007;60(1):63-91. doi:10.1111/j.1744-6570.2007.00065.x
22. McKinley SK, Petrusa ER, Fiedeldey-Van Dijk C, et al. A multi-institutional study of the emotional intelligence of resident physicians. *Am J Surg*. 2015;209(1):26-33.
23. Jensen AR, Wright AS, Lance AR, et al. The emotional intelligence of surgical residents: a descriptive study. *Am J Surg*. 2008;195(1):5-10.
24. McKinley SK, Phitayakorn R. Emotional intelligence and simulation. *Surg Clin North Am*. 2015;95(4):855-867.
25. Hollis RH, Theiss LM, Gullick AA, et al. Emotional intelligence in surgery is associated with resident job satisfaction. *J Surg Res*. 2017;209:178-183.
26. Barrick MR, Mount MK. The Big Five personality dimensions and job performance: a meta-analysis. *Pers Psychol*. 1991;44(1):1-26. doi:10.1111/j.1744-6570.1991.tb00688.x
27. Tett RP, Jackson DN, Rothstein M. Personality measures as predictors of job performance: a meta-analytic review. *Pers Psychol*. 1991;44(4):703-742. doi:10.1111/j.1744-6570.1991.tb00696.x
28. Beagrie S. How to... excel at psychometric assessments. *Personnel Today*. 2005:25.
29. Lievens F, Sackett PR. The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *J Appl Psychol*. 2012;97(2):460-468.
30. Lievens F, Coetsier P. Situational tests in student selection: an examination of predictive validity, adverse impact, and construct validity. *Int J Sel Assess*. 2002;10(4):245-257. doi:10.1111/1468-2389.00215
31. Cooke S, Lemay JF. Transforming medical assessment: integrating uncertainty into the evaluation of clinical reasoning in medical education. *Acad Med*. 2017;92(6):746-751. doi:10.1097/ACM.0000000000001559
32. Gardner AK, Ritter EM, Paige JT, Ahmed RA, Fernandez G, Dunkin BJ. Simulation-based selection of surgical trainees: considerations, challenges, and opportunities. *J Am Coll Surg*. 2016;223(3):530-536.
33. Association of American Medical Colleges. Situational judgment test (SJT) research. <https://www.aamc.org/admissions/admissionslifecycle/409100/situationaljudgmenttest.html>. Published 2017. Accessed March 6, 2017.
34. Gardner AK. How can best practices in recruitment and selection improve diversity in surgery? [published online September 15, 2017]. *Ann Surg*. 2017. doi:10.1097/SLA.0000000000002496
35. Outtz JL. The role of cognitive ability tests in employment selection. *Hum Perform*. 2002;15(1-2):161-171. doi:10.1207/S15327043HUP1501&02_10
36. Whetzel DL, McDaniel MA, Nguyen NT. Subgroup differences in situational judgment test performance: a meta-analysis. *Hum Perform*. 2008;21(3):291-309. doi:10.1080/O8959280802137820
37. Williams RG, Mellinger JD, Dunnington GL. A problem-oriented approach to resident performance ratings. *Surgery*. 2016;160(4):936-945.
38. Gardner AK, Scott DJ. Repaying in kind: examination of the reciprocity effect in faculty and resident evaluations. *J Surg Educ*. 2016;73(6):e91-e94.
39. Sklar DP. Who's the fairest of them all? meeting the challenges of medical student and resident selection. *Acad Med*. 2016;91(11):1465-1467.