2022 APDS SPRING MEETING

Identifying Potential Attrition during the Residency Applicant Screening Process Using a Situational Judgment Test

Ross E. Willis, PhD, Jason W. Kempenich, MD, Ronit Patnaik, MD, and Daniel L. Dent, MD

Department of Surgery, UT Health San Antonio, San Antonio, Texas

OBJECTIVE: As USMLE Step 1 moves to pass/fail, residency programs are seeking alternate interview selection processes. Attrition in general surgery is reported as high as 26%. Thus, it is important to ensure that programs are selecting and matching applicants with shared values. Situational judgment tests (SJTs) measure educational and cultural values by posing ambiguous situations and individuals rate the effectiveness of possible reactions. SJTs have previously been shown to identify job applicants with shared values while promoting diversity. Scoring categories are high, moderate, or low values congruence. We sought to explore predictive validity of the SJT relative to program attrition.

DESIGN: Residents who matched into our program between 2018 and 2021 completed the SJT. We tracked attrition.

SETTING: UT Health San Antonio, Texas

PARTICIPANTS: Fifty-six categorical general surgery residents

RESULTS: Per SJT ratings, the numbers of residents who had high, moderate, and low values congruence were 27, 16, and 13, respectively. Attrition numbers for residents who scored high and moderate congruence were similar, indicating that these ratings were indistinguishable. As such, we combined those 2 categories to create a 2×2 matrix and used signal detection theory as a framework for analysis. Overall attrition was 16.1% (9/56). Of the 43 residents who scored high or moderate congruence, 90.7% remained in the program. There was a 9.3% chance of attrition for these residents. Of the 13 residents who scored low congruence, 38.5% attrited. While scoring as low congruence on the SJT does not definitively indicate attrition, it does indicate that attrition is 4.14 times more likely for these residents (chi-square, $p = 0.0121$).

CONCLUSIONS: One of the most important aspects of residency applicant selection and interviewing is mitigating risk by identifying applicants who carry a high risk of attrition. The SJT significantly identifies at-risk applicants. (J Surg Ed 000:1–6. © 2022 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: selection, interviewing, situational judgment test, attrition

COMPETENCIES: Professionalism

INTRODUCTION

Attrition in general surgery is estimated to be between 17% and 40%[1-4] for categorical residents. Attrition presents hardships for the individual resident, fellow residents, program directors, and the training program. Attriting residents may feel they have "wasted time" in the wrong specialty.[2] Peer residents may experience disruption and increased workload. Program Directors may have to expend additional effort recruiting replacement residents at variable points in the residency program.[1] Programs may suffer in terms of reputation and low morale.[5]

Methods of reducing attrition seem to fall into 2 philosophies. The first philosophy, improving retention, may be accomplished via improved support systems to help struggling residents, modifying the work environment, and giving early exposure to surgical procedures.[1,2,6] The focus of this paper is on a second philosophy, improving selection.

Improving selection focuses on finding the right people for the program. The current selection process relies heavily on USMLE scores. The National Resident Matching Program conducted a survey in which program

Correspondence: Inquiries to Ross E. Willis, PhD, Department of Surgery, UT Health San Antonio, 7703 Floyd Curl Drive, Mail Code 7742, San Antonio, TX 78229.; e-mail: willisr@uthscsa.edu

 © 2022 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved. 1931-7204/$30.00 https://doi.org/10.1016/j.jsurg.2022.07.020

directors were asked to specify which of 33 factors they used when selecting applicants for interview as well as the importance of each factor.[7] The vast majority of surgery program directors reported using USMLE Step 1 and Step 2 (94% and 81%, respectively). In terms of ranking applicants, surgery program directors also reported using USMLE Step 1 and 2 scores (81% and 78%, respectively).

The USMLE was designed for the purpose of licensure, not selection.[8,9] Perhaps the most concerning adverse issue with USMLE being used for selection is that studies have shown sex[10] and racial biases,[10,11] which may make it difficult for programs to achieve diversity goals. Now that USMLE Step 1 has been changed to pass/fail, more than 88% of general surgery program directors say they will require Step 2 for residency applicants and will place more weight on this exam.[12] Doing so will further perpetuate the problems of using exam scores that spuriously predict performance and potentially jeopardize diversity.

Perhaps a more meaningful selection strategy is to identify individuals with shared values.[13,14] Situational judgment tests (SJTs) measure educational and organization-specific cultural values by posing ambiguous situations and individuals rate the effectiveness of possible reactions. An example STJ scenario is shown in Table 1. SJT scores range from high to low values congruence and are based on concordance of the applicant's ratings of reactions compared to those of individuals currently working within the organization. Scoring algorithms can be customized to an individual organization to maximize relevance. As such an individual applicant can receive a score of high values congruence for one organization while simultaneously receiving a score of low values congruence for another organization, even when the organizations use the exact same SJT questions and responses.

SJTs have been used for medical school selection in the United Kingdom and Belgium and for residency selection in Belgium, Singapore, Canada, and Australia.[15] SJTs have been shown to predict medically-related job performance,[16] scores on a clinical skills Objective Structured Clinical Examination,[17] and ACGME milestones scores.[18] SJTs have also been shown to promote diversity.[19]

We sought to explore predictive validity of a SJT relative to our program attrition over a 4-year period.

METHODS

In 2018, we employed the services of 2 Industrial/Organizational Psychologists at SurgWise Consulting, Houston, Texas, USA. The psychologists spent several days conducting a job analysis via structured interviews with several key faculty and senior residents at our program. The goal was to obtain data about our program's core values. Eleven themes emerged from the interview transcripts:

1 Communication skills
2 Interpersonal skills
3 Motivational fit
4 Adaptability
5 Integrity
6 Perseverance
7 Problem solving
8 Professionalism
9 Self-directed learning
10 Stress management
11 Teamwork

A SJT with 20 scenarios was developed around the 11 values themes. A program-specific scoring algorithm was

TABLE 1. Example of SJT Scenario

	Not at All Effective	Slightly Effective	Moderately Effective	Very Effective	Extremely Effective
Tell your attending the complaints are without merit	○	○	○	○	○
Speak with the EM faculty to inquire more about how your behavior is being perceived	○	○	○	○	○
Apologize to EM faculty and monitor your behavior closely	○	○	○	○	○
Do nothing and keep to yourself when in the ED	○	○	○	○	○
Speak with the Program Director about these issues	○	○	○	○	○

created based on input from our faculty and senior residents' responses to the scenarios. Each applicant's responses are scored in terms of high, moderate, or low values congruence.

We began asking residency applicants to complete the SJT as part of our interview screening process in academic year 2018. In academic years 2019, 2020, and 2021, 22 applicants matched into our program as categorical residents. To assess concurrent validity, we asked 54 categorical residents to retroactively complete the SJT in 2018 and 34 complied (63% completion rate). Thus, we have SJT data for 56 residents. We tracked attrition of these residents from 2018 to 2021. Data were analyzed with a chi-square statistical test.

RESULTS

During the time period of this study, our program was comprised of 143 categorical residents; 51.7% female (74/143) and 48.3% male (69/143). Fifty-six residents completed the SJT, 53.6% were female, 46.4% were male. Chi-square revealed that the distribution of residents who completed the SJT was not significantly different than the distribution of all categorical residents in our program (p = .82).

Overall attrition was 16.1% (9/56). The gender distribution of attrition was 55.6% female (5/9) and 44.4% male (4/9). Again, chi-square revealed that the distribution of residents who completed the SJT was not significantly different than the distribution of all categorical residents in our program (p = .83).

Per SJT ratings, 27 residents had high values congruence, 2 of whom left the program, 16 had moderate values congruence, 2 of whom attrited, and 13 had low values congruence, 5 of whom attrited. Attrition numbers for residents who scored high and moderate congruence were equal indicating that these SJT ratings were indistinguishable. As such, we combined these 2 categories to create a 2 × 2 matrix and used signal detection theory as a framework for analysis as shown in Table 2.

TABLE 2. Frequencies of All Residents Who Completed the SJT and Whether They Remained in Program or Attrited

	High & Moderate Values Congruence N = 43	**Low Values Congruence N = 13**
Remained in Program	**HIT** (higher is better) 39/43 = 90.7%	**False alarm** (lower is better) 8/13 = 61.5%
Attrited	**MISS** (lower is better) 4/43 = 9.3%	**Correct rejection** (higher is better) 5/13 = 38.5%

A signal detection theory matrix is comprised of 2 axes: Actual state of affairs and predicted state of affairs. "Hits" and "correct rejections" are defined as instances where the predicted state of affairs matches the actual state of affairs. Specifically, a hit occurs when the SJT predicted an individual would remain in the program and the individual actually did stay in the program. A correct rejection occurs when the SJT predicted an individual will attrite and the individual actually did attrite. Higher values in these quadrants of the matrix are desirable. "Misses" and "false alarms" are defined as instances where the predicted state of affairs does not match the actual state of affairs. A "miss" occurs when the SJT predicted an individual will remain in the program but the individual actually attrited. A false alarm occurs when the SJT predicted an individual would attrite but the individual actually remained in the program. Lower values in these quadrants of the matrix are desirable.

Of the 43 residents who scored high or moderate values congruence, 90.7% remained in the program (i.e., "hit") and 9.3% left the program (i.e., "miss"). Of the 13 residents who scored low values congruence, 61.5% remained in the program (i.e., "false alarm") and 38.5% attrited (i.e., "correct rejection"). Chi-Square indicated a significant difference in attrition as a function of SJT values congruence score (p = 0.0121). Residents who were identified as having low values congruence on the SJT were 4.14 times more likely to attrite (i.e., miss vs correct rejection). An alternative way of looking at the accuracy of the SJT is that it was correct 78.6% of the time (i.e., hits and correct rejections) and incorrect 21.4% of the time (i.e., misses and false alarms). The SJT was right 3.7 times more often than it was wrong.

Because the SJT is intended to be used as a screening tool to select candidates for interviews, we analyzed the data for only the 22 residents who took the SJT as part of the application process and subsequently matched into our program. Results are presented in Table 3 and indicate a significant difference in attrition as a function of SJT values congruence score (p = 0.0062). The SJT correctly identified applicants who would remain in our program 100% of the time and correctly identified applicants who would attrite 40% of the time for a correct identification ratio of 86.4% (i.e., hits and correct rejections). Conversely, the SJT made an incorrect prediction 13.6% of the time (i.e., misses and false alarms). The SJT was right 6.3 times more often than it was wrong when only examining the data from the residents who completed it during the application process.

For the 34 residents who took the SJT retroactively (i.e., while they were residents in our program), the SJT was not significantly accurate at determining attrition

TABLE 3. Frequencies of the 22 Residents Who Completed the SJT During the Application Process and Whether They Remained in Program or Attrited

	High & Moderate Values Congruence N = 17	Low Values Congruence N = 5
Remained in program	**HIT** (higher is better) 17/17 = 100%	**False alarm** (lower is better) 3/5 = 60%
Attrited	**MISS** (lower is better) 0/17 = 0%	**Correct rejection** (higher is better) 2/5 = 40%

(p = 0.1761, Table 4). For these residents, the SJT predicted specific individuals would remain in our program 84.6% of the time and correctly identified individuals who would attrite 37.5% of the time for a correct identification ratio of 73.5% (i.e., hits and correct rejections), which was not statistically significant.

DISCUSSION

With reported attrition rates as high as 40% in general surgery, many educators have noted a need to match trainees that "fit" specific programs' characteristics.[13,14,20,21] We are careful to avoid the term "fit" because it is often used to mean "like-me/us." Modest and colleagues point out that fit may be the antithesis of diversity and call for a change in terminology that incorporates values congruence and cultural enhancement.[13]

There is increasing evidence which suggests that SJTs can help identify individuals with shared values,[15] predict performance,[16-18] correlate with a traditional method of rank list creation[22] and improve diversity.[19] Our data indicate that the SJT did not definitively identify attrition. Indeed, we experienced approximately 60% false alarms depending upon whether examining all residents who completed the SJT or only residents who completed the SJT as applicants to our program. These are instances wherein the SJT predicted an individual would leave our program when the individual did not. While it may alarm some program directors to know that the SJT is incorrectly marking applicants as high-risk, we posit that one of the most important aspects of residency applicant selection and interviewing is to mitigate risk by identifying applicants who carry a higher risk of attrition. Given that the SJT correctly identified high-risk for attrition individuals at a rate 4 times greater than it incorrectly identified high-risk for attrition individuals, we feel these false alarms are acceptable. This is especially true when the SJT results are viewed within the context of other data such as personal statements, letters of recommendation, or having observed the individual in the clinical setting.

TABLE 4. Frequencies of the 34 Residents Who Completed the SJT Retroactively and Whether They Remained in Program or Attrited

	High & Moderate Values Congruence N = 26	Low Values Congruence N = 8
Remained in program	**HIT** (higher is better) 22/26 = 84.6%	**False alarm** (lower is better) 5/8 = 62.5%
Attrited	**MISS** (lower is better) 4/26 = 15.4%	**Correct rejection** (higher is better) 3/5 = 37.5%

This study is not without limitations. First, 39.3% of the residents in this study completed the SJT as applicants (i.e., before becoming residents in our program) whereas 60.7% completed the SJT retroactively (i.e., when they were residents in our program). There are no published data on whether performance on a SJT remains constant or variable over time. However, our data indicated that the SJT was more accurate in terms of predicting attrition when applied to interview candidates rather than existing residents. This is encouraging since SJTs are intended to serve as one of many screening tools. Since a goal of this study was to examine concurrent validity in terms of attrition, we opted to include data from study participants who completed the SJT as residents and provide both analyses.

The second limitation is that these data represent the experience of only 1 institution. While other institutions are slowly adopting SJTs into their screening process, we were unable to find an institution that wanted to collaborate. This may have been because during the timeframe of this study, we interviewed and matched applicants who were identified as having poor values congruence per the SJT. Other institutions may have more or less of a threshold for considering applicants who scored as having low values congruence on the SJT.

Program directors may be leery of implementing a process outside of the traditional application process for fear that applicants will be disinclined to apply to their programs. Gardner et al.[23] created SJTs for 7 general surgery programs. A total of 1625 applicants applied to at least one of these programs, which represents approximately 70% of all United States medical students applying to general surgery. Completion rate was 97%. It is evident that an additional "hurdle" in the application process does not dissuade candidates.

CONCLUSION

The SJT significantly identified individuals at-risk for attrition. An SJT may be used as a screening tool to select individuals who have moderate or high values congruence rather than relying heavily on USMLE scores.

REFERENCES

1. Dodson TF, Webb AL. Why do residents leave general surgery? The hidden problem in today's programs. *Curr Surg*. 2005;62:128–131.
2. Farley DR, Cook JK. Whatever happened to the General Surgery graduating class of 2001? *Curr Surg*. 2001;58:587–590.
3. Kwakwa F, Jonasson O. Attrition in graduate surgical education: an analysis of the 1993 entering cohort of surgical residents. *J Am Coll Surg*. 1999;189:602–610.
4. Longo WE, Seashore J, Duffy A, Udelsman R. Attrition of categoric general surgery residents: results of a 20-year audit. *Am J Surg*. 2009;197:774–778. ; discussion 9-80.
5. 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.
6. Yeo H, Bucholz E, Ann Sosa J, et al. A national study of attrition in general surgery training: which residents leave and where do they go? *Ann Surg*. 2010;252:529–534. discussion 34-6.
7. National Residency Matching Program. Results of the 2018 NRMP Program Director Survey. Washington, DC2018. 156-8 p.
8. McGaghie WC, Cohen ER, Wayne DB. Are United States Medical Licensing Exam Step 1 and 2 scores valid measures for postgraduate medical residency selection decisions? *Acad Med*. 2011;86:48–52.
9. Nungester RJ, Dawson-Saunders B, Kelley PR, Volle RL. Score reporting on NBME examinations. *Acad Med*. 1990;65:723–729.
10. Rubright JD, Jodoin M, Barone MA. Examining demographics, prior academic performance, and United States Medical Licensing Examination Scores. *Acad Med*. 2019;94:364–370.
11. Edmond MB, Deschenes JL, Eckler M, Wenzel RP. Racial bias in using USMLE step 1 scores to grant internal medicine residency interviews. *Acad Med*. 2001;76:1253–1256.
12. Pontell ME, Makhoul AT, Ganesh Kumar N, Drolet BC. The change of USMLE step 1 to Pass/Fail: perspectives of the Surgery Program Director. *J Surg Educ*. 2021;78:91–98.
13. Modest JM, Cruz AI Jr., Daniels AH, Lemme NJ, Eberson CP. Applicant fit and diversity in the orthopaedic surgery residency selection process: defining and melding to create a more diverse and stronger residency program. *JB JS Open Access*. 2020;5:1–4.
14. Kelz RR, Mullen JL, Kaiser LR, et al. Prevention of surgical resident attrition by a novel selection strategy. *Ann Surg*. 2010;252:537–541.
15. Patterson F, Zibarras L, Ashworth V. Situational judgement tests in medical education and training: research, theory and practice: AMEE Guide No. 100. *Med Teach*. 2016;38:3–17.
16. 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.
17. Patterson F, Lievens F, Kerrin M, Munro N, Irish B. The predictive validity of selection for entry into postgraduate training in general practice: evidence from three longitudinal studies. *Br J Gen Pract*. 2013;63:e734–e741.
18. Cullen MJ, Zhang C, Marcus-Blank B, et al. Improving our ability to predict resident applicant performance: validity evidence for a situational judgment test. *Teach Learn Med*. 2020;32:508–521.
19. Gardner AK, Cavanaugh KJ, Willis RE, Dunkin BJ. Can better selection tools help us achieve our diversity goals in postgraduate medical education? Comparing use of USMLE Step 1 scores and situational judgment tests at 7 surgical residencies. *Acad Med*. 2020;95:751–757.
20. Dudas LM, Khan US, Borgstrom DC. Mastering the Match: recruiting a successful surgery resident. *Curr Surg Rep*. 2021;9:27.
21. Rajesh A, Asaad M. Alternative strategies for evaluating general surgery residency applicants and an interview limit for MATCH 2021: an impending necessity. *Ann Surg*. 2021;273:109–111.

22. Shipper ES, Mazer LM, Merrell SB, et al. Pilot evaluation of the computer-based assessment for sampling personal characteristics test. *J Surg Res*. 2017; 215:211–218.

23. Gardner AK, Cavanaugh KJ, Willis RE, Dunkin BJ. If you build it, will they come? Candidate completion of preinterview screening assessments. *J Surg Educ*. 2019;76:1534–1538.