

Making progress on identifying those who aren't making progress: Using situational judgment tests to predict those at risk for remediation and attrition

Aimee Gardner[1], Brian Dunkin[2]

Corresponding author: Dr Aimee Gardner aimee.gardner@bcm.edu

Institution: 1. Baylor College of Medicine, SurgWise Consulting, 2. Houston Methodist Hospital, SurgWise Consulting

Categories: Assessment, Postgraduate (including Speciality Training), Selection

Received: 30/11/2018

Published: 07/12/2018

Abstract

Introduction

There is little evidence on tools that can be used to identify individuals most at risk for remediation and attrition in postgraduate surgical training. We explore the extent to which a situational judgment test (SJT) can predict which trainees are likely to require remediation or leave the program.

Methods

Postgraduate trainees in a single general surgery residency program in the United States completed a 50-item SJT. Data regarding remediation actions and attrition were retrieved from the program for the two years following completion of the assessment. United States Medical Licensing Examination (USMLE) scores were also included to examine their ability to predict remediation and attrition.

Results

Complete data was available from fifty-two of 56 (92%) residents in the program. Twenty-one percent (11/52) of residents required at least one remediation intervention within the one (for PGY5s) to two years after completing the SJT. Mann-Whitney U tests revealed a significant difference in SJT performance between those needing remediation versus those who did not require remediation, such that those requiring remediation performed worse on the assessment ($p < 0.05$). Further exploration of remediation frequency by SJT score quartile reveals that trainees performing below top quartile of the SJT is associated with a 3-5 times higher likelihood of receiving a remediation intervention. Eight percent (4/52) of residents left the program within the one to two years after completing the SJT. There was not a significant difference in SJT scores between those who remained in the program 1-2 years after completing the assessment ($p = 0.41$). There were no differences in USMLE scores between the two groups in either remediation interventions (245 versus 244, $p = 0.75$) or attrition (243 versus 246, $p = 0.67$).

Conclusions

These data demonstrate that SJTs can be created to effectively identify surgery residents most at risk for remediation across a two-year timeframe. These data provide an additional layer of validity evidence to support the role of SJTs in surgical education and align with other studies failing to find linkages between USMLE scores and residency performance criterion.

Keywords: selection; situational judgment test; USMLE; remediation; attrition; surgery

Introduction

Modern surgical training is intense, fast-paced, and rigorous. To succeed in this environment, trainees must be able to quickly adapt to new environments, master new surgical techniques and technologies, and successfully work within dynamic teams. As a result of this unique environment, many postgraduate trainees fail to meet program expectations during their training and require additional focused interventions to get back on track. In the United States (U.S.), nearly one third of general surgery trainees require remediation during their five years of post-medical school training, with 25% of those remediated requiring more than one intervention (Yaghoubian et al., 2012). Attrition rates are high as well, ranging from 20% - 30% and 62% of those who leave general surgery training go on to pursue a non-surgical specialty (Longo et al., 2009; Yeo et al., 2017).

Many training programs have performed retrospective reviews of the reasons and types of remediation in their programs, in hopes of identifying early on those trainees at risk. Interestingly, despite the technical demands of surgical training, deficiency in technical skills is an uncommon reason for remediation and a rarely a reason for dismissal (Khoushhal et al., 2017). When examining non-technical factors, Yaghoubian et al. (2012) reported a deficiency in medical knowledge as the most frequent reason for remediation in six US training programs (74%), while others have found that problems with professionalism and interpersonal skills were more frequent (Bergen et al., 2000). Williams et al. (2009) examined the performance of surgical trainees in their program over a 30-year period and found that serious performance problems fell into three main categories: interpersonal skills, insufficient knowledge, and communication skills. Unfortunately, despite specific attempts at helping these trainees get back on track, 88% were refractory to remediation.

Given the high incidence of non-technical and non-cognitive issues requiring remediation in surgical training, it is not surprising that current methods of evaluating trainees or screening applicants for training fail to identify those at risk. In the U.S., traditional screening of residency applicants includes review of U.S. Medical Licensing Examination (USMLE) scores, medical school grades, letters of recommendation, personal statements, and unstructured interviews. However, these tools have been criticized not only for their inability to predict performance in residency, but also to identify those most at risk for remediation and attrition (Gardner, Grantcharov, and Dunkin, 2018; McGaghie, Cohen and Wayne, 2011; Fryer et al., 2012; Stohl, Hueppchen, and Bienstock, 2010; Sutton et al., 2014; Zuckerman et al., 2018). Thus, there is a need to identify and develop tools that can predict who may be likely to struggle in surgical training.

In this study, we explore the extent to which a situational judgment test (SJT) can predict which trainees are likely to require remediation or leave the program. SJTs present candidates with a variety of situations they would likely to encounter if they were brought into the organization, and are asked to make judgments about the effectiveness of possible responses. SJTs have become widespread across organizations outside of medicine, likely because of the robust body of validity evidence that exists showing their ability to predict on-the-job performance (Clevenger et al. 2001, Weekly and Ployhart 2001), reduce potential for adverse impact (Hough, Oswald, and Ployhart, 2001), and provide applicants with realistic job previews (Clevenger et al., 2001). Their multi-dimensional nature and capacity

to measure the ability *to do* rather than the ability *to know* by placing candidates "in the situation" separates SJTs from more traditional, straight-forward assessments such as multiple-choice questions, short answer questions, and the like. As such, we explore the following research questions in this study:

1. *Can SJT scores be used to predict remediation and attrition in residency?*
2. *To what extent can SJT scores be used to complement examination scores in predicting resident remediation and attrition?*

Methods

Sample

Participants included postgraduate trainees in a single general surgery residency program in the United States. The program was chosen because of its relatively large number of residents and robust assessment database.

Situational Judgment Test

A pool of 50 situational judgment test items were created by a team of MDs and PhDs. Semi-structured interviews with subject matter experts (SMEs) and review of historical resident performance data were used to identify which competencies were critical for residents to be successful in the program. Data from these interviews informed the development of SJT items, similar to prior studies (Gardner and Dunkin 2018a; Lievens, Peeters, and Schollaert, 2008). Faculty SMEs central to the residency education program (Program Director, Associate Program Directors, members of the Clinical Competency Committee, etc.) reviewed all items and provided input regarding effectiveness of each item to inform which items ended up on the final tool and the scoring algorithm.

The final SJT was given to all postgraduate year (PGY) 1 – 5 residents at the beginning of the academic year. Results from the assessment were neither used for any developmental or assessment purpose nor were they shared with program faculty.

Outcome Measures

Remediation data, defined as any documented action taken by program leadership to improve a resident deficit, were collected for the following two years. All remediation events were captured as dichotomous (yes/no) variables and categorized according to the Accreditation Council for Graduate Medical Education (ACGME) core competencies: medical knowledge, patient care, interpersonal and communication skills, professionalism, systems-based practice, and problem-based learning and improvement.

Resident attrition data were also collected. Residents who left the program prior to graduation for any reason (involuntary termination, transfer to another surgery program, transfer to another specialty, exiting medicine for another occupation, etc.) were included in the turnover group according to a dichotomous (yes/no) labeling.

Analyses

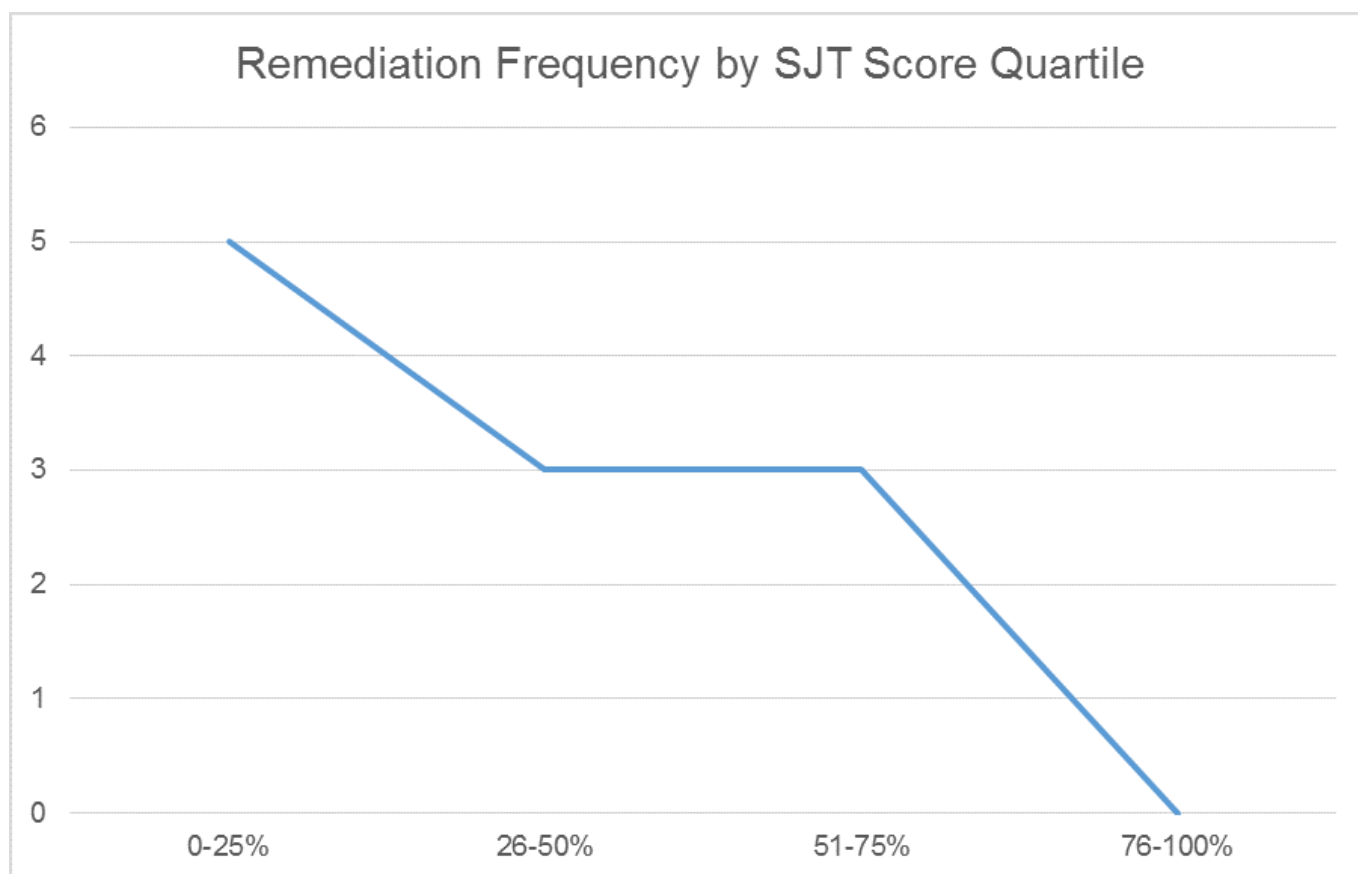
Mann-Whitney U tests and χ^2 were used to compare the SJT score distributions for trainees who were marked positive for either remediation or attrition over the two-year time frame. As USMLE 1 scores are traditionally used as the primary initial screening tool among residency Program Directors in the United States (Program Director Survey, 2018), these data were also included as a potential predictor for remediation and turnover.

Results

Complete data were available for 93% (52/56) of residents in the program. Residents consisted of categorical PGY1s (25%), PGY2s (19.2%), PGY3s (25%), PGY4s (17.3%), and PGY5s (13.5%).

Twenty-one percent (11/52) of residents required at least one remediation intervention within the one (for PGY5s) to two years after completing the SJT. The majority of these interventions were related to deficits in professionalism (56%), with the remaining related to decision-making (33%) or technical skills (11%). Mann-Whitney U tests revealed a significant difference in SJT performance between those needing remediation versus those who did not require remediation, such that those requiring remediation performed worse on the assessment ($p = 0.04$). Further exploration of remediation frequency by SJT score quartile reveals that trainees performing below top quartile of the SJT is associated with a 3-5 times higher likelihood of receiving a remediation intervention. Figure 1 displays these data based on performance on the SJT.

Figure 1. Remediation Frequency by SJT Score Quartile



Eight percent (4/52) of residents left the program within the one to two years after completing the SJT. Two residents (PGY2 and PGY3) transferred to another general surgery program and the remaining two (both PGY1s) joined another medical specialty. There was not a significant difference in SJT scores between those who remained in the program 1-2 years after completing the assessment ($p = 0.41$).

Finally, we explore the relationship among USMLE1 scores and these variables. SJT and USMLE1 scores were not correlated. Additionally there were no differences in USMLE scores between the two groups in either remediation interventions (245 versus 244, $p = 0.75$) or attrition (243 versus 246, $p = 0.67$).

Discussion

Identifying residents at risk for performance deficiencies and turnover is an important, but complex, task for training program leaders. Problem residents can put a strain on the program director and clinical faculty, increase the workload of other healthcare providers, and complicate patient care activities (Williams et al., 2009). Additionally, program morale and reputation can suffer with high levels of attrition (Gardner, Grantcharov, and Dunkin, 2018). Thus, identifying residents at risk for performance deficiencies or turnover is critical not only for trainee and program success, but also so adequate interventions can be implemented in time to avoid potential patient care issues.

We explored the extent to which SJTs could be created and used to accurately identify residents at risk for remediation, and compared its utility to the most common screening tool currently used – licensing examination scores. By following residents in a single institution over time, we found that SJTs can indeed predict those at risk for remediation. Similar to other work (Cousans et al., 2017), our data suggests that high scores on an SJT assessment is associated with about a five times reduction in likelihood of remediation. For a specialty that remediates approximately 1 in 3 trainees at least once (Bergen et al., 2000), and of which efforts bring substantial financial costs (Gardner et al., 2018), these findings have considerable practical value to surgical educators. For example, the program involved in this study could administer their customized SJT to new interns to identify individuals who are more likely to struggle during clinical training and develop educational interventions accordingly. Additionally, these data contribute to the increasingly robust chain of validation reasoning that suggests SJTs may play an important role in postgraduate medical education selection (Gardner and Dunkin, 2018a; Gardner and Dunkin, 2018b; Patterson et al. 2017).

This study failed to find support in this sample for using the SJT to identify who might leave residency prior to graduation. Although the four individuals who did leave the program within the two-year timeframe did have lower scores on the SJT, the differences among this small group were not robust enough to reach significance. Additionally, given that the majority of attrition occurs within the first few years of surgical training (Leibrandt, Fassler, and Morris, 2004), the PGY4s and PGY5s within our sample likely represent those least at risk for turnover. Multi-institutional studies with more robust sample sizes are likely needed to properly explore these relationships.

Our findings also contribute to the burgeoning literature showing the limited value of using USMLE scores for selection (McGaghie, Cohen, and Wayne, 2011). Although some studies have been able to link USMLE1 scores to other test scores during and after training (Fryer et al. 2012; de Virgilio et al. 2010), many have been unable to identify a positive relationship between these scores and outcomes of interest in surgical training, including faculty evaluations (Brothers and Wetherholt 2007; Fryer et al. 2012; McGaghie, Cohen, and Wayne, 2011; Stohl, Hueppchen, and Bienstock, 2010; Sutton et al., 2014; Zuckerman et al., 2018), professionalism metrics (Gardner and Dunkin, 2018a), and operative experience (Gardner and Dunkin 2018a). This lack of validity evidence, combined with concerns about how reliance on USMLE for selection may be at odds to increase diversity in surgical training (Gardner, 2018; Gardner and Dunkin, 2018b; Higgins, 2018), suggests program leaders should explore development and adoption of tools that can be used for residency selection.

As with any study, these findings are not without limitations. First, these data include just one sample of surgery trainees within a single institution. Although the institution examined represents one of the largest general surgery training programs in the U.S., and thus includes the most robust data available from just one program, it is nonetheless a single institution study. However, given that the prevalence of remediation and attrition in this sample was lower than the national rate, these data may actually underestimate the role of the SJTs in predicting these outcomes. Further, the USMLE scores in this sample are well above the national mean. It is unknown if inclusion of a wider array of USMLE scores would have resulted in different outcomes. Finally, although the remediation data

included here was retrieved from program records and thus reflect the most serious instances that were addressed, it may not fully capture all instances wherein a trainee may have received coaching from a number of different supervisors or colleagues. As with any exercise in measuring efficacy of interventions or tools, we are at the mercy of the assessment data available.

Conclusion

These data demonstrate that SJTs can be created to effectively identify surgery residents most at risk for remediation across a two-year timeframe. These data provide an additional layer of validity evidence to support the role of SJTs in surgical education and align with other studies failing to find linkages between USMLE scores and residency performance criterion.

Take Home Messages

- Tools are lacking to identify surgery residents most at risk for remediation and attrition
- SJTs can be created to effectively identify surgery residents most at risk for remediation across a two-year timeframe

Notes On Contributors

Aimee K. Gardner is Co-Founder of SurgWise Consulting and Assistant Dean of Research & Evaluation at Baylor College of Medicine.

Brian J. Dunkin, MD is Co-Founder of SurgWise Consulting and Professor of Surgery at Houston Methodist Hospital.

Acknowledgements

None.

Bibliography/References

- Bergen PC, Littlefield JH, O'Keefe GE, et al. (2000) 'Identification of high-risk residents.' *Journal of Surgical Research*, 92, 239-244. <https://doi.org/10.1006/jsre.2000.5924>
- Brothers TE, Wetherholt S. (2007) 'Importance of the faculty interview during the resident application process.' *Journal of Surgical Education*, 64, 378-385. <https://doi.org/10.1016/j.jsurg.2007.05.003>
- Clevenger J., Pereira GM, Wiechmann D, Schmitt N., Schmidt-Harvey V. (2001) 'Incremental validity of situational judgment tests.' *Journal of Applied Psychology*, 86: 410-417. <https://doi.org/10.1037/0021-9010.86.3.410>
- Cousans F, Patterson F, Edwards H, Walker K, McLachlan JC, et al. (2017) 'Evaluating the complementary roles of an SJT and academic assessment for entry into clinical practice.' *Advances in Health Science Education*, 22: 401-413. <https://doi.org/10.1007/s10459-017-9755-4>

De Virgilio C, Yaghoubian A, Kaji A, Collins JC, Deveney K, et al. (2010) 'Predicting performance on the American Board of Surgery qualifying and certifying examinations.' *Archives of Surgery*, 145: 852-856.

<https://doi.org/10.1001/archsurg.2010.177>

Fryer JP, Corcoran N, George B, Wang E, DaRosa D. (2012) 'Does resident ranking during recruitment accurately predict subsequent performance as a surgical resident?' *Journal of Surgical Education*, 69: 724-730.

<https://doi.org/10.1016/j.jsurg.2012.06.010>

Gardner AK, Dunkin BJ. (2018a) 'Evaluation of validity evidence for personality, emotional intelligence, and situational judgment tests to identify successful residents.' *JAMA Surgery*, 153: 409-416.

<https://doi.org/10.1001/jamasurg.2017.5013>

Gardner AK, Dunkin BJ. (2018b) 'Pursuing excellence: The power of selection science to enhance diversity, increase efficiency, and provide meaningful data for selecting surgical trainees.' *Annals of Surgery*,

<https://doi.org/10.1097/SLA.0000000000002806>

Gardner AK, Grantcharov T, Dunkin BJ. (2018) 'The science of selection: Using best practices from industry to improve success in surgery training.' *Journal of Surgical Education*, 75: 278-285.

<https://doi.org/10.1016/j.jsurg.2017.07.010>

Gardner AK, Smink DS, Scott BG, Korndorffer JR, Harrington D, et al. (2018) 'Money talks: How much are we spending on resident selection?' *Journal of Surgical Education*, <https://doi.org/10.1016/j.jsurg.2018.10.001>

Gardner AK. (2018) 'How can best practices in recruitment and selection improve diversity in surgery?' *Annals of Surgery*, 267: e1-e2. <https://doi.org/10.1097/SLA.0000000000002496>

Higgins R. (2018) 'Inside baseball – leveling the playing field in the surgical residency selection process.' *Annals of Surgery*, 267: e3. <https://doi.org/10.1097/SLA.0000000000002522>

Hough LM, Oswald FL, Ployhart RE. (2001) 'Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned.' *International Journal of Selection and Assessment*, 9: 152-194. <https://doi.org/10.1111/1468-2389.00171>

Khoushhal Z, Hussain MA, Greco E, Mamdani M, Verma S, et al. (2017) 'Prevalence and causes of attrition among surgical residents. A systematic review and meta-analysis.' *JAMA Surgery*, 152: 265-272.

<https://doi.org/10.1001/jamasurg.2016.4086>

Leibrandt TJ, Fassler SA, Morris JB. (2004) 'Attrition and replacement of general surgery residents.' *Surgical Clinics of North America*, 84: 1525-1535. <https://doi.org/10.1016/j.suc.2004.06.011>

Lievens F, Patterson F, Corstjens J, Martin S, Nicholson S. (2016) 'Widening access in selection using situational judgment tests: Evidence from the UKCAT.' *Medical Education*, 50: 624-636. <https://doi.org/10.1111/medu.13060>

Lievens F, Peeters H, Schollaert E. (2008) 'Situational judgment tests: A review of recent research.' *Personnel Review*, 37: 426-441. <https://doi.org/10.1108/00483480810877598>

Longo WE, Seashore J, Duffy A, Udelsman R. (2009) 'Attrition of categorical general surgery residents: results of a 20-year audit.' *American Journal of Surgery*, 197: 774-778. <https://doi.org/10.1016/j.amjsurg.2008.06.038>

McGaghie WC, Cohen ER, Wayne DB. (2011) 'Are United States Medical Licensing Exam step 1 and 2 scores valid measures for postgraduate medical residency selection decisions?' *Academic Medicine*, 86: 48-52.

<https://doi.org/10.1097/ACM.0b013e3181ffacdb>

Patterson F, Cousans F, Edwards H, Rosselli A, Nicholson S, et al. (2017) 'The predictive validity of a text-based

situational judgment test in undergraduate medical and dental school admissions.’ *Academic Medicine*, 92: 1250-1253. <https://doi.org/10.1097/ACM.0000000000001630>

National Resident Matching Program. (2018) *Results of the 2018 NRMP Program Director Survey*.
<https://www.nrmp.org/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf>
Accessed on November 30, 2018.

Stohl HE, Hueppchen NA, Bienstock JL. (2010) ‘Can medical school performance predict residency performance? Resident selection and predictors of successful performance in obstetrics and gynecology.’ *Journal of Graduate Medical Education*, 322-326. <https://doi.org/10.4300/JGME-D-09-00101.1>

Sutton E, Richardson JD, Ziegler C, Bond J, Burke-Poole M, et al. (2014) ‘Is USMLE step 1 score a valid predictor of success in surgical residency?’ *American Journal of Surgery*, 208: 1029-1034.
<https://doi.org/10.1016/j.amjsurg.2014.06.032>

Weekly JA, Ployhart RE. (2001) ‘Situational judgment: Antecedents and relationships with performance.’ *Human Performance*, 18: 81-104. https://doi.org/10.1207/s15327043hup1801_4

Williams RG, Roberts NK, Schwind CJ, Dunnington GL. (2009) ‘The nature of general surgery resident performance problems.’ *Surgery*, 145: 651-658. <https://doi.org/10.1016/j.surg.2009.01.019>

Yaghoubian A, Galante J, Kaji A, et al. (2012) ‘General surgery resident remediation and attrition.’ *Archives of Surgery*, 147: 829-833. <https://doi.org/10.1001/archsurg.2012.1676>

Yeo H, Abelson J, Mao J, Lewis F, Michelassi F, et al. (2017) ‘Who makes it to the end? A novel predictive model for identifying surgical residents at risk for Attrition.’ *Annals of Surgery*, 266, 499-507.
<https://doi.org/10.1097/SLA.0000000000002384>

Zuckerman SL, Kelly PD, Dewan MC, Morone PJ, Yengo-Kahn AM, et al. (2018) ‘Predicting resident performance from preresidency factors: A systematic review and applicability to neurosurgical training.’ *World NeuroSurgery*, 110: 475-484. <https://doi.org/10.1016/j.wneu.2017.11.078>

Appendices

None.

Declarations

The author has declared the conflicts of interest below.

Both authors are co-founders of SurgWise Consulting, which provides advice on assessment to surgical training programs and organisations.

This has been published under Creative Commons "CC BY 4.0" (<https://creativecommons.org/licenses/by-sa/4.0/>)

Ethics Statement

Approved by the University of Texas Institutional Review Board, STU: 032015104.

External Funding

This paper has not had any External Funding

AMEE MedEdPublish: rapid, post-publication, peer-reviewed papers on healthcare professions' education. For more information please visit www.mededpublish.org or contact mededpublish@dundee.ac.uk.