CASE STUDY

Making progress on identifying those who aren’t making progress: Using situational judgment tests to predict those at risk for remediation and attrition [version 1]

Aimee Gardner¹, Brian Dunkin²

¹Baylor College of Medicine
²Houston Methodist Hospital

Abstract
This article was migrated. The article was marked as recommended.

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

Conclusions

Open Peer Review

Migrated Content
"Migrated Content" refers to articles submitted to and published in the publication before moving to the current platform. These articles are static and cannot be updated.

John Cookson, University of Worcester
Ken Masters, Sultan Qaboos University
Ariela Marshall, Mayo Clinic
Megan Anakin, University of Otago
Barbara Griffin, Macquarie University
Balakrishnan Nair, Centre for Medical Professional Development and University of Newcastle

Please see the article web page for additional reviewers.

Any reports and responses or comments on the article can be found at the end of the article.
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 (Khoushal 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 χ² 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 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.

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).

**Figure 1. Remediation Frequency by SJT Score Quartile**
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
SJT}s 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.

**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 organizations.

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

**External Funding**
This article has not had any External Funding

**Bibliography/References**

Bergen, PC, Littlefield, JH, O’Keefe, GE, et al. (2000) Identification of high-risk residents. *Journal of Surgical Research.* 92: 239-244.  
Reference Source

Brothers, TE, Wetherholt, S. (2007) Importance of the faculty interview during the resident application process. *Journal of Surgical Education.* 64: 378-385.  
Reference Source

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.  
Reference Source

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: 401-418.  
Reference Source

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.  
Reference Source

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.  
Reference Source

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.  
Reference Source

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.* 267: e3.  
Reference Source

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.  
Reference Source

Gardner, AK, Smink, DS, Scott, BG, Kenndorff, JR, Harrington, D, et al. (2018) Money talks: How much are we spending on resident selection? *Journal of Surgical Education.*  
Reference Source

Gardner, AK. (2018) How can best practices in recruitment and selection improve diversity in surgery? *Annals of Surgery.* 267: e1-e2.  
Reference Source

Higgins, R. (2018) Inside baseball – leveling the playing field in the surgical residency selection process. *Annals of Surgery.* 267: e3.  
Reference Source

Hooge, LM, Oswald, FL, Plyhart, 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.  
Reference Source

Khousshal, 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.  
Reference Source

Leibrandt, TJ, Fassler, SA, Morris, JB. (2004) Attrition and replacement of general surgery residents. *Surgical Clinics of North America.* 84: 1525-1535.  
Reference Source

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.  
Reference Source

Lievens, F, Peeters, H, Schollaert, E. (2008) Situational judgment tests: A review of recent research. *Personnel Review.* 37: 426-441.  
Reference Source

Longo, WE, Seashore, J, Dufly, A, Udelman, R. (2009) Attrition of categorical general surgery residents: results of a 20-year audit. *American Journal of Surgery.* 197: 774-778.  
Reference Source

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.  
Reference Source

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.  
Reference Source

National Resident Matching Program. (2018) Results of the 2018 NRMP Program Director Survey.  
Reference Source

Accessed on November 30, 2018.

Stohl, HE, Huesphcn, NA, Biensstock, 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.  
Reference Source
Open Peer Review

Migrated Content

Version 1

Reviewer Report 23 July 2019

https://doi.org/10.21956/mep.19734.r29569

© 2019 Rogers G. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gary D. Rogers
Griffith University

This review has been migrated. The reviewer awarded 5 stars out of 5

Thank you, Aimee and Brian, for this well-written paper reporting what appears to have been a carefully-executed and important study. The positive findings, that your situational judgement tests (SJTs) predict which trainees will require remediation in relation particularly to professionalism issues in the subsequent one to two years, are truly impressive and surprising, given the concerns that have been raised about the susceptibility of SJTs to faking. This effect could be expected to be particularly prevalent among intellectually able candidates, such as those selected for surgical training. Accordingly, it would be really interesting to know more about how the SJTs were developed and how they appear to have avoided this effect – though I certainly take Barbara’s point about the risks of revealing too much detail in an environment that includes predatory commercial coaching companies. On the question of ‘what to do’ prospectively with the trainees for whom SJTs predict a high risk of professionalism breaches (apart from not selecting them!), there is some evidence pointing to the value of small-group educational interventions that promote affective domain learning through reflection on emotionally impactive clinical experiences (simulated or real) and deliberate connection with the professional values they highlight. It would be interesting to see if providing this kind of intervention to all trainees, prospectively, ‘flattens’ the effect of the SJT scores on the rate of requiring professionalism remediation.

Competing Interests: No conflicts of interest were disclosed.
© 2018 Griffin B. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Barbara Griffin
Macquarie University

This review has been migrated. The reviewer awarded 4 stars out of 5

Given the predatory nature of the commercial coaching industry, I am not surprised that the authors did not share details of their SJT items. However, even though the small number of participants in this study precludes generalisation, I really like the robust process of item development that was described. Using data from historical cases together with SME experience to both develop and review the items may well have contributed to the significant findings. An interesting question would be whether this particular SJT relates in the same way to performance at a different institution - such information would help us know if a bespoke test is needed for each training program's specific context.

Competing Interests: No conflicts of interest were disclosed.

Reviewer Report 17 December 2018

https://doi.org/10.21956/mep.19734.r29565

© 2018 Nair B. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Balakrishnan Nair
Centre for Medical Professional Development and University of Newcastle

This review has been migrated. The reviewer awarded 4 stars out of 5

It is good to see research on non-cognitive aspects of trainees. At the end of the day, unprofessional behaviour is what gets trainees into difficulty with regulatory bodies and institutions. As the authors state 1/3 of trainees in surgical training will require remediation. Some of these trainees are likely to abandon training. Any tool which may help to identify these trainees in difficulty, early enough, would be most welcome. Most of medicine is not “black and white”. It is in the grey area where most trainees have difficulty. The script concordance tests have been popular for this reason. In this interesting study, trainees took part in Situation Judgement Testing and were followed up to assess requirements for remediation; poor test scores correlated with higher rates of remediation. Examples of SJT would have been useful, as were provided in the studies on script concordance testing. The study raises the question when to intervene: should we intervene with a low score or wait till something goes wrong? This is a
single institution study with a small sample size; it has internal validity, but generalisation may be difficult. It will need a larger sample size and multicentre involvement to address this.

**Competing Interests:** No conflicts of interest were disclosed.

Ariela Marshall
Mayo Clinic

This review has been migrated. The reviewer awarded 4 stars out of 5

This is an interesting and well done study and shows the potential utility of the situational judgment test for analyzing which residents may be at risk for remediation. Overall I think this makes a positive contribution to the literature regarding assessment. There are a few points which could be clarified/strengthened for future work:

1. The authors describe the creation of the SJT by a “team of MDs and PhDs” and the identification of “which competencies were critical for residents to be successful in the program.” It would be helpful to have a list of the competencies and the questions for other programs (especially in other specialties) who may wish to create their own SJTs. In addition, having this information accessible would be helpful in terms of the breakdown of the competencies – are they mostly professional, or are they technical/specialty-specific? This may help address how generalizable the findings are to non-surgical specialties.

2. The authors found that most of the remediation interventions were related to deficits in professionalism, but that 33% were related to decision making and 11% to technical skills. Again, it would be useful to have lists of the exact type of deficiencies (what types of “decision making” errors – medication choice? Test ordering? Level of severity of illness? Etc) to better determine the generalizability of the study.

3. It would be helpful, as another reviewer suggested, for the authors to describe “next steps” – I could see this expanding either (a) to other specialties within the same center or (b) to surgical specialties at other centers – the overall goal of course would be to all specialties at all centers, but it would be helpful to think about the next immediate goal.

**Competing Interests:** No conflicts of interest were disclosed.
Megan Anakin
University of Otago

This review has been migrated. The reviewer awarded 4 stars out of 5

Thank you for inviting me to review this article. Like the other three reviewers, I appreciated the well-written case study that explored the use of situational judgment tests to identify at-risk learners. This article will be of interest to educators working with learners in health professional education at any stage training. For example, at my medical school, we are seeking to enhance the identification of “non-technical and non-cognitive issues” in our students earlier in the programme. We have not discussed the use and benefits of situational judgment tests as a possible solution. As reviewer Ken Masters suggested, sharing the situational judgment test constructed by the authors would have given me something concrete to present to my colleagues for discussion. Nevertheless, this article will be useful to stimulate conversation about how we might approach building an instrument fit for the purpose of our educational programme that will enable us to detect students who would benefit from more support earlier in their training.

Competing Interests: No conflicts of interest were disclosed.

Ken Masters
Sultan Qaboos University

This review has been migrated. The reviewer awarded 4 stars out of 5

A succinctly and well-written case study, clearly laying out the issues (and their importance) to be addressed, and situating the study within the literature and practical context. In particular, as the authors note, the study highlights the weakness of relying on USMLE scores for selection. I hope that the authors are planning to expand their research into other centres, in order to determine the generalisability of their SJT. On that note, it is a pity that the authors have not shared their SJT; I believe that it would of value to others working in the field.
**Competing Interests:** No conflicts of interest were disclosed.

**Reviewer Report 07 December 2018**

https://doi.org/10.21956/mep.19734.r29566

© 2018 Cookson J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**John Cookson**  
University of Worcester

This review has been migrated. The reviewer awarded 4 stars out of 5

Thank you. I was interested in this as there has been much discussion of SJTs on this (Europe) side of the pond. One issue is whether the assessors can agree on the correct answers. Another is what action to take as a result of this kind of study. This paper found an association between the SJT and remediation events; most of which were about professionalism. There was no correlation with USMLE results. I think one could have predicted this; indeed if there was no association with events one would wonder if the SJT was doing anything at all. Neither is it surprising that there was no correlation with the USMLE as this and similar assessments are clearly assessing a different domain to the SJT. The problem is what to do? The authors suggest early intervention with those with a low score which certainly seems reasonable, assuming that remediation works. The wider question is whether SJTs can predict failure. We need a prediction more than a association. How well does a low score on the SJT predict subsequent problems and how good does that prediction need to be before we take action such as terminating the course. This paper suggests that that is still some way off.

**Competing Interests:** No conflicts of interest were disclosed.