The Predictive Validity of a Text-Based Situational Judgment Test in Undergraduate Medical and Dental School Admissions

Fiona Patterson, MSc, PhD, CPsychol, Fran Cousans, MSc, Helena Edwards, MSc, CPsychol, Anna Rosselli, MSc, Sandra Nicholson, MRCGP, MSc, PhD, and Barry Wright, MRCPsych, MD

Abstract

Problem
Situational judgment tests (SJTs) can be used to assess the nonacademic attributes necessary for medical and dental trainees to become successful practitioners. Evidence for SJTs' predictive validity, however, relates predominantly to selection in postgraduate settings or using video-based SJTs at the undergraduate level; it may not be directly transferable to text-based SJTs in undergraduate medical and dental school selection. This preliminary study aimed to address these gaps by assessing the validity of the UK Clinical Aptitude Test (UKCAT) text-based SJT.

Approach
Study participants were 218 first-year medical and dental students from four UK undergraduate schools who completed the first UKCAT text-based SJT in 2013. Outcome measures were educational supervisor ratings of in-role performance in problem-based learning tutorial sessions—mean rating across the three domains measured by the SJT (integrity, perspective taking, and team involvement) and an overall judgment of performance—collected in 2015.

Outcomes
There were significant correlations between SJT scores and both mean supervisor ratings (uncorrected $r = 0.24$, $P < .001$; corrected $r = 0.34$) and overall judgments (uncorrected $r_s = 0.16$, $P < .05$; corrected $r_s = 0.20$). SJT scores predicted 6% of variance in mean supervisor ratings across the three nonacademic domains.

Next Steps
The results provide evidence that a well-designed text-based SJT can be appropriately integrated, and add value to, the selection process for undergraduate medical and dental school. More evidence is needed regarding the longitudinal predictive validity of SJTs throughout medical and dental training pathways, with appropriate outcome criteria.

Problem

> [T]he way we select physicians must change.… In addition to students who are academically prepared, we need to select those who also have strong interpersonal and social skills, and can demonstrate various intrapersonal competencies necessary to learn clinical skills.
>
> —Association of American Medical Colleges[1]

The Association of American Medical Colleges[1] has astutely summarized considerations at the forefront of current theory and practice in health care professions education selection: In addition to academic and clinical skills, nonacademic attributes such as empathy and the ability to work in teams are necessary for students and trainees to become successful practitioners. Therefore, it is critical to select for these attributes at the undergraduate level for medical and dental school.

Situational judgment tests (SJTs) are a selection method designed to assess candidates' judgment regarding scenarios encountered in a specific role,[2] and they can be used to measure candidates' nonacademic attributes. They are useful during high-volume admissions processes because they can be delivered via computer and marked by machine. There is a general consensus among researchers that SJTs—when appropriately designed and implemented—are favorably received by stakeholders and can form reliable and valid components of medical and dental school selection.[3]

Currently, there is limited evidence regarding the longitudinal predictive validity of SJTs for undergraduate medical and dental school selection. Instead, predictive validity evidence for SJTs relates predominantly to selection at the postgraduate level, usually for entry to specialty training,[4] and may not be transferable to undergraduate settings because of differences in test specifications between these contexts.

In a notable exception to this trend of postgraduate selection research, Lievens[5] reported good predictive and incremental validity of a video-based SJT for selection into medical school in Belgium. However, these findings may not directly relate to the *text*-based SJT added in 2013 to the UK Clinical Aptitude Test (UKCAT), a suite of tests used in the selection process for entry into a number of undergraduate (baccalaureate-level) medical and dental schools in the United Kingdom. Research by Lievens and Sackett[6] showed text-based SJTs to have less predictive validity for nonacademic outcomes than comparable video-based SJTs. This may be because video-based SJTs include more nuanced and nonverbal cues, which are important in making judgments about interpersonal situations. However, given the high volume of applicants in the United Kingdom, the expense of developing video-based SJTs (including actors, videographers, and studios) is prohibitive

Please see the end of this article for information about the authors.

Correspondence should be addressed to Fiona Patterson, Work Psychology Group, The Stanley Building, 7 Pancras Square, London, N1C 4AG; telephone: (+44) 203-8597700; e-mail: f.patterson@workpsychologygroup.com; Twitter: @WorkPsychGroup.

Acad Med. XXXX;XX:00–00.
First published online
doi: 10.1097/ACM.0000000000001630

compared with developing relatively cost-effective text-based SJTs.

The present research aims to address the gaps in the literature identified above by assessing the validity of the UKCAT text-based SJT in predicting educational supervisor ratings for students' in-role performance. In this preliminary study, we posed the following research questions:

1. What is the predictive validity of the text-based SJT used for selection into medical and dental school for supervisor ratings of in-role performance?
2. Are there significant differences between SJT scores of students rated by their supervisors as being "particularly promising" versus those rated as "likely to struggle"?

Approach

The participants in this 2015 study were 218 first-year medical and dental students (medical: n = 197 [90.4%]; dental: n = 21 [9.6%]) from four undergraduate schools (three medical, one dental) that volunteered to take part through the UKCAT Consortium. These participants represented 32.6% of the first-year intake of the four schools. They began their course in 2014, having completed the first live version of the UKCAT SJT in 2013.

Of the study participants, 119 (54.6%) were female, 136 (62.4%) were white, and 65 (29.8%) were black, Asian, or "minority ethnic"; 17 (7.8%) did not disclose their race/ethnicity. The mean age when taking the UKCAT was 18.6 years (SD = 2.5). These demographic data are broadly representative of the whole student cohort that completed the SJT in 2013.

The study participants (N = 218) had a higher mean SJT score (M = 211.3, SD = 13.2) compared with the overall population that took the test in 2013 (N = 25,587; M = 204.2, SD = 16.5). The study sample also contained fewer very low SJT scorers compared with the overall population, as can be seen through the minimum SJT scores (study sample: min = 168.5; overall population: min = 118.5; possible score range for the SJT: 0–245.5).

All four schools received ethical approval to participate in the research through their own institutional review boards. Immediately prior to data collection, students and the educational supervisors for students' problem-based learning (PBL) tutorial sessions were verbally briefed by a senior researcher (H.E.) at their school about the nature of the study and how data would be used before consenting to participate. All students signed an information sheet to indicate their informed consent for their selection data to be matched with their supervisor ratings and used on aggregate in publications. Students' data were received anonymously from supervisors, with UKCAT ID numbers used to match the in-training performance and SJT data. Where supervisors did not have access to student UKCAT ID numbers, school administrators matched supervisor ratings to UKCAT ID numbers based on the students' names.

Situational judgment test

The UKCAT text-based SJT targets three nonacademic attributes identified as important for success in the role of a medical or dental student: integrity, perspective taking, and team involvement. It uses a multiple-response format where candidates are asked to rate the appropriateness of various responses to a given situation. (Full details of the SJT format and example items can be found at http://www.ukcat.ac.uk/about-the-test/situational-judgement.)

For the UKCAT SJT in 2013, each student completed 66 items (16 scenarios) from one of six different test forms, which were equated to ensure equivalence of difficulty. As expected in selection contexts, there was evidence of range restriction for the SJT scores, where the study sample had a narrower spread of scores and an absence of extreme low scorers compared with the distribution of the overall population. Therefore, true score corrections for restriction of range were applied to the correlations (as outlined by Hunter and colleagues[7]) to estimate the magnitude of the correlation between predictor and outcome variables in a nonrestricted dataset. Mean internal consistency of the SJT across forms was $\alpha = .77$.

In-role performance

Students' educational supervisors completed an in-training performance questionnaire, either online or in text-based format, at a single time point during March–April 2015 (approximately seven months after the start of the students' undergraduate education). Supervisors were instructed to take performance across numerous PBL tutorial sessions into account when rating students (so the data, although collected once, captured students' overall performance). Previous research shows that educational supervisor ratings are a useful way to gather job performance outcome data.[8]

The questionnaire was designed using the relative percentile method (RPM). Each student's educational supervisor was asked to provide a single rating (out of 100) for the student's performance for each of the three SJT domains (integrity, perspective taking, and team involvement), rating the student *in comparison* with all of the first-year medical or dental students at their school. (This differs from more traditional methods of collecting supervisor ratings, which require supervisors to rate students *independently* of other students.) Evidence shows that the RPM can achieve a greater range of—and more accurate—scores when compared with rating each student independently.[9]

The three domain ratings were strongly correlated with one another ($r = 0.71$–0.84, $P < .001$). A principal components factor analysis of the three domain ratings showed that a single factor could explain 85% of the variance. Therefore, mean supervisor rating was used as the main criterion-matched outcome variable.

Additionally, supervisors were asked to make an overall judgment of whether they would describe the student as "particularly promising," "average," or "likely to struggle," allowing supervisors to identify weaker students more generally, beyond the criterion-matched outcome measures.

Outcomes

Histograms and *z* scores of students' SJT scores were examined for outliers. None were identified; therefore, all 218 cases were included in the analyses. Descriptive statistics for and correlations between the SJT and in-training performance measures are reported in Table 1. Frequencies of overall judgment categories are provided in Table 2. All correlations reported are Pearson *r*

Table 1

Correlations Between Situational Judgment Test (SJT) Scores and Educational Supervisors' In-Role Performance Ratings of First-Year Medical and Dental Students (N = 218)

Measure	Mean	SD	Correlation 1[a]	2	3
1. SJT score[b]	211.26	13.21	—		
2. Mean supervisor rating[c]	62.58	18.04	.24[d] (.34)	—	
3. Overall judgment[e]	—	—	.16[f] (.20)	.63[d]	—

Abbreviation: SD indicates standard deviation.
[a]Correlations in parentheses were corrected based on Hunter and colleagues' true score correction for indirect restriction of range.[7]
[b]Students took the text-based SJT as part of the UK Clinical Aptitude Test (UKCAT) in 2013. The range of SJT scores in the study sample was 168.5 to 240.0. The possible score range was 0 to 245.5.
[c]Educational supervisors for problem-based learning session tutorials rated student performance on the three nonacademic domains measured by the UKCAT SJT (integrity, perspective taking, and team involvement) using a scale of 1 to 100, in March–April 2015. The mean supervisor rating is the mean of these three domain ratings.
[d]$P < .001$.
[e]Educational supervisors rated students' overall performance as being "particularly promising," "average," or "likely to struggle," in March–April 2015.
[f]$P < .05$.

correlations, except for overall judgment where Spearman rho was used because data were nominal.

Predictive validity of the SJT for supervisor ratings of in-role performance

SJT scores demonstrated significant correlations with both mean supervisor ratings (uncorrected $r = 0.24$, $P < .001$; corrected $r = 0.34$) and overall judgments (uncorrected $r_s = 0.16$, $P < .05$; corrected $r_s = 0.20$).

After ensuring that assumptions for parametric analyses were met, we conducted a linear regression analysis to assess the predictive relationship between SJT scores and mean supervisor ratings. Table 3 shows the results of the linear regression analysis: SJT scores significantly predicted mean supervisor ratings, explaining 6% of the variance, which is a small effect size ($R^2 = 0.06$, $P < .001$).

Table 2

Frequency of, and Mean Situational Judgment Test (SJT) Scores by, Educational Supervisors' Overall Judgment Categories, First-Year Medical and Dental Students (N = 218)[a]

Overall judgment category	No. (%) of students	Mean SJT score
Particularly promising	80 (36.53)	213.39
Average	109 (49.77)	211.03
Likely to struggle	30 (13.70)	206.41

[a]Students took the text-based SJT as part of the UK Clinical Aptitude Test in 2013. The range of SJT scores in the study sample was 168.5 to 240.0; possible scores ranged from 0 to 245.5. Educational supervisors for problem-based learned session tutorials rated students' overall in-role performance in March–April 2015.

Differences between SJT scores for students rated as "particularly promising" vs. "likely to struggle"

A one-way between-groups ANOVA showed significant differences between SJT scores for students in the three overall judgment categories ($F[2,216] = 3.13$, $P < .05$). Bonferroni post hoc comparisons indicated that students rated as "likely to struggle" had significantly lower SJT scores than those rated as being "particularly promising" ($P < .01$). Those rated as "average" did not differ significantly in their SJT scores compared with students in either of the other two groups.

Next Steps

This is the first study to examine the predictive validity of an SJT in the context of UK medical and dental school admissions and, to our knowledge, the first validation study exploring a text-based SJT for undergraduate medical and dental school admissions. The results demonstrate predictive validity of the SJT for both a criterion-matched outcome (mean supervisor rating, matched to the three domains targeted by the UKCAT SJT) and a more general judgment about the students' in-role performance (overall judgment). Importantly, these findings also indicate that text-based SJTs may be appropriate for use in high-stakes, high-volume selection processes.

The magnitude of the relationships found between the text-based SJT scores and supervisor ratings of students' nonacademic performance (uncorrected $r = 0.24$ and $r_s = 0.16$ for mean supervisor rating and overall judgment, respectively), although relatively small, are comparable to those found by Lievens and colleagues.[5,6] Those authors reported uncorrected $r = 0.15$–0.34 between video-based SJT scores (for entry into undergraduate medical education in Belgium) and outcome measures similar to those used in this study. Our findings show that a text-based SJT may be equally as effective as a video-based SJT in predicting important nonacademic outcomes for medical and dental students. This has practical relevance because the time, resources, and expense required for designing and developing text-based SJTs are considerably less than for video-based SJTs. Text-based SJTs are a more financially viable and practicable approach to assessing candidates' nonacademic attributes for selection into undergraduate medical or dental training in the United Kingdom and internationally.

Nonetheless, the two SJT delivery formats undoubtedly have their differences, and it is likely that one may be more appropriate than the other in some settings as they may tap into different constructs. To date, theory development in this area has been limited; however, we suggest that text-based SJTs require candidates to make judgments based on imagination and inference from the text rather than from direct observation of interactions as in video-based SJTs. Current evidence suggests that both text- and video-based SJTs predict subsequent performance in clinical practice, indicating the value of both delivery formats. Future research should systematically explore differences between the two delivery formats in terms of the constructs they implicitly assess and, subsequently, their relative appropriateness for different settings. A potential avenue for future research and practice is a "hybrid" SJT, which combines both text- and video-based scenarios and responses in a single test.

Table 3

Linear Regression Analysis of 218 First-Year Medical and Dental Students' Situational Judgment Test (SJT) Scores With Mean Supervisor Ratings[a]

	Unstandardized *b*	SE *b*	*β*	*t*	*F*	R^0
Mean supervisor rating	0.33	0.09	0.24	3.61	12.04	0.06[b]

[a]Educational supervisors for problem-based learning session tutorials rated student in-role performance on the three nonacademic domains measured by the UK Clinical Aptitude Test (UKCAT) text-based SJT (integrity, perspective taking, and team involvement) using a scale of 1 to 100, in March–April 2015. The mean supervisor rating is the mean of these three domain ratings. Students took the UKCAT SJT in 2013. The range of SJT scores in the study sample was 168.5 to 240.0.
[b]$P < .001$.

This study measured the performance of a single cohort of undergraduate medical and dental students during the first year of their degree program, and as such, the predictive validity data reported are over a one-year interval. This relatively short-term approach has value because selection methods may be differentially predictive at different points of the medical training pathway,[10] and so it is beneficial to obtain early measures of predictive validity. Future research should track performance as the cohort progresses through undergraduate education, into postgraduate training, and beyond, to assess the longer-term predictive validity of the UKCAT SJT.

Mean supervisor rating was criterion-matched to the domains assessed by the SJT, in line with best practices for validation study design. However, choice of outcome measure in validation studies is a challenge for health care education researchers. The "criterion problem" (where researchers must determine what constitutes a "competent" physician to identify appropriate outcome criteria) means it is difficult to evaluate important aspects of in-role performance using a single measure. Ours was an imperfect outcome measure because it relied on a single supervisor's rating of each student's performance; this rating was subjective and taken at a single time point (although supervisors were encouraged to consider students' overall performance).

Nevertheless, this study contributes to existing knowledge on the use of text-based SJTs in undergraduate selection settings and provides preliminary evidence from which to design further longitudinal studies with additional outcome measures. Appropriate outcome measures for future research will vary depending on the students' stage of training and the specialty eventually chosen; these may include combinations of peer ratings of performance, self-reported performance, and patient feedback during clinical practice.

Acknowledgments: In addition to the UK Clinical Aptitude Test (UKCAT) Research Board members who reviewed the draft manuscript, the authors would like to thank the individuals at the four schools involved in the study who helped support data access and gathering and without whom this work would have not been possible.

Funding/Support: The authors gratefully acknowledge the funding provided by the UKCAT Research Board to support this study.

Other disclosures: F.P., F.C., H.E., and A.R. provide advice on selection methodology to Health Education England and UKCAT through the Work Psychology Group. S.N. is a board member and research panel lead for UKCAT. None of the authors receive royalties for medical schools' use of the UKCAT SJT.

Ethical approval: All four schools received ethical approval to participate in the research through their own institutional review boards. In agreement with the UKCAT Consortium and the participating schools, individual institutions are not named in this paper.

Previous presentations: A summary of this research was presented at the 2016 Ottawa Conference, Perth, Western Australia, March 19–23, 2016.

F. Patterson is founding director, Work Psychology Group Ltd., London, United Kingdom, and visiting researcher, Department of Psychology, Cambridge University, Cambridge, United Kingdom.

F. Cousans is associate research psychologist, Work Psychology Group, Derby, United Kingdom, and teaching fellow in occupational psychology, Department of Neuroscience, Psychology & Behaviour, University of Leicester, Leicester, United Kingdom.

H. Edwards is senior consultant psychologist, Work Psychology Group Ltd., Derby, United Kingdom.

A. Rosselli is consultant psychologist, Work Psychology Group Ltd., Derby, United Kingdom.

S. Nicholson is head, Centre for Medical Education, and head of student progression, Institute of Health Sciences Education, Queen Mary University of London, London, United Kingdom.

B. Wright is professor of child psychiatry, University of York, and academic lead for student support, Hull York Medical School, University of York, York, United Kingdom.

References

1 Association of American Medical Colleges. Situational judgment test research. https://www.aamc.org/admissions/admissionslifecycle/409100/situationaljudgmenttest.html. Accessed February 24, 2017.
2 Clevenger J, Pereira GM, Wiechmann D, Schmitt N, Harvey VS. Incremental validation of situational judgment tests. J Appl Psychol. 2001;86:410–417.
3 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.
4 Patterson F, Knight A, Dowell J, Nicholson S, Cousans F, Cleland J. How effective are selection methods in medical education? A systematic review. Med Educ. 2016;50:36–60.
5 Lievens F. Adjusting medical school admission: Assessing interpersonal skills using situational judgement tests. Med Educ. 2013;47:182–189.
6 Lievens F, Sackett PR. Video-based versus written situational judgment tests: A comparison in terms of predictive validity. J Appl Psychol. 2006;91:1181–1188.
7 Hunter JE, Schmidt FL, Le H. Implications of direct and indirect range restriction for meta-analysis methods and findings. J Appl Psychol. 2006;91:594–612.
8 Viswesvaran C, Ones DS, Schmidt FL. Comparative analysis of the reliability of job performance ratings. J Appl Psychol. 1996;81(5):557–574.
9 Goffin RD, Gellatly IR, Paunonen SV, Jackson DN, Meyer JP. Criterion validation of two approaches to performance appraisal: The behavioral observation scale and the relative percentile method. J Bus Psychol. 1996;11(1):23–33.
10 Ferguson E, Semper H, Yates J, Fitzgerald JE, Skatova A, James D. The "dark side" and "bright side" of personality: When too much conscientiousness and too little anxiety are detrimental with respect to the acquisition of medical knowledge and skill. PLoS One. 2014;9(2). doi: 10.1371/journal.pone.0088606.