Using a situational judgement test for selection into dental core training: a preliminary analysis

E. Rowett,*[1] F. Patterson,[1,2] F. Cousans[1] and K. Elley[3]

In brief

Describes the evaluation of a pilot situational judgement test for selection into UK Dental Core Training.	Demonstrates the potential suitability of using a situational judgement test for Dental Core Training selection, with good psychometric properties found.	Highlights that no significant group differences were found for gender and only minimal differences for ethnicity.

Objective and setting This paper describes the evaluation of a pilot situational judgement test (SJT) for selection into UK Dental Core Training (DCT). The SJT's psychometric properties, group differences based on gender and ethnicity, and candidate reactions were assessed. **Methods** The SJT targets four non-academic attributes important for success in DCT. Data were collected alongside live selection processes from five Health Education England local teams in the UK (N = 386). Candidates completed the pilot SJT and an evaluation questionnaire to examine their reactions to the test. **Results** SJT scores were relatively normally distributed and showed acceptable levels of internal reliability ($\alpha = 0.68$). Difficulty level and partial correlations between scenarios and SJT total score were in the expected ranges (64.61% to 90.03% and r = 0.06 to 0.41, respectively). No group differences were found for gender, and group differences between White and BME candidates were minimal. Most candidates perceived the SJT as relevant to the target role, appropriate and fair. **Conclusions** This study demonstrated the potential suitability of an SJT for use in DCT selection. Future research should replicate these preliminary findings in other cohorts, and assess the predictive validity of the SJT for predicting key training and practice-based outcomes.

Introduction

Many dentists in the UK enter Dental Core Training (DCT) after the one year Dental Foundation Training (DFT). DCT provides an opportunity to develop competencies to enhance their skills for general dental practice or to enter speciality training. DCT helps to consolidate skills acquired during undergraduate foundation training, and to develop and enhance new skills in specific areas of clinical practice across different environments. Entry to DCT is highly competitive, and until recently (2017), the selection process for DCT has varied by training region. Typically, however, shortlisting has comprised 'white space' application forms which have been scored against the person specification.

Successful candidates have then been invited to a selection centre which assesses clinical and academic skills. There is increasing recognition that non-academic attributes are important to assess as part of the selection process.

A popular method of assessing non-academic attributes as part of selection is a situational judgement test (SJT), the use of which has been gaining popularity across a range of occupations in recent years.[1,2] To the authors' knowledge, no previous study has explored the use of SJTs for entry into dental postgraduate training beyond foundation training. SJTs are used to assess individuals' reactions to a number of hypothetical job relevant scenarios, which reflect situations candidates are likely to encounter in a job role. SJTs are scored using pre-determined keys, developed in consultation with subject matter experts (SMEs).[3] Since SJTs are machine-marked, they have the potential to be cost-effective and efficient as either a shortlisting tool, or as an element of selection.[4] Machine marking also reduces concerns around inter-rater reliability inherent in other selection methods such as references or 'white space' application forms.

An SJT was successfully implemented for selection into DFT in the UK,[5] which performed well psychometrically and received largely positive candidate feedback,[5] despite concerns from some stakeholders.[6] DFT is a one year period of training including experiential learning within general dental practice, compared to the more advanced level of training completed during DCT. As a result, the success of the DFT SJT suggested that an SJT may also be suitable as part of DCT selection, however, since there is a difference in experience level and context across DFT and DCT, it would not be appropriate to use the same SJT for both.

SJTs have been used extensively in medical selection, showing good evidence of predictive validity at the point of entry to medical school,[7] foundation training[8] and speciality training.[9,10] The research evidence showing the utility of SJTs in other postgraduate healthcare settings prompted consideration of an SJT for DCT selection to assess non-academic attributes alongside the existing selection centres. With any new selection methodology, various psychometric and legal criteria must be satisfied

[1]Work Psychology Group, Derby, UK; [2]University of Cambridge, Cambridge, UK; [3]Health Education England – West Midlands, UK
*Correspondence to: Emma Rowett
Email: e.rowett@workpsychologygroup.com

including standardisation, reliability and fairness.[5] This paper reports the development and evaluation of a pilot SJT for DCT and explores the psychometric properties, fairness in relation to demographic differences in test score, and face validity of the SJT for candidates, in order to ascertain its suitability for use in selection for DCT.

Specifically, we addressed the following research questions:

- Are the psychometric properties of the SJT robust?
- Does the SJT produce scoring differences based on demographic groups (gender and ethnicity)?
- Do candidates react to the SJT positively and view it as fair – does the SJT have good face validity?

Method

Materials

Situational judgement test

The SJT targets four key professional attributes which were identified following a thorough job analysis of the role of a DCT in collaboration with SMEs; these are outlined in Table 1, with behavioural indicators listed for each.

SJT content was designed in line with best practice principles (see Figure 1 and Patterson *et al.*[4]). The SJT scenarios were set in the context of DCT, and were developed by trained psychologists in collaboration with SMEs, such as trainers. SMEs provided scenario material and initial scoring keys during telephone interviews, which were subsequently developed into full SJT scenarios and response options by psychologists. Scenarios were then reviewed
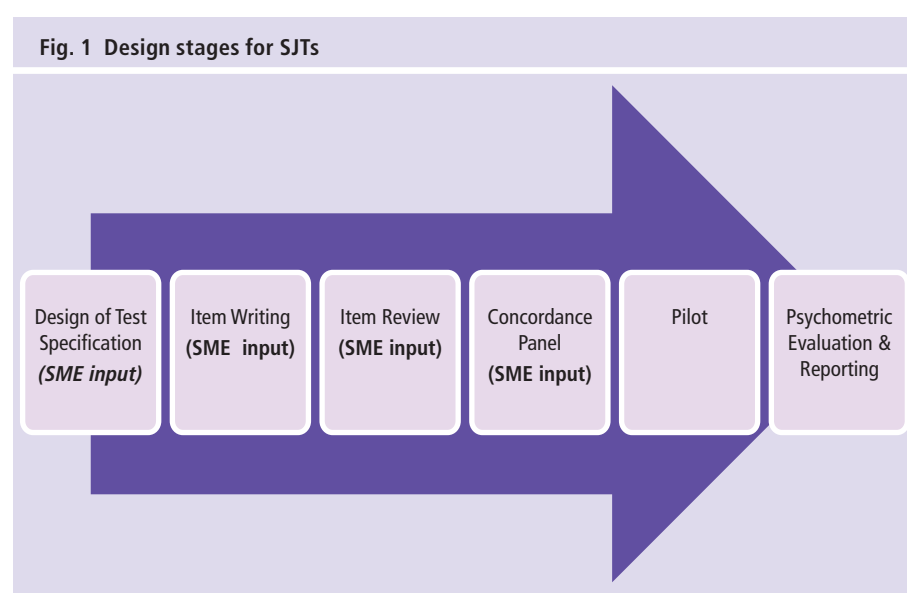
in focus groups with SMEs, to ensure their appropriateness and relevance to the DCT role. Following this, a different group of SMEs answered all scenarios in test conditions to finalise scenario material and scoring keys before piloting.

Two response formats were used, both of which require relative judgements to be made across response options:

- Ranking response format: candidates are required to rank five possible responses in order of appropriateness
- Multiple response format: candidates are required to select the three most appropriate response options from eight possible actions.

Response instructions were knowledge-based (candidates respond in terms of what

they should do in the given situation), which have been shown to have a higher cognitive loading than behavioural tendency instructions (that is, what they would do),[11] and are less susceptible to biases and faking.[12,13]

The final pilot SJT consisted of 35 scenarios (22 ranking response format with a total possible score of 20 points per scenario, and 13 multiple choice response format with a total possible score of 12 points per scenario) (Fig. 1).

Candidate feedback questionnaire

All candidates who completed the SJT were invited to respond to a previously validated feedback questionnaire based on procedural justice theory, presenting eight statements about their experience of completing the SJT. Candidates rated their level of agreement with each statement on a five-point Likert scale (1 =

Fig. 1 Design stages for SJTs

Design of Test Specification *(SME input)* → Item Writing *(SME input)* → Item Review *(SME input)* → Concordance Panel *(SME input)* → Pilot → Psychometric Evaluation & Reporting

Table 1 Professional attributes targeted by the SJT	
Critical thinking, appraisal and decision making	Synthesises and assimilates large amounts of information rapidly; accurately evaluates information, identifying key issues and risks; explores different options, weighing up the appropriateness of various courses of action; makes decisions and diagnoses based on sound clinical judgement & application of knowledge; recognises when to use own initiative to make independent decisions; seeks further relevant information where appropriate, from suitable sources; demonstrates attention to detail.
Professionalism	Demonstrates sensitivity to equality and diversity, respecting the perspectives of others; protects the confidentiality of patients' information; maintains professional and personal boundaries; demonstrates maturity, courtesy and professionalism in their approach and appearance; demonstrates honesty and integrity; values and adheres to formal standards, guidelines and procedures, for example, GDC requirements, and follows these as appropriate; proactively takes responsibility where appropriate, and is accountable for own mistakes; shows insight into their own strengths and limitations, willing to seek advice or assistance where appropriate; shows confidence in own abilities, working independently within these boundaries of competence, where appropriate.
Coping with Pressure	Demonstrates resilience and perseverance when faced with challenging circumstances; employs effective coping mechanisms to handle difficulties and challenges; attends to own physical and mental wellbeing; adapts quickly to changing circumstances, demonstrating flexibility; maintains a high level of performance even when working under pressure; keeps emotions under control in difficult or challenging situations; manages own time efficiently; takes proactive measures to ensure multiple priorities can be handled effectively.
Patient-centred care	Focuses on the patient's safety and best interests, complying with the duty of candour; delivers high quality patient care in a timely manner; shows commitment to providing excellent standards of patient care, demonstrating enthusiasm to exceed expectations; takes a holistic view to delivering patient care, considering each patient's personal circumstances and tailoring the treatment plan accordingly; demonstrates a non-judgemental, empathetic, courteous and sensitive approach, providing reassurance where appropriate.

Table 2 Descriptive and demographic data	
SJT	
Mean	442.76
SD	23.28
Min	314
Max	488
DCT Level	
1	202 (59.24%)
2/3 or Academic	103 (30.21%)
Missing	36 (10.56%)
Ethnicity	
White	123 (36.07%)
Black & Minority Ethnic	208 (61.00%)
Missing/do not wish to disclose	10 (2.93%)
Gender	
Male	115 (33.72%)
Female	209 (61.29%)
Missing/do not wish to disclose	12 (4.99%)

Fig. 2 Histogram of distribution of SJT scores
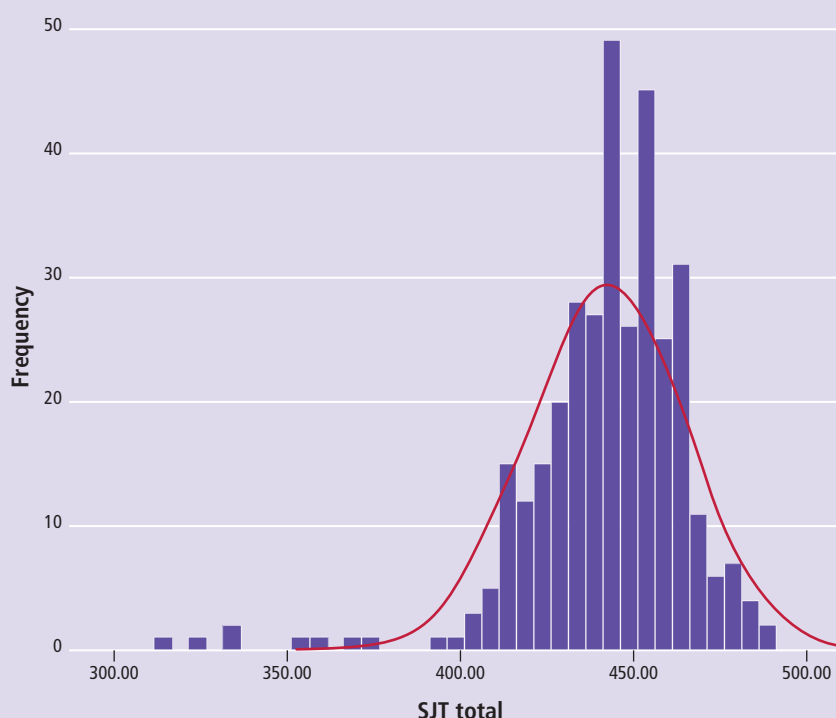
strongly disagree, 5 = strongly agree). We were interested in candidates' responses to individual statements such as: 'The content of the situational judgement test was clearly relevant to Dental Core Training' and 'The content of the situational judgement test appeared to be fair' in order to determine the level of agreement for these concepts. As in previous research,[13] this would enable the recruiters to better understand the candidates' perceptions of and reactions to the SJT. This, in turn, helps to facilitate any necessary interventions from the candidates' perspective.

Sample and procedure

Data collection took place during the 2015 selection process for DCT years 1, 2, 3 and academic positions in four Health Education England local teams (Yorkshire and the Humber, London, South West and West Midlands) and Scotland. All candidates at the five centres were invited to complete the SJT on the day of the selection in invigilated conditions, and they had 60 minutes to complete it. A total of 386 candidates completed the pilot SJT.

Ethics

All candidates included in the analysis signed a consent form indicating that they understood that their SJT scores and feedback forms would be used for research purposes only, that their participation was voluntary and would not impact their chances of being offered a training place.

Results

Candidates with a high number of missing data (>3 scenarios, as identified by analysis of the mean SJT scores by number of missing responses; N = 37) and those who did not complete consent forms (N = 8) were removed, leaving a total of 341 remaining participants for analysis (analyses were conducted to compare the data of candidates who were or were not included in the final sample; no significant differences were found). Descriptive and demographic data for candidates and the SJT are provided in Table 2.

Are the psychometric properties of the SJT robust?

An initial psychometric examination of the scenarios showed that three were performing badly and were therefore removed from the pilot analysis. Here we examined each scenario's partial correlation with the total SJT: this measures the extent to which the candidates' performance on that scenario correlates with their overall performance on the pilot SJT. We removed scenarios that had partial correlations of $r < 0.02$. The remaining 32 scenarios had partial correlations ranging from $r = 0.06–0.41$; with 20 (62.50%) of these having partial correlations that would be considered within the 'good' range ($r > 0.17$), and a further 7 (21.88%) in the 'moderate' range ($r > 0.13–0.16$). The remaining 5 items (15.63%) were of limited quality, as assessed by their partial correlations ($r < 0.12$).[15]

The difficulty of the scenarios (that is, how achievable it is to gain full marks on a scenario) was calculated as a percentage of the candidates' mean score on the scenario divided by the total possible score as follows:

Scenario difficulty =

$$\frac{\text{(Mean candidate score on scenario)}}{\text{(Maximum possible scenario score)}}$$

This ranged from 64.61% to 90.03% across the 32 scenarios. This range is comparable to those found in SJTs with similar test specifications that are used in other postgraduate settings.[5,16,17]

SJT total scores showed a close to normal distribution with a slight negative skew (see Figure 2; skewness = 1.86, kurtosis = 7.12).[18] This indicates that the test is capable of differentiating between candidates, which is an important requirement in selection contexts.

Results show the reliability of the SJT to be $\alpha = 0.68$, comparable to values reported for SJTs in other settings[19] and close to the accepted standard of $\alpha = 0.70$ as an indicator of good internal reliability.[20]

Does the SJT produce scoring differences based on demographic groups?

There were no significant differences between total SJT scores for males and females ($t(322) = 0.54$, P = .59). White candidates scored significantly higher than Black and Minority Ethnic (BME) candidates ($t(329) = 2.12$, P=.03), $d = 0.26$), although the effect size was small,[21] indicating minimal differences between groups on SJT performance.

Do candidates react to the SJT positively and view it as fair – does the SJT have good face validity?

All DCT candidates who completed the SJT also completed the feedback questionnaire (N = 386). The majority of candidates agreed or strongly agreed with the following statements: the SJT content was relevant to DCT (84.20%); the content of the SJT was appropriate for their training level (84.72%); the content of the SJT appeared to be fair (79.79%); and the level of difficulty of the SJT was appropriate (82.12%). These results provide good evidence for the face validity of the SJT and are commensurate with other findings of positive candidate perceptions towards SJTs for use in selection.[5,22]

Discussion

This study provides initial evidence for the suitability of an SJT to assess professional attributes during selection into UK Dental Core Training. A level of internal reliability close to that widely deemed acceptable[20] was found for this pilot test in line with similar findings from SJTs in other settings.[11] Additionally, partial correlations between scenarios and total SJT scores showed that 84.38% of scenarios had moderate or good associations with total SJT scores, indicating that the vast majority of SJT scenarios were high quality. Scenarios which were categorised as being of limited quality will be amended and re-piloted. Similarly, the difficulty of scenarios was within the range we would expect to see for a postgraduate SJT for use to 'sift out' low performing candidates.[5,16,17] That is; most candidates are scoring fairly well (but not perfectly) across the majority of scenarios. The distribution of the pilot SJT was slightly negatively skewed, which is a common finding among SJTs in the wider research literature.[23]

The group difference analyses indicate that the SJT is fair. No gender differences were found for the SJT and although White candidates performed better than BME candidates, the effect was marginal. This is important because ethnic group differences are common among selection methods, and the effect size found in this study was lower than that often reported in other settings.[24]

Candidates' reactions to the pilot SJT were very positive, indicating good levels of face validity. Over three quarters of candidates rated the SJT as relevant to dentistry, appropriate for the training level, suitably difficult and fair.

Implications for practice

The distribution of test scores showed that the pilot DCT SJT can differentiate between test takers. Practically, this is important because a selection method is only valuable if it allows selectors to differentiate between candidates on the basis of key attributes or skills relevant to the target role. Taken together with the findings of the difficulty ranges of the scenarios, this suggests that the SJT provides greatest value to selectors towards the bottom end of the distribution, as most candidates tend to perform relatively well on the test. Emerging evidence in the UK from the medical education literature suggests that SJTs may have greater predictive validity at the lower end of performers.[4,8] This suggests that in practice SJTs may be best used to 'select out' candidates in the earlier stages of a selection process and other, more academically-focused, methods may be more suited to 'select in' candidates during the later stages of selection.[25] This is in line with the intended use of the DCT SJT, which, if implemented, would be used to select out candidates who do not demonstrate the required non-academic attributes for DCT. It is likely that this will be followed by other selection tools, designed to target clinical and academic skills. While we note that this is a preliminary study, since we did not assess the predictive validity of the SJT, the psychometric evaluation does provide initial support for the use of the SJT to select out lower performing candidates.

Finally, given the minimal group differences found in this pilot study, we can tentatively conclude that the SJT has little adverse impact on candidates based on their group membership. This is important in a high-stakes selection context as in DCT.

Recommendations for future research

Future pilots should be completed to assess whether these findings can be replicated with other cohorts, where psychometric and group difference analyses could be conducted for other contexts. It would also be useful to further examine the model of 'selecting out' candidates scoring at the lower end of the distribution on important non-academic attributes in the early stages of selection in the dental context. Future studies should also examine the longitudinal predictive validity of the SJT with measures of performance at key stages in the postgraduate dental training pathway including during DCT and into qualified practice, and investigate whether relationships between the SJT and outcome measures show the nonlinear trend that may be expected if SJTs are better at predicting performance towards the lower end of the distribution.

1. Wyatt M R R, Pathak S B, Zibarras L. Advancing selection in an SME: Is best practice methodology applicable? *Int Small Bus J* 2010; **28**: 258–273.
2. Pangallo, A., Zibarras, L. & Patterson, F. Measuring resilience in palliative care workers using the situational judgement test methodology. *Med Educ*, 2016; **50**: 1131-1142.
3. Lievens F, Sackett P R P. Situational judgement tests in high-stakes settings: Issues and strategies with generating alternate forms. *J Appl Psychol* 2007; **92**: 1043–1055.
4. 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.
5. Patterson F, Ashworth V, Mehra S, Falcon H. Could situational judgement tests be used for selection into dental foundation training? *Br Dent J* 2012; **213**: 23–26.
6. Affleck P, Bowman M, Wardman M, Sinclair S, Adams R. Can we improve on situational judgement tests? *Br Dent J* 2016; **220**: 9–10.
7. Lievens F, Sackett P R. The validity of interpersonal skills assessment via situational judgement tests for predicting academic success and job performance. *J Appl Psychol* 2012; **97**: 460–468.
8. Cousans F, Patterson F, Edwards H, McLaghlan J, Good D. Evaluating the Complementary Roles of an SJT and Academic Assessment for Entry into Clinical Practice. *Adv Health Sci Educ Theory Pract* 2017; DOI: 10.1007/s10459-017-9755-4.
9. 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**: 734–741.
10. Patterson F, Carr V, Zibarras L et al. New machine-marked tests for selection into core medical training: evidence from two validation studies. *Clin Med* 2009; **9**: 417–420.
11. Nguyen N T, Mcdaniel M A. Response instructions and racial differences in a situational judgement test. *Pers Psychol* 2003; **8**: 33–44.
12. Nguyen N T, Biderman M D, McDaniel M. Effects of response instructions on faking a situational judgement test. *Int J Sel Assess* 2005; **13**: 250–260.
13. Whetzel D L, McDaniel M a. Situational judgement tests: An overview of current research. *Hum Resour Manag Rev* 2009; **19**: 188–202.
14. Patterson F, Zibarras L L, Carr V, Irish B, Gregory S. Evaluating candidate reactions to selection practices using organizational justice theory. *Med Educ* 2011; **45**: 289–297.
15. Everitt BS. *The Cambridge Dictionary of Statistics*. 2nd

edition. Cambridge: Cambridge University Press, 2002.

16. Patterson F, Murry H, Baron H, Aitkenhead A, Flaxman C. Analysis of the situational judgement test for selection to the foundation programme 2014. Technical Report. 2014.

17. Patterson F, Edwards H, Aitkenhead A, Flaxman C, Shaw R, Rosselli A. Analysis of the situational judgement test for selection to the foundation programme 2015. Technical Report. 2015.

18. Gravetter F, Wallnau L. Essentials of statistics for the behavioural sciences. 8th edition. Belmont, CA: Wadsworth, 2014.

19. McDaniel M A, Nguyen N T N T, Cabrera M A M, Nguyen N T N T. Situational judgement tests: A review of practice and constructs assessed. *Int J Sel Assess* 2001; **9:** 103–113.

20. Kline P. *The handbook of psychological testing*. London: Routledge, 2000.

21. Cohen J. *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Erlbaum Associates, 1988.

22. Patterson F, Baron H, Carr V, Plint S, Lane P. Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. *Med Educ* 2009; **43:** 50–57.

23. St-Sauveur C, Girouard S, Goyette V. Use of Situational Judgement Tests in Personnel Selection : Are the different methods for scoring the response options equivalent ? *Int J Sel Assess* 2014; DOI: 10.1111/ijsa.12072

24. Woolf K, Potts H W W, McManus I C. Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis. *BMJ* 2011; **342:** 1–14.

25. Patterson F, Knight A, Dowell J, Nicholson S, Cousans F, Cleland J. How effective are selection methods in medical eduction and training? Evidence from a systematic review. *Med Educ* 2016; **50:** 36–60.