Work Psychology Group
Thinking differently

UKCAT Situational Judgement Test 2014

Abridged Technical Report

February 2015

Professor Fiona Patterson
Vicki Ashworth
Stuart Martin
Helen Baron
Karen Fung
Rachael Shaw

Background

The UK clinical aptitude test (UKCAT) is used in the selection process by a consortium of UK university medical and dental schools. The test helps universities to make more informed choices from amongst the many highly qualified candidates who apply for the medical and dental degree programmes.

A situational judgement test (SJT) was piloted in 2012 and introduced live in 2013 to evaluate important non-academic attributes as part of the UKCAT. The purpose of the SJT is to enable the UKCAT to assess a broader range of constructs outside those relating to cognitive ability. The use of SJTs in medical and dental selection is widening. Following role analyses of numerous medical specialities, it is becoming widely acknowledged that non-academic or professional attributes are essential requirements for a doctor or a dentist.

SJTs are designed to assess individuals' judgement regarding situations encountered in a target role. Candidates are presented with a set of hypothetical but relevant scenarios and asked to make judgements about possible responses. Candidates' responses are evaluated against a pre-determined scoring key to provide a picture of their situational judgement in that particular context. SJT scenarios are based on extensive analysis of the target role, to ensure that test content reflects the most important situations in which to evaluate candidates' judgement and are concerned with testing attitudes and ethical values rather than knowledge or clinical skills.

SJTs have become increasingly popular over the last 20 years and are used mostly in large-scale selection processes, often at shortlisting, but can also form part of workplace assessment to highlight development needs. In the UK, SJTs are used nationally to select GP's, Foundation doctors and in other high stakes selection. Research literature indicates that SJTs have significant validity in predicting job performance and can offer substantial incremental validity (added-value) over methods such as ability and personality tests.

Design of the SJT

Six parallel test forms were constructed that each consisted of 20 scenarios (within 67 items nested within them). 19 of these were live 'operational' scenarios (62 items), and one was a pre-test pilot scenario (5 items). Each operational form consisted of 9 'appropriateness' scenarios and 10 'importance' scenarios. The item order and items within each form were both set, although candidates were allocated to forms at random, unless they had requested extra time. The six test forms had different items with the maximum possible score ranging between 235 and 243. Thus scores on different forms were equated onto a single scale, based on a classical pre-equating method.

Candidate performance

23,884 applicants sat the UKCAT SJT in 2014. Raw scores were scaled to a common reporting scale. Table 1 below shows the results of candidates' scaled scores. The minimum and maximum possible

scaled scores were 300 and 900 respectively. Please note that the scaled scores presented throughout this report are pre-equated to a single scale and thus are comparable across different forms.

Table 1: SJT Total Scale Score Summary Statistics

N Candidates	Mean	SD	Min	Max
23,884	597.54	80.80	300	793

Test analysis

Table 2 below provides the descriptive statistics using the raw score for the six forms separately.

Table 2: Raw Score Test Statistics for the SJT

Form	N Items	N Candidates	Mean	SD	Min	Max	Max Possible Score	Total Facility (mean as % of max poss score)	Alpha	SEM
1	62	4,690	178.86	19.03	88	221	236	75.79	.806	8.38
2	62	3,905	184.57	21.56	33	232	240	76.90	.852	8.29
3	62	3,817	186.24	22.30	65	238	243	76.64	.859	8.37
4	62	3,845	184.70	19.90	43	225	238	77.61	.838	8.01
5	62	3,856	184.15	20.88	0	222	235	78.36	.856	7.92
6	62	3,771	178.51	20.43	64	222	236	75.64	.833	8.35

The standard deviations range between 19.03 and 22.30. The standard deviation indicates how much variation there is from the mean. A low SD indicates that the data points tend to be very close to the mean, whereas a higher SD indicates that the data are spread out over a large range of values. In terms of scoring distribution, the scores range from 0 to 238. This represents a good spread of scores, and indicates that the SJT is able to differentiate effectively between applicants. A highly desirable degree of internal consistency reliability is demonstrated for each form with a Cronbach's alpha of above .80 for each form.

Table 3 below provides the descriptive statistics using the scaled score for the six forms separately. The equated scaled mean scores are used because raw means are affected by minor differences in difficulty of test forms, so should not be compared directly. Scale means take account of calibration of tests onto the same scale so differences reflect sample differences only.

Table 3: Scale Score Descriptives and Standard Error of Measurement for the SJT

Form	N Items	N Candidates	Mean	SD	Min	Max	SEM
1	62	4,690	592.52	78.37	300	767	34.5
2	62	3,905	598.51	87.22	300	793	33.6
3	62	3,817	595.00	83.16	300	790	31.2
4	62	3,845	594.00	79.20	300	756	31.9
5	62	3,856	608.42	80.58	300	756	30.6
6	62	3,771	597.86	75.05	300	759	30.7

The scale score means are broadly similar across the forms. The highest is for Form 5 (608.42) and the lowest is for Form 1 (592.52). Table 3 shows the SD's per form are also broadly similar, and the greatest SD difference is between Forms 2 (87.22) and 6 (75.05).

The standard error of measurement (SEM) is a range around the mean that estimates the error when interpreting an individual's test score and tells us the range within which a person's 'true' score may fall. As such, the SEM provides some guidance with respect to the importance placed on mean score differences (e.g. differences in mean score of less than 1 SEM between forms should not be regarded as meaningfully different). Table 5 shows that the largest scale score mean difference is between Forms 1 and 5, with a difference of 15.90. This is approximately half of an SEM so this is well within the acceptable range.