UK Clinical Aptitude Test (UKCAT) Consortium
UKCAT Examination

Technical Report
Testing Interval: 2 July 2018 – 2 October 2018

Prepared by:
Pearson VUE
31st January 2019

Non-disclosure and Confidentiality Notice

This document contains confidential information concerning Pearson's services, products, data security procedures, data storage parameters, and data retrieval processes. You are permitted to view and retain this document provided that you disclose no part of the information contained herein to any outside agent or employee, except those agents and employees directly charged with reviewing this information. These agents and employees should be instructed and agree not to disclose this information for any purposes beyond the terms stipulated in the agreement of your company or agency with Pearson.

Table of Contents

List of Tables

Executive Summary

The UK Clinical Aptitude Test (UKCAT) was administered in 2018 from 2 July to 2 October. During this period, a total of 27,469 exams were administered. Each exam consisted of four scored cognitive subtests: Verbal Reasoning (VR), Quantitative Reasoning (QR), Abstract Reasoning (AR), and Decision Making (DM). The cognitive subtests were followed by a Situational Judgement Test (SJT).

Each exam was composed of 164 items on the cognitive subtests of which 148 were operational items and 16 were pretest items. In addition, there were 69 SJT items of which 63 were operational and 6 were pretest. The exam was administered via computer in a 120-minute time period including administration of instructions. Each of the five sections was timed separately. There were four groups of candidates who received time accommodation in 2018. Candidates with special educational needs (SEN) were allotted 150 minutes (UKCATSEN) or 180 minutes (UKCATSEN50) based on UKCAT's pre-approval, and candidates with special accommodation (UKCATSA) were allotted 5 minutes for each display screen. Candidates who had special educational needs and who qualified for special accommodation (UKCATSENSA) were allotted 150 minutes for the exam in addition to 5 minutes display screen time. Results were provided to the candidates at the conclusion of testing and then later to schools to which the candidates had applied.

Overall candidate performance was broadly consistent with previous years. The average VR score was slightly lower by 3 points, and the average AR score was slightly higher by 8 points. There were falls in average scores in both QR and DM by 37 and 23 points respectively. These were due to differences in scaling method of QR and the benchmark population for DM, and they were in line with expectations. The 2018 SJT band distribution is broadly similar to that observed in 2017.

The performance patterns for different subgroups (e.g., ethnicity, gender, age, and National Statistics Socio-Economic Classification [NS-SEC]) closely paralleled those of the previous years. Males performed somewhat better than females on all cognitive subtests. Female candidates outperformed male candidates on the SJT, as observed in previous years. Ethnic-group performance trends also closely paralleled those of 2017.

In terms of candidate performance by social-economic group (UK candidates only; based on parental profession), Category 1 (Managerial and Professional Occupations) was consistently associated with higher mean scaled scores in the cognitive subtests. The lowest average cognitive subtest scaled scores occurred for Category 5 (Semi-routine or Routine Occupations) for all subtests except QR, for which the lowest average score was achieved by Category 4 (Lower Supervisory and Technical Positions) candidates. The social-economic trends are similar to those of 2017 with the mean scaled score differences between Category 1 and Category 5 decreasing for VR, AR and DM. For QR, Class 5 candidates performed marginally better than Class 4. Historically the SJT shows a weaker correlation with social-economic group than the cognitive sections. In 2018, there was a small difference in SJT scaled score by socio-economic classification, although similarly to QR, Class 5 candidates slightly outperformed Class 4.

Candidate age was broken into five groups: ≤15, 16 to 19, 20 to 24, 25 to 34, and ≥35. Performance across various age groups was examined separately by the candidates'

highest educational qualification. For candidates with Honours degrees, the age group 20 to 24 showed the highest scores across all cognitive sections. For candidates with school-leaving qualifications (i.e., below Honours degrees), the age group 16 to 19 had the highest scores.

The report also includes the performance analysis by the candidates' first language (English vs. non-English for UK and non-UK residents). The results indicated that candidates who reported English as their first language performed significantly better on all cognitive sections than candidates who did not list English as their first language. This is consistent with the 2017 cohort. The SJT showed similar trends to the cognitive sections by first language.

Background

The UK Clinical Aptitude Test (UKCAT) Consortium was formed by various medical and dental schools of higher-education institutions in the United Kingdom. The purpose of the UKCAT examination is to help select and/or identify more accurately those individuals with the innate ability to develop professional skills and competencies required to be a good clinician. The test results are to be used by institutions of higher education as part of the process of determining which applicants are to be accepted into the courses for which they have applied. The test results are also used by the Consortium for research to improve educational services. The goals of the Consortium are to use the UKCAT to widen access for students who desire to study Medicine and Dentistry at university level and to admit those candidates who will become the very best doctors and dentists of the future.

The UKCAT examination was first administered in July 2006 through the Pearson VUE Test Delivery System in testing centres in the United Kingdom and other countries. The 2018 testing period began on 2 July and ended on 2 October. During this period, a total of 27,469 exams were administered. Three forms each of the Verbal Reasoning (VR), Quantitative Reasoning (QR), Abstract Reasoning (AR), Decision Making (DM), and Situational Judgement Test (SJT) subtests were used to generate three UKCAT forms (Table 1). Each candidate was randomly assigned one of the three operational (scored) versions of the cognitive tests and a set of pretest (unscored) items.

Table 1. Composition of the Three UKCAT Forms

UKCAT Form	Verbal Reasoning	Quantitative Reasoning	Abstract Reasoning	Decision Making	Situational Judgement
Form 1	VR1	QR1	AR1	DM1	SJT1
Form 2	VR2	QR2	AR2	DM2	SJT2
Form 3	VR3	QR3	AR3	DM3	SJT3

The cognitive test forms were developed from the operational items used in the 2006 to 2017 administrations and also from items that had been pretested during these years. The SJT items were developed from operational and pretest items used from 2013 to 2017. All items (operational and pretest) were analysed, and those with acceptable item statistics were saved as the active item bank.

Design of Exam

The UKCAT is an aptitude exam and is designed to measure innate cognitive abilities in addition to individuals' judgement regarding situations encountered in a target role. It is not an exam that measures student achievement and therefore it does not contain any curriculum or science content.

The 2018 exam contained one SJT subtest and four scored cognitive subtests: VR, QR, AR and DM. All sections contained both operational (scored) and pretest (unscored) items. Candidates were given 120 minutes to answer a total of 232 items from the five subtests. There were four groups of candidates with time accommodation in 2018. Candidates with special educational needs (SEN) were allotted 150 minutes (UKCATSEN) or 180 minutes (UKCATSEN50) based on UKCAT's pre-approval, and candidates with special accommodation (UKCATSA) were allotted 120 minutes for the entire exam with flexible breaks, or 180 minutes for the entire exam with flexible breaks (UKCATSENSA). The design of the exam is shown in Table 2.

Table 2. UKCAT Exam Design

Subtest	Scored Items	Unscored Items	Total Number of Items	Test Time
VR	10 testlets of 4 items	1 testlet of 4 items	44	21 minutes allowed on items and 1 minute for instruction
QR	8 testlets of 4 items	1 testlet of 4 items	36	24 minutes allowed on items and 1 minute for instruction
AR	10 testlets of 5 items	1 testlet of 5 items	55	13 minutes allowed on items and 1 minute for instruction
DM	1 testlet of 26 items	3 items	29	31 minutes allowed on items and 1 minute for instruction
SJT	20 testlets of 2 to 5 items	1 testlet of 5 items 1 testlet of 1 item	69	26 minutes allowed on items and 1 minute for instruction

Examination Results

Cognitive Subtests

Scaled scores are reported for each of the four cognitive subtests based on all scored items in each subtest. Cognitive subtest scaled scores range from 300 to 900. Universities receive subtest scaled scores plus a total score (a sum of the four subtest scores) with a range of 1,200 to 3,600. An IRT calibration model and IRT true-score equating methods were used to transform raw scores from each form into a common reporting scale.

Table 3 presents summary statistics for each of the cognitive subtests plus the total summed scaled score for the total group. There were 27,469 candidate scores collected during 2018 testing. The scaled score means vary across the four cognitive subtests.

Table 3 Cognitive Subtest and Total Scaled Score Summary Statistics: Total Group

Test	Total *N*	Mean	*SD*	Min	Max
VR	27,469	566.58	77.64	300	900
AR	27,469	636.79	88.06	300	900
QR	27,469	657.75	75.47	300	900
DM	27,469	624.13	81.1	300	900
Total	27,469	2485.25	255.65	1,200	3,550

Table 4 summarises the scaled score statistics for UKCAT non-SEN candidates and SEN candidates. SEN candidates were allocated additional time and outperformed non-SEN candidates in all four subtests.

Table 4 Cognitive Subtest and Total Scaled Score Summary Statistics: SEN vs. non-SEN

Exam	Test	Total *N*	Mean	*SD*	Min	Max
UKCAT	AR	26,298	635.66	87.87	300	900
	DM	26,298	623.18	81.12	300	900
	QR	26,298	656.99	75.36	300	900
	VR	26,298	565.21	77.12	300	900
	Total	26,298	2,481.05	255.08	1,200	3,550
UKCATSEN	AR	1,093	663.77	88.21	360	890
	DM	1,093	645.95	78.24	300	880
	QR	1,093	675.55	76.1	430	900
	VR	1,093	598.53	82.9	300	890
	Total	1,093	2,583.8	250.17	1,460	3,370

Situational Judgement Test

For the Situational Judgement Test candidates are awarded one of four bands to reflect their performance on the operational items in the SJT. The bands are determined using the scaled score calculated for each candidate, as shown in Table 5. In previous years, score band boundaries were based on a normal distribution of scores, however, the observed distribution is not fully normal. This, combined with a rise in standardised scores year-on-year led to a deviation of the proportion in each band, compared from the intended proportion. Therefore, for 2018, the SJT was rescaled based on the observed 2017 candidate distribution to adjust for candidate performance.

The scaled score, not issued to candidates, ranges from 300 to 900 and is designed to place proportions of candidates into each band based on the 2017 score distribution. A classical pre-equating model was used to transform the raw scores from each form onto a common reporting scale. As the psychometric model used for the SJT is different to that used for the cognitive subtests, the scores are not directly comparable.

Table 5. SJT Band Scaled Score Range and Description (Base in 2017)

Bands	Scaled Score Range	Intended Band Proportions	Narrative
Band 1	664–900	22%	Those in Band 1 demonstrated an excellent level of performance, showing similar judgement in most cases to the panel of experts.
Band 2	604–663	38%	Those in Band 2 demonstrated a good, solid level of performance, showing appropriate judgement frequently, with many responses matching model answers.
Band 3	522–603	30%	Those in Band 3 demonstrated a modest level of performance, with appropriate judgement shown for some questions and substantial differences from ideal responses for others.
Band 4	300–521	10%	The performance of those in Band 4 was low, with judgement tending to differ substantially from ideal responses in many cases.

Table 6 presents the number and percentage of candidates in each band for the 27,469. The proportions observed in the 2018 SJT are similar to the intended percentages. Bands 2 and 4 show the most notable divergence from the intended proportions which may be due to the 2018 candidate population having a different distribution of scores to 2017.

Table 6. SJT Band Distribution in 2018

SJT Band	Number of Candidates	Percentage of Candidates
Band 1	5,715	20.8 %
Band 2	9,369	34.1 %
Band 3	8,805	32.1 %
Band 4	3,580	13 %
Total	27,469	100 %

Table 7 summarises the percentage by band and the scaled score statistics for SEN and non-SEN candidates. As observed in the past two years, SEN candidates broadly outperformed non-SEN candidates on the SJT.

Table 7. SJT Percentage by Band and Summary Statistics for SEN and non-SEN Candidates

Exam	Total *N*	Percentage of Candidates				Scaled Score			
		Band 1	Band 2	Band 3	Band 4	Mean	*SD*	Min	Max
UKCAT	26,298	20.3%	34.1%	32.3%	13.3%	602.11	72.64	300	768
UKCATSEN	1,093	31.8%	32.3%	28%	7.9%	622.85	68.03	300	756

Examination Results by Demographic Variables

The differential patterns of group performance for these variables in 2018 were similar to those from 2006 to 2017. Please note that group differences may be attributed to various factors. Readers are advised to interpret any difference with caution.

For the purpose of the demographic analysis, SJT scaled score summary statistics are included to illustrate trends. As mentioned previously, these scores are not issued to candidates and are not directly comparable to the cognitive subtests scaled scores.

Gender

Table 8 presents scaled score summary statistics for males and females for each of the subtests. On average, males outperformed females on VR (mean difference of 6 scaled score points), QR (mean difference of 21 scaled score points), AR (mean difference of 7 scaled score points) and DM (mean difference of 7 scaled score points). Female candidates outperformed male candidates on the SJT, as observed in previous years.

Table 8. Subtest and Total Scaled Score Summary Statistics by Gender

Test	Gender	Total *N*	Total %	Mean	*SD*	Min	Max
Verbal Reasoning	Female	17,138	62	564.46	76.98	300	890
	Male	10,147	37	570.50	78.56	300	900
Quantitative Reasoning	Female	17,138	62	650.08	73.94	300	900
	Male	10,147	37	671.25	76.21	300	900
Abstract Reasoning	Female	17,138	62	634.36	86.78	300	900
	Male	10,147	37	641.25	89.96	300	900
Decision Making	Female	17,138	62	621.10	80.67	300	900
	Male	10,147	37	629.84	81.46	300	890
Total Cognitive Scaled Score	Female	17,138	62	2,470.00	253.93	1,230	3,480
	Male	10,147	37	2,512.83	256.12	1,200	3,550
Situational Judgement Test	Female	17,138	62	608.63	71.37	300	768
	Male	10,147	37	593.76	73.39	300	767

Ethnicity

Table 9 summarises the performance of the various ethnic groups on each of the four cognitive subtests. Only UK candidates are asked to provide an ethnic group. The categories have been collated as follows:

- UK–White: White
- UK–Asian: Asian Indian; Asian Pakistani; Asian Bangladeshi; Asian Other
- UK–Black: Black Caribbean; Black African; Black Other
- UK–Mixed Race: Mixed White and Black Caribbean; Mixed White and Black African; Mixed White and Asian; Other Mixed
- UK–Chinese: Asian - Chinese
- UK–Other: Other e.g. gypsy, traveller, or Irish traveller, or not specified

Table 9. Subtest and Total Scaled Score Summary Statistics by Ethnic Group

Test	Ethnic Group	Total *N*	Total %	Mean	*SD*	Min	Max
Verbal Reasoning	Non UK	5,161	19%	542.26	77.27	300	870
	UK - Asian	7,854	29%	553.72	69.44	300	890
	UK - Black	2,169	8%	542.14	68.66	300	870
	UK - Chinese	416	2%	581.3	80.42	350	870
	UK - Mixed Race	1,184	4%	581.77	77.91	320	890
	UK - Other	1,089	4%	534.63	69.85	300	760
	UK - White	9,417	34%	597.99	75.48	300	900
Quantitative Reasoning	Non UK	5,161	19%	642.54	82.25	300	900
	UK - Asian	7,854	29%	659.69	74.26	300	900
	UK - Black	2,169	8%	624.62	69.05	300	900
	UK - Chinese	416	2%	701.9	81.85	460	900
	UK - Mixed Race	1,184	4%	662.7	76.16	390	900
	UK - Other	1,089	4%	639.35	71.75	430	900
	UK - White	9,417	34%	672.13	68.85	300	900
Abstract Reasoning	Non UK	5,161	19%	615.74	88.93	300	900
	UK - Asian	7,854	29%	641.34	87.09	300	900
	UK - Black	2,169	8%	606.51	82.4	300	890
	UK - Chinese	416	2%	680.36	97.28	380	900
	UK - Mixed Race	1,184	4%	647.58	87.71	350	900
	UK - Other	1,089	4%	622.75	83.29	300	890
	UK - White	9,417	34%	650.3	85.11	300	900
Decision Making	Non UK	5,161	19%	608.92	85.38	300	890
	UK - Asian	7,854	29%	611.76	77.05	300	890
	UK - Black	2,169	8%	588.02	75.62	300	880
	UK - Chinese	416	2%	648.44	83.61	350	890
	UK - Mixed Race	1,184	4%	636.68	79.21	310	880
	UK - Other	1,089	4%	595.21	79.77	300	880
	UK - White	9,417	34%	652.28	73.81	300	900
Total Cognitive Scaled Score	Non UK	5,161	19%	2,409.45	268.99	1,200	3,310
	UK - Asian	7,854	29%	2,466.51	243.1	1,490	3,480
	UK - Black	2,169	8%	2,361.30	232.46	1,310	3,410
	UK - Chinese	416	2%	2,611.99	267.28	1,860	3,400
	UK - Mixed Race	1,184	4%	2,528.73	256.92	1,680	3,300
	UK - Other	1,089	4%	2,391.93	242.44	1,490	3,180
	UK - White	9,417	34%	2,572.69	229.56	1,200	3,550
Situational Judgement Test	Non UK	5,161	19%	567.19	83.91	300	760
	UK - Asian	7,854	29%	602.64	67.78	300	760
	UK - Black	2,169	8%	594.36	71.72	300	767
	UK - Chinese	416	2%	609.53	65.96	337	746
	UK - Mixed Race	1,184	4%	618.41	67.08	300	760

Test	Ethnic Group	Total *N*	Total %	Mean	*SD*	Min	Max
	UK - Other	1,089	4%	591.75	76.42	300	749
	UK - White	9,417	34%	624.83	60.07	300	768

For VR and DM, the highest-performing group was UK-White. For QR and AR, the highest-performing group was UK-Chinese. The ethnic group with the lowest scaled score was UK-Black for QR, AR and DM. For VR, the lowest scoring candidates were UK-Other. The differences between the highest and lowest average scores were all significant.

There was considerable variation among the mean SJT scaled scores by ethnic group. The highest-performing group for the SJT was UK-White with Non-UK candidates having the lowest mean scaled score. This is consistent with the results from 2017.

NS-SEC

Table 10 provides scaled score summary statistics for all UK candidates by NS-SEC class (occupation and employment status). For all subtests, the means generally trended downwards in order of the occupational classes, from Class 1 to Class 5, Class 1 had the highest mean and Class 5 had the lowest mean for all subtests, except QR and SJT where the lowest mean was Class 4.

Table 10. Subtest and Total Scaled Score Summary Statistics by NS-SEC Class for UK Candidates

Test	NS-SEC Group	Total *N*	Total %	Mean	*SD*	Min	Max
Verbal Reasoning	1	14,328	65%	582.59	76.62	300	900
	2	1,065	5%	577.68	71.79	370	890
	3	1,316	6%	556.03	71.94	300	890
	4	687	3%	549.34	70.1	320	820
	5	1,456	7%	545.78	66.71	340	870
	NA	3,277	15%	550.06	74.84	300	890
Quantitative Reasoning	1	14,328	65%	669.48	73.45	300	900
	2	1,065	5%	656.83	72.24	330	900
	3	1,316	6%	650.45	67.4	390	900
	4	687	3%	643.84	70.65	300	880
	5	1,456	7%	644.96	68.93	430	900
	NA	3,277	15%	643.62	71.77	300	900
Abstract Reasoning	1	14,328	65%	650.66	87.77	300	900
	2	1,065	5%	631.19	84.05	300	900
	3	1,316	6%	628.72	79.2	380	900
	4	687	3%	623.61	80.35	380	890
	5	1,456	7%	623.44	82.95	300	900
	NA	3,277	15%	624.33	85.11	300	900
Decision Making	1	14,328	65%	639.67	78.13	300	900
	2	1,065	5%	627.05	74.58	300	880
	3	1,316	6%	608.56	74.1	310	880
	4	687	3%	604.45	80.09	310	870
	5	1,456	7%	598.83	72.03	310	880
	NA	3,277	15%	602.18	80.63	300	880

Test	NS-SEC Group	Total *N*	Total %	Mean	*SD*	Min	Max
Total Cognitive Scaled Score	1	14,328	65%	2,542.40	245.65	1,490	3,550
	2	1,065	5%	2,492.75	229.05	1,680	3,310
	3	1,316	6%	2,443.76	225.83	1,600	3,220
	4	687	3%	2,421.25	233.76	1,610	3,270
	5	1,456	7%	2,413.01	222.19	1,660	3,350
	NA	3,277	15%	2,420.19	250.42	1,200	3,410
Situational Judgement Test	1	14,328	65%	617.21	64.39	300	767
	2	1,065	5%	616.11	63.12	306	756
	3	1,316	6%	602.89	67.77	300	768
	4	687	3%	597.24	70.08	365	755
	5	1,456	7%	600.75	66.88	300	741
	NA	3,277	15%	597.67	72.07	300	753

Note. Codes for NS-SEC Groups
1 – Managerial and Professional Occupations
2 – Intermediate Occupations
3 – Small Employers and Own Account Workers
4 – Lower Supervisory and Technical Occupations
5 – Semi-routine and Routine Occupations
NA – Could not calculate SEC group i.e. information withheld

Age and Education

Table 11 provides scaled score summary statistics for the total group both by age group and the candidates' highest educational qualification.

Table 11. Subtest and Total Scaled Score Summary Statistics by Age Group and Highest Qualification

Test	Highest Qualification	Age Group	Total *N*	% Total *N*	Mean	*SD*	Min	Max
Verbal Reasoning	Below Honours degree level	Up to 15	32	0%	544.06	84.85	300	720
		16-19	19,981	97%	567.02	75.29	300	900
		20-24	524	3%	536.32	81.54	320	820
		25-34	126	1%	523.25	89.97	300	790
		>=35	36	0%	503.06	98.56	300	760
	Honours degree level or above	Up to 15	NA[a]	NA	NA	NA	NA	NA
		16-19	624	10%	534.81	73.54	300	760
		20-24	4,269	68%	578.24	78.25	300	890
		25-34	1,238	20%	573.09	90.63	300	890
		>=35	187	3%	539.73	94.83	300	870
Quantitative Reasoning	Below Honours degree level	Up to 15	32	0%	610.94	64.02	460	780
		16-19	19,981	97%	663.4	74.48	330	900
		20-24	524	3%	626.32	81.1	300	900
		25-34	126	1%	586.35	80.64	300	800
		>=35	36	0%	571.67	81.12	330	740
	Honours degree level or above	Up to 15	NA[a]	NA	NA	NA	NA	NA
		16-19	624	10%	639.5	74.35	390	900
		20-24	4,269	68%	655.21	71.48	300	900
		25-34	1,238	20%	633	75.66	330	900

Test	Highest Qualification	Age Group	Total *N*	% Total *N*	Mean	*SD*	Min	Max
		>=35	187	3%	586.79	78.06	300	870
Abstract Reasoning	Below Honours degree level	Up to 15	32	0%	575	81.16	300	740
		16-19	19,981	97%	640.69	86.39	300	900
		20-24	524	3%	609.52	94.79	300	890
		25-34	126	1%	574.21	100.84	300	890
		>=35	36	0%	548.06	98.41	320	710
	Honours degree level or above	Up to 15	NA[a]	NA	NA	NA	NA	NA
		16-19	624	10%	627.5	88.65	300	890
		20-24	4,269	68%	639.92	86.94	300	900
		25-34	1,238	20%	615.39	92.99	300	890
		>=35	187	3%	558.98	94.89	300	870
Decision Making	Below Honours degree level	Up to 15	32	0%	590	68.53	450	760
		16-19	19,981	97%	629.77	79.16	300	900
		20-24	524	3%	589.87	85.18	300	840
		25-34	126	1%	547.7	83.58	300	740
		>=35	36	0%	529.44	114.32	310	800
	Honours degree level or above	Up to 15	NA[a]	NA	NA	NA	NA	NA
		16-19	624	10%	593.89	81.83	310	830
		20-24	4,269	68%	623.89	77.9	300	880
		25-34	1,238	20%	601.25	87.96	300	880
		>=35	187	3%	550.27	91.28	300	870
Total Cognitive Scaled Score	Below Honours degree level	Up to 15	32	0%	2,320	247.96	1,510	2,800
		16-19	19,981	97%	2,500.88	247.98	1,330	3,550
		20-24	524	3%	2,362.02	279.13	1,560	3,150
		25-34	126	1%	2,231.51	292.45	1,200	2,980
		>=35	36	0%	2,152.22	310.7	1,490	2,760
	Honours degree level or above	Up to 15	NA[a]	NA	NA	NA	NA	NA
		16-19	624	10%	2,395.71	254.18	1,510	3,130
		20-24	4,269	68%	2,497.26	246.03	1,200	3,480
		25-34	1,238	20%	2,422.73	285.36	1,400	3,320
		>=35	187	3%	2,235.78	300.31	1,460	3,410
Situational Judgement Test	Below Honours degree level	Up to 15	32	0%	544.78	78.92	388	699
		16-19	19,981	97%	600.81	70.09	300	768
		20-24	524	3%	580.24	87.84	300	751
		25-34	126	1%	568.36	95.58	300	724
		>=35	36	0%	563.11	107.58	300	707
	Honours degree level or above	Up to 15	NA[a]	NA	NA	NA	NA	NA
		16-19	624	10%	573.03	81.93	300	762
		20-24	4,269	68%	624.19	66.15	300	767
		25-34	1,238	20%	619.12	77.74	300	756
		>=35	187	3%	581.92	98.19	300	758

[a]There was only 1 candidate in the Above Honours degree level and Up to 15 age group. For confidentiality, these scores are not reported.

Candidates were divided into five age groups: ≤15, 16 to 19, 20 to 24, 25 to 34, and ≥35. Two categories of educational qualification were examined: Below Honours Degree level and Honours Degree level or above. Candidates in the Honours Degree level or above category were mostly in the 20 to 24 age group, which also represented the highest mean

scores across all four cognitive subtests. Candidates in the Below Honours Degree level category were mostly in the 16 to 19 age group, which showed the highest mean scores across all four cognitive subtests.

Similar to cognitive subtests, the Below Honours Degree level had the highest mean SJT scaled scores at ages 16 to 19 and the Honours Degree level or above category had the highest mean SJT score at ages 20 to 24. These trends are consistent with those observed in 2017.

First Language

Scaled score analysis by the candidates' first language (English vs. Other for UK and non-UK candidates) is presented in Table 12. Candidates whose first language is English performed better on all four cognitive sections compared to candidates whose first language is not English for both UK and non-UK candidates. UK candidates generally outperformed non-UK candidates, with the exception of AR, where non-UK candidates whose first language was not English did slightly better than UK candidates whose first language was not English.

Table 12. Subtest and Total Scaled Score Summary Statistics by Country of Residence and First Language

Test	Country of Residence	First Language	Total *N*	% Total *N*	Mean	*SD*	Min	Max
Verbal Reasoning	UK	English	16,076	59%	585.58	74.89	300	900
		Other	6,019	22%	537.72	69.51	300	890
	Non-UK	English	2,091	8%	568.47	75.17	300	870
		Other	3,061	11%	524.46	73.43	300	820
Quantitative Reasoning	UK	English	16,076	59%	648.39	86.41	300	900
		Other	6,019	22%	624.45	86.19	300	900
	Non-UK	English	2,091	8%	620.90	85.90	300	890
		Other	3,061	11%	612.47	90.89	300	900
Abstract Reasoning	UK	English	16,076	59%	640.41	76.28	300	900
		Other	6,019	22%	594.46	78.52	300	890
	Non-UK	English	2,091	8%	627.37	81.90	300	890
		Other	3,061	11%	596.40	85.21	300	880
Decision Making	UK	English	16,076	59%	619.42	62.15	300	768
		Other	6,019	22%	591.17	73.29	300	760
	Non-UK	English	2,091	8%	586.97	73.10	300	760
		Other	3,061	11%	553.99	87.92	300	741
Total Cognitive Scaled Score	UK	English	16,076	59%	2,543.26	237.81	1,200	3,550
		Other	6,019	22%	2,398.42	246.31	1,310	3,400
	Non-UK	English	2,091	8%	2,472.82	256.19	1,570	3,260
		Other	3,061	11%	2,366.81	268.79	1,200	3,310
Situational Judgement Test	UK	English	16,076	59%	619.44	62.11	300	768
		Other	6,019	22%	591.04	73.38	300	760
	Non-UK	English	2,091	8%	586.94	73.01	300	760
		Other	3,061	11%	553.72	88.09	300	741

Test and Item Analysis

Test analysis for the operational forms included computation of the raw score means, standard deviations, internal consistency reliabilities, and standard errors of measurement for each form of each cognitive subtest. Similar test analyses were performed and reported for the scaled scores for the cognitive subtests. For the SJT, although scaled scores are not issued to candidates, they are used to determine bands and summary statistics are presented for reference.

Item analysis for the cognitive subtests included a complete classical analysis of item characteristics including *p* values and point biserial (item discrimination). IRT analyses included estimation of item difficulty, or *b*, parameter, and differential item functioning.

SJT item responses are graded using a partial credit model; candidates being awarded different marks depending on their response. Furthermore, the maximum score available varies by items depending on the key with some items having available score points of 0, 1, 3, 4 and others using score points of 0, 1, 2, 3.

Test Analysis Cognitive Subtests

The raw score means, standard deviations, ranges, internal consistency reliabilities (Cronbach's alpha), and standard errors of measurement for each form of each subtest are summarised in Table 13. The mean raw score differences across forms were within one point for all four cognitive subtests. Variation in score reliability and *SEM* across the cognitive subtests can be attributed to test length, amount of information in item types, scoring methods applied, and range of discrimination and difficulty amongst the items.

The highest raw score reliabilities were found for AR. This can be attributed to the test length as AR has the largest number of items; generally, reliability increases with test length. The score reliability pattern in 2018 is similar to the 2017 exam. All reliability indices are in line with expectations for comparable tests of this type and length.

Table 13. Raw Score Test Statistics

Test	Form	*N* Items	*N* Candidates	Mean	*SD*	Min	Max	Alpha	*SEM*
VR	1	40	10,058	22.14	6.04	2	39	0.77	2.90
	2	40	8,758	21.25	5.91	1	39	0.76	2.90
	3	40	8,653	22.18	5.96	0	40	0.77	2.86
QR	1	32	10,058	18.31	5.65	0	32	0.80	2.53
	2	32	8,758	17.86	5.48	0	32	0.78	2.57
	3	32	8,653	18.52	5.75	1	32	0.81	2.51
AR	1	50	10,058	30.88	7.69	0	50	0.84	3.08
	2	50	8,758	30.10	7.72	0	50	0.83	3.18
	3	50	8,653	30.86	7.80	0	50	0.84	3.12
DM	1	26	10,058	19.55	4.98	0	34	0.70	2.73
	2	26	8,758	18.63	5.31	2	34	0.72	2.81
	3	26	8,653	19.02	5.07	2	32	0.72	2.68

Candidates receive a scaled score for each cognitive subtest (Table 14). Unlike raw score reliability (in which the reliability index (Cronbach's alpha) was generated based on the intercorrelations or internal consistency among the items) the overall reliability of the scaled scores depends on the conditional reliability at each scaled score point instead of on item scores. For this reason, the two reliability indices (Cronbach's alpha and marginal reliability of scaled scores) are not directly comparable. Scaled score reliabilities are in line with expectations given test lengths for VR, QR, AR and DM.

Table 14. Scaled Score Reliability and Standard Error of Measurement for Cognitive Subtests

Tests	Form	*N* Items	*N* Candidates	Mean	*SD*	Min	Max	Scaled Score Reliability	*SEM*
VR	1	40	10,058	570.26	78.88	300	890	0.75	39.44
	2	40	8,758	559.06	75.99	300	890	0.74	38.75
	3	40	8,653	569.92	77.29	300	900	0.74	39.41
QR	1	32	10,058	658.38	75.13	300	900	0.77	36.03
	2	32	8,758	652.80	72.8	300	900	0.76	35.66
	3	32	8,653	662.03	78.21	330	900	0.78	36.68
AR	1	50	10,058	639.36	87.47	300	900	0.81	38.13
	2	50	8,758	629.79	86.19	300	900	0.81	37.57
	3	50	8,653	640.87	90.17	300	900	0.81	39.30
DM	1	26	10,058	628.34	78.96	300	900	0.69	43.96
	2	26	8,758	617.17	81.7	300	900	0.72	43.23
	3	26	8,653	626.27	82.5	300	880	0.71	44.43

Table 15 contains ranges and means of reliabilities and standard errors for the total scaled score. These values were computed as a composite function of the standard errors and reliabilities of the cognitive test forms contributing to the total. Each total scaled score is a simple sum (linear composite) of the forms of the cognitive tests that were administered to a given candidate. There were three combinations of cognitive test forms and therefore there were three estimates of total scaled score reliability and standard error. Each form had a reliability of 0.90, therefore the average reliability for total scaled score was 0.90, reflecting good overall reliability. The average standard error was 95.80, which is very reasonable for the range of total scaled score. The average reliability were similar to those observed in 2017 (0.89), the average *SEM* was somewhat higher than 2017 (81.71).

Table 15. Scaled Score Reliability and Standard Error of Measurement for Total Score

Reliability		*SEM*	
Range[a]	Mean	Range	Mean
0.90	0.90	94.02-95.80	95.80

[a]Based on three combinations of cognitive test forms.

Score reliabilities of the four cognitive subtests in the 2018 UKCAT ranged from moderate to high. Reliability for the total score was good. Variation in score reliability across the four subtests can be partially attributed to the length of subtests. Improvement of score reliability can be attributed to a stronger item bank. A strong item bank provides higher flexibility in selecting better-fitted (more discriminative and reasonably challenging) items.

Item Analysis Cognitive Subtests

Since 2007, the item development and pretesting plan for the cognitive tests has been implemented in order to strengthen the UKCAT item pool. Improvement of the active item pool is achieved through rounds of item writing, pretesting, data analysis and statistical screening. Each year, new items are developed through item-writing workshops. These newly developed items are then pretested with operational items. At the end of each testing window, both operational and pretest items are analysed. The purpose of item analysis is to examine item quality and determine suitability of items for future use.

Test Analysis Situational Judgment Test

The raw score means, standard deviations, ranges, internal consistency reliabilities and *SEM* for each form of the SJT are summarised in Table 16. The test statistics are computed based on all candidates who took the SJT. The maximum number of available score points is 232 for all the three forms in 2018, however, it has varied in previous years. Therefore, the mean raw score as a percentage of the maximum available score is used to compare the raw score. The reasonably high percent correct and skewed scaled score distribution indicates that the SJT is capable of identifying the weakest candidates. Currently the SJT does not differentiate as well between the more able candidates. If the SJT were to be used to differentiate across more of the scale rather than to screen out candidates at the lower end, the difficulty of the test would need to be increased further.

The reliabilities for all SJT forms are good and comparable to 2017. As expected, the increase in the difficulty of the forms has not impacted on the reliability of the SJT. The *SEM* was based on the raw score metric and ranged from 8.45 to 8.86.

Table 16. SJT Raw Score Test Statistics (all candidates)

Form	*N* Items	*N* Candidates	Mean	*SD*	Min	Max	Mean Percent Raw Score	Alpha	*SEM*
1	63	10,058	166.32	21.12	0	212	72%	0.84	8.45
2	63	8,758	164.37	21.35	0	212	71%	0.84	8.54
3	63	8,653	164.51	19.82	0	211	71%	0.80	8.86

Bands are based on SJT scaled scores. Test statistics for scaled scores are provided in Table 17. The scaled scores are linearly related to the raw scores and therefore the raw score reliability applies equally to the scaled scores. This is in contrast to the cognitive tests where the scaled scores are a transformation of the IRT ability values. Differences in average scaled score between the forms are small and within one *SEM*.

Table 17. SJT Scaled Score Test Statistics (all candidates)

Form	*N* Items	*N* Candidates	Mean	*SD*	Min	Max	*SEM*
1	63	10,058	604.63	74.25	300	767	29.70
2	63	8,758	602.98	73.62	300	768	29.45
3	63	8,653	601.13	69.36	300	765	31.02

Item Analysis SJT

Each year, new SJT items are developed and reviewed. The SJT items are analysed using classical test theory. A review of the SJT following the 2013 test window showed that an IRT approach is not appropriate. Unlike IRT, classical test statistics are sample dependent, meaning that they are calculated based on the sample of candidates who respond to each item and are not linked back to a common benchmark group. Therefore, the item statistics presented for the SJT are not comparable to those of the cognitive sections due to the different measurement models used.

Differential Item Functioning

Introduction

Differential Item Functioning (DIF) refers to the potential for items to behave differently for different groups. DIF is generally an undesirable characteristic of an examination because it means that the test is measuring not only the construct it was designed to measure but also an additional characteristic or characteristics of performance that depend on classification or membership in a group, usually a gender or ethnic group classification. For instance, if female and male candidates of the same ability level perform very differently on an item, then the item may be measuring something other than the ability of the candidates, possibly some aspect of the candidates that is related to gender. The principles of test fairness require that examinations undergo scrutiny to detect and remove items that behave in significantly different ways for different groups based solely on these types of demographic characteristics. In DIF, the terms “reference group” and “focal group” are used for group comparisons and generally refer to the *majority* and the *minority* demographic groupings of the exam population, respectively.

Detection of DIF

There are a number of procedures that can be used to detect DIF. One of the most frequently used is the Mantel-Haenszel procedure (Zwick, Thayer, Lewis, 1999). The Mantel-Haenszel procedure compares reference and focal group performance for candidates within the same ability strata. If there are overall differences between reference group and focal group performance for candidates of the same ability levels, then the item may not be fitting the psychometric model and may be measuring something other than what it was designed to measure.

The Mantel-Haenszel procedure requires a criterion of proficiency or ability that can be used to match (group) candidates to various levels of ability. For the UKCAT examination, matching is carried out using the raw score on each subtest associated with the item under study.

Items were classified for DIF using the Mantel-Haenszel delta statistic. This DIF statistic (hereafter known as MH D-DIF) is expressed as *differences* on the delta scale, which is commonly used to indicate the difficulty of test items. For example, an MH D-DIF value of 1.00 means that one of the two groups being analysed found the question to be one delta point more difficult than did *comparable* members of the other group. (Except for extremely difficult or easy items, a difference of one delta point is approximately equal to a difference of 10 points in percent correct between groups.) The convention of having negative values of MH D-DIF reflect an item that is differentially more difficult for the focal group (generally, females or the ethnic minority group) has been adopted. Positive values of MH D-DIF indicate the item is differentially more difficult for the reference group (generally white or male candidates). Both positive and negative values of the DIF statistic are found and are taken into account by these procedures.

Criteria for Flagging Items

For the UKCAT examination, MH D-DIF items were classified into one of three categories: A, B, or C. Category A contains items with negligible DIF, Category B contains items with

slight to moderate DIF, and Category C contains items with moderate to large DIF. These categories are derived from the DIF classification categories developed by Educational Testing Service (ETS) and are defined below:

A: MH D-DIF is not significantly different from zero or has an absolute value < 1.0

B: MH D-DIF is significantly different from zero and has an absolute value >= 1.0 and < 1.5

C: MH-D-DIF is significantly larger than 1.0 and has an absolute value >= 1.5

Items flagged in Category C are typically subjected to further scrutiny. Items flagged in Categories A and B are not reviewed because of the minor statistical significance. The principal interpretation of Category C items is that—based on the present samples—items flagged in this category appear to be functioning differently for the reference and focal groups under comparison. If an item functions differently for two different groups, then content experts may (or may not) be able to determine from the item itself whether the item text contains language or content that may create a bias for the reference or focal group. Therefore, Category C flagging for DIF is necessary but not sufficient grounds for revision and possible removal of the item from the pools.

Comparison Groups for DIF Analysis

DIF analyses were conducted for the pretest and operational items when sample sizes were large enough. The UKCAT DIF comparison groups are based on gender, age, ethnicity and social-economic status. Age was separated into groups less than 20 years old and greater than 35 years old. There are 17 ethnic categories in the UKCAT database. For the DIF analyses, several of these categories were collapsed into meaningful, broader groups. The DIF ethnic categories used for these analyses (collapsed where indicated) were as follows:

White: White – British
Black: Black – Black/British – African, Black – Black/British – Caribbean, Black – Black/British Other
Asian: Asian – Asian/British – Bangladeshi, Asian – Asian/British – Indian, Asian – Asian/British – Other Asian, Asian – Asian/British – Pakistani.
Chinese: Asian – Asian/British – Chinese
Mixed: Mixed – Mixed – Other, Mixed – White/Asian, Mixed – White/Black African, Mixed – White/Black Caribbean
Other: Other ethnic group

Sample Size Requirements

Minimum sample-size requirements used for the UKCAT DIF analyses were at least 50 focal group candidate responses and at least 200 total (focal plus reference) candidate responses. Because pretest items were distributed across multiple versions of the forms, fewer responses are available per item than for operational items. As a result, it was not possible to compute DIF for many of the pretest items for some group comparisons (e.g., between White and Black, Asian, Chinese, and Mixed race).

DIF Results Cognitive Subtests

Table 18 (Cognitive test operational items), Table 19Table 19 (cognitive test pretest items),

Table 20 (operational items) and Table 21 (pretest items) in Appendix A show the number and percentages of items classified into each of the three DIF categories along with the quantities for which insufficient data were available to compute DIF (Category NA).

For the cognitive tests, in operational DIF analysis, comparisons between age groups did not meet sample size requirements to compute DIF.

For cognitive test pretest items, comparisons between age groups; ethnic groups; and NS-SEC Class 1 and the other four NS-SEC groups failed to meet the minimum sample size requirements. For SJT pretest items, comparisons between White and Black, White and Chinese, White and Mixed, and between NS-SEC Class 1 and Class 4 did not meet minimal sample size requirements. These items will be re-evaluated for DIF when they are used in future operational forms.

For the operational pools, there were 14 occurrences of Category C DIF across all cognitive subtests and comparisons. The proportion of Category C DIF out of all possible comparisons across the four cognitive tests was extremely low. Of these 14 occurrences, three occurred in the Age <20 / >35 comparison; six in the White/Black comparison; four in the White/Chinese comparison; and one in White/Asian comparison. For the pretest items, there was two occurrences of Category C DIF in the Male/Female comparison group. It should be noted that as pretest items are seen by fewer candidates, a significant number of comparisons could not be made due to low sample numbers in the focal groups.

For the operational SJT pool, there were no occurrences of Category C DIF and 32 instances of Category B DIF. For the pretest items, there were two occurrences of Category C DIF. It should be noted that as pretest items are seen by fewer candidates, a significant number of comparisons could not be made due to low sample numbers in the focal groups.

Taken together, the results indicated very little DIF occurrence in the UKCAT SJT items.

Appendix A. DIF Summary Tables

Table 18. DIF Classification: Cognitive Subtests, Operational Pool

Comparison Group	MH D-DIF Code	Verbal Reasoning		Quantitative Reasoning		Abstract Reasoning		Decision Making	
		N Items	%	*N* Items	%	*N* Items	%	*N* Items	%
Male/ Female	A	119	99%	96	100%	149	99%	77	99%
	B	1	1%	0	0%	1	1%	1	1%
	C	0	0%	0	0%	0	0%	0	0%
	NA	0	0%	0	0%	0	0%	0	0%
	Total	120	100%	96	100%	150	100%	78	100%
Age <20/>35	A	80	67%	64	67%	100	67%	48	62%
	B	0	0%	0	0%	0	0%	1	1%
	C	0	0%	0	0%	0	0%	3	4%
	NA	40	33%	32	33%	50	33%	26	33%
	Total	120	100%	96	100%	150	100%	78	100%
White/ Black	A	116	97%	94	98%	148	99%	73	94%
	B	3	3%	1	1%	2	1%	1	1%
	C	1	1%	1	1%	0	0%	4	5%
	NA	0	0%	0	0%	0	0%	0	0%
	Total	120	100%	96	100%	150	100%	78	100%
White/ Asian	A	116	97%	94	98%	150	100%	77	99%
	B	3	3%	2	2%	0	0%	1	1%
	C	1	1%	0	0%	0	0%	0	0%
	NA	0	0%	0	%	150	100%	77	99%
	Total	120	100%	96	100%	150	100%	78	100%
White/ Chinese	A	118	98%	96	100%	149	99%	76	97%
	B	0	0%	0	0%	0	0%	1	1%
	C	2	2%	0	0%	1	1%	1	1%
	NA	0	0%	0	0%	0	0%	0	0%
	Total	120	100%	96	100%	150	100%	78	100%
White/Mixed	A	120	100%	96	100%	150	100%	78	100%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	0	0%	0	0%	0	0%	0	0%
	Total	120	100%	96	100%	150	100%	78	100%
NS-SEC Class 1/2	A	120	100%	96	100%	150	100%	78	100%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	0	0%	0	0%	0	0%	0	0%
	Total	120	100%	96	100%	150	100%	78	100%
NS-SEC Class 1/3	A	120	100%	96	100%	150	100%	78	100%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	0	0%	0	0%	0	0%	0	0%
	Total	120	100%	96	100%	150	100%	78	100%
NS-SEC Class 1/4	A	120	100%	96	100%	150	100%	78	100%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	0	0%	0	0%	0	0%	0	0%
	Total	120	100%	96	100%	150	100%	78	100%
NS-SEC Class 1/5	A	119	99%	96	100%	150	100%	78	100%
	B	1	1%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	0	0%	0	0%	0	0%	0	0%
	Total	120	100%	96	100%	150	100%	78	100%

Table 19. DIF Classification: Cognitive Subtests, Pretest Pool

Comparison Group	MH D-DIF Code	Verbal Reasoning		Quantitative Reasoning		Abstract Reasoning		Decision Making	
		N Items	%	*N* Items	%	*N* Items	%	*N* Items	%
Male/ Female	A	230	99%	218	100%	296	100%	210	100%
	B	0	0%	0	0%	0	0%	0	0%
	C	2	1%	0	0%	0	0%	0	0%
	NA	0	0%	0	0%	0	0%	0	0%
	Total	232	100%	218	100%	296	100%	210	100%
Age <20/>35	A	0	0%	0	0%	0	0%	0	0%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	232	100%	218	100%	296	100%	210	100%
	Total	232	100%	218	100%	296	100%	210	100%
White/ Black	A	2	1%	1	0%	5	2%	0	0%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	230	99%	217	99%	291	98%	210	100%
	Total	232	100%	218	100%	296	100%	210	100%
White/ Asian	A	232	100%	217	100%	296	100%	205	98%
	B	0	0%	1	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	0	0%	0	0%	0	0%	5	2%
	Total	232	100%	218	100%	296	100%	210	100%
White/ Chinese	A	0	0%	0	0%	0	0%	0	0%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	232	100%	218	100%	296	100%	210	100%
	Total	232	100%	218	100%	296	100%	210	100%
White/ Mixed	A	0	0%	0	0%	0	0%	0	0%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	232	100%	218	100%	296	100%	210	100%
	Total	232	100%	218	100%	296	100%	210	100%
NS-SEC Class 1/2	A	0	0%	0	0%	0	0%	0	0%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	232	100%	218	100%	296	100%	210	100%
	Total	232	100%	218	100%	296	100%	210	100%
NS-SEC Class 1/3	A	0	0%	0	0%	0	0%	0	0%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	232	100%	218	100%	296	100%	210	100%
	Total	232	100%	218	100%	296	100%	210	100%
NS-SEC Class 1/4	A	0	0%	0	0%	0	0%	0	0%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	232	100%	218	100%	296	100%	210	100%
	Total	232	100%	218	100%	296	100%	210	100%
NS-SEC Class 1/5	A	0	0%	0	0%	0	0%	0	0%
	B	0	0%	0	0%	0	0%	0	0%
	C	0	0%	0	0%	0	0%	0	0%
	NA	232	100%	218	100%	296	100%	210	100%
	Total	232	100%	218	100%	296	100%	210	100%

Note. NA: Insufficient data to compute MH D-DIF

Table 20. DIF Classification: SJT, Operational Pool

Comparison Group	Degree of DIF					
	A		B		C	
	N Items	Percentage	*N* Items	Percentage	*N* Items	Percentage
Male/Female	147	97%	4	3%	0	0%
Age <20/>35	149	99%	2	1%	0	0%
White/Black	144	95%	7	5%	0	0%
White/Asian	134	89%	17	11%	0	0%
White/Chinese	149	99%	2	1%	0	0%
White/Mixed	151	100%	0	0%	0	0%
NS-SEC Class 1/2	151	100%	0	0%	0	0%
NS-SEC Class 1/3	151	100%	0	0%	0	0%
NS-SEC Class 1/4	151	100%	0	0%	0	0%
NS-SEC Class 1/5	151	100%	0	0%	0	0%

Table 21. DIF Classification: SJT, Pretest Pool

Comparison Group	Degree of DIF							
	A		B		C		N<200	
	N Items	Percentage	*N* Items	Percentage	*N* Items	Percentage	*N* Items	Percentage
Male/Female	190	95%	11	5%	0	0%	0	0%
Age <20/>35	188	94%	13	6%	0	0%	0	0%
White/Black	184	92%	7	3%	1	0%	9	4%
White/Asian	192	96%	9	4%	0	0%	0	0%
White/Chinese	169	84%	2	1%	1	0%	29	14%
White/Mixed	183	91%	4	2%	0	0%	14	7%
NS-SEC Class 1/2	201	100%	0	0%	0	0%	0	0%
NS-SEC Class 1/3	197	98%	4	2%	0	0%	0	0%
NS-SEC Class 1/4	198	99%	3	1%	0	0%	3	1%
NS-SEC Class 1/5	191	95%	10	5%	0	0%	0	0%