



SQE2 Pilot

June 2020

Table of Contents

1. Aims of the pilot	3
2. The SQE2 pilot	3
2.1 Structure of the pilot	3
2.2 Candidate Recruitment	4
2.3 Delivery of the pilot	6
3. Stakeholder Engagement and Feedback	6
4. Results of the pilot	7
4.1 Candidate Performance	7
4.2 Test quality	8
5. Analysis	9
5.1 Number of stations for adequate reliability and precision	9
5.2 Compensation between skills	9
5.3 Compensation between skills (advocacy)	10
5.4 Fairness: Equality Issues	12
5.5 Work Experience	15
5.6 Specialisms/common core/uniform exam	16
6. Recommendations	20
7. Appendices	21

1. Aims of the pilot

The aim of the SQE2 pilot was to review the SRA's 2017 draft Assessment Specification with a view to making recommendations for an SQE2 assessment which would be reliable, accurate, valid, cost effective and manageable.

In particular the 2017 draft Assessment Specification said answers were needed to the following questions to ensure this overarching aim was met:

What is the most valid, reliable and manageable way to assess Rights of Audience? For example, could Rights of Audience be assessed at stage 1?

What is the minimum number of separate assessments required in stage 2 to reliably, validly and cost-effectively assess the SoSC? For example, could fewer more synoptic assessments, covering a broader range of competences, be used at stage 2?

In what range of practice contexts should the stage 2 practical legal skills assessments be assessed? What is the impact of candidate choice and prior work experience on the reliability of assessments set in different practice areas?

What are the benefits and risks of retaining a non-compensatory standard setting model, as currently proposed, as opposed to a total compensatory or a partial compensatory (within clusters of competence) standard setting model?

2. The SQE2 pilot

2.1 Structure of the pilot

The practice areas tested in the SQE2 legal skills pilot were:

- Criminal Litigation (including advising clients at the police station)
- Dispute Resolution
- Property Practice
- Wills and Intestacy, Probate Administration and Practice
- Business organisations, rules and procedures.

Questions in these practice areas could draw on underlying black letter law in the Functioning Legal Knowledge (FLK) as follows:

- Criminal Litigation: Criminal liability
- Dispute Resolution: Contract law and tort
- Property Practice: Land law
- Wills and Intestacy, Probate Administration and Practice: Trusts
- Business organisations, rules and procedures: Contract law.

Questions on ethics were pervasive throughout the SQE2 pilot. Questions involving taxation could arise in Property Practice; Wills and Intestacy, Probate Administration and Practice; and Business organisations, rules and procedures.

The assessments in the SQE2 legal skills pilot were:

- client interview
- attendance note/legal analysis (linked to the client interview)
- oral presentation/advocacy
- case and matter analysis
- legal research
- legal writing
- legal drafting.

For the purposes of the pilot, candidates completed these seven assessments twice (i.e. they did a total of 14 assessments) as follows:

- Seven assessments in a practice area chosen by the candidate (the “specialism”). For the purposes of the pilot, candidates were asked to choose between (1) Business organisations rules, and procedures and (2) Criminal Litigation
- Seven assessments in practice contexts which ranged across the five practice areas above (the “common core”), (the same seven assessments were taken by all candidates irrespective of the specialism they had chosen).

These are shown pictorially in **Appendix 1**.

One of the principal reasons for this design was to use it to evaluate the extent to which SQE2 could incorporate specialisation while still performing its function of being a credentialing exam for a generic qualification with a universal standard. We were evaluating three options:

- a) The model the pilot followed of a universal part that everybody takes (the common core) and a test of skills in a choice of one out of five specialisms. The aim was to try to use the pilot to see if the common core could be used as a yardstick to measure the difficulty of the specialisms and so arrive at a universal standard even though candidates were not taking the same assessments.
- b) All candidates take the same exam (referred to as a uniform exam).
- c) Candidates select two areas of legal practice out of the five mentioned above and are examined in each of the skills, only in those practice areas.

In addition to the legal skills test outlined above, candidates were asked to answer 120 single best answer multiple choice questions. This was done to enrich the analysis of the legal skills data. There is no proposal to include single best answer multiple choice questions in SQE2.

2.2 Candidate Recruitment

The SRA publicised the opportunity for people to be involved in the SQE2 pilot through a range of communications channels, including social media and its website. It also directly engaged law firms, representative groups and training providers to highlight the opportunity to potential candidates. People could then apply on the SRA website. We wanted, as far as possible, a selection of candidates who would be broadly representative of those who would take the SQE.

In terms of education and prior experience we were therefore seeking candidates who either:

- had completed Stage 1 (the compulsory element) of the Legal Practice Course (LPC); or
- had completed a period of study and/or work experience equivalent to Stage 1 (the compulsory element) of the LPC; or
- were qualified lawyers in a recognised jurisdiction eligible to qualify via the Qualified Lawyers Transfer Scheme; or
- were barristers of England and Wales.

The application process resulted in 633 applications to take the pilot assessments. 541 candidates were invited to take part in the SQE2 pilot. 335 accepted their place. 103 of them cancelled in the run up to the examinations and there were 42 no-shows (who did not attend any day of the pilot), leaving 190 active participants. 23 of these did not sit the full pilot resulting in 167 full participants.

Attracting candidates to take a pilot assessment is always challenging (despite financial and other incentives), particularly where, as in this case, it involves attendance over 4 days. Numbers were not as high as for the SQE1 pilot. Nevertheless at 167 candidates the pilot was able to provide strong indicative evidence¹. Further, the results of the pilot were not interpreted in isolation. Analysis and interpretation was undertaken by, and in the light of, the experience of the Kaplan Advisory Board with world leading expertise in professional assessment for licensure including extensive experience of professional assessment for licensure in law². In addition, results of the pilot were interpreted in the light of comprehensive data, collected since 2011, from the Qualified Lawyers Transfer Scheme (QLTS), which has similar assessments to SQE2. This QLTS data covers over 5,000 legal skills test candidate attempts across 18 sittings. These elements, considered together, provide a solid basis for the recommendations made in this report.

Candidate demographics are set out in **Appendix 2**. A comparison with the most recently published LPC data on key demographics is informative:

- 66% of candidates identified as female compared with 34% male (LPC 2017/18, 64% female, 36% male)
- 25% of candidates identified as Asian/Asian British (LPC 2017/18, 24%)
- 19% of candidates identified as Black/Black British (LPC 2017/18, 11%)
- 5% of candidates identified as Mixed/Multiple ethnic groups (LPC 2017/8, 3%)
- 7% of candidates declared a disability within the meaning of the Equality Act 2010 (LPC 2017/18, 15% identified themselves as having a disability).

¹ These numbers are higher than for some pilots of similar national exams. For instance the pilot for the GPs' Clinical Skills Assessment (the equivalent of SQE2) was held in September 2006 and involved 98 participants. The mandatory exam for all GP trainees was implemented in 2007.

² The Advisory Board comprises Dr Susan Case, formerly Director of Testing, National Conference of Bar Examiners; [Dr David Swanson](#), American Board of Medical Specialities; [Richard Wakeford](#), Hughes Hall, University of Cambridge.

It is also noteworthy that there were significant disparities between the groups electing to sit the business and criminal specialities. The final columns in the Tables in **Appendix 2** show whether these differences were statistically significant or not.

In the group choosing the criminal speciality there was a significantly higher number of Muslim candidates; a significantly higher number who had more than three months' work experience in the speciality practice area; a significantly lower number who had completed the GDL; and a significantly lower number of LPC distinctions and more passes.

The fact that there were significant differences between the demographic make-up of the groups will be referred to later.

2.3 Delivery of the pilot

The SQE2 pilot was delivered as follows:

- 10 December – written tests for the common core
- 11 December – written tests for the chosen specialisation
- 13 December – oral tests for the criminal specialisation
- 14 December – oral tests for the common core
- 15 December – oral tests for the business specialisation

SQE2 pilot oral assessments were delivered at two centres in London, Kaplan Spring House and the Royal College of General Practitioners premises in Euston. SQE2 pilot written assessments were delivered at 29 UK Pearson VUE test centres.

A variety of reasonable adjustments were accommodated including additional time, individual/smaller testing rooms, enlarged font, screen magnifier and stop the clock rest breaks.

The pilot ran smoothly and there were no incidents requiring the use of the Local Escalation Plan.

In summary delivery of the oral assessments simultaneously at more than one venue, and delivery of the written assessments at Pearson VUE were both feasible and manageable. Some further automation is needed to systems for delivery efficiently at scale but there is ample time to test and document the further systems that are still required.

3. **Stakeholder engagement and feedback**

Stakeholder views fed into the framing of the key questions mentioned above and the approach taken by the pilot. In addition, Kaplan and the SRA have sought stakeholder views on key issues including:

- the extent of specialisation that should be allowed in SQE2;
- whether candidates should be required to pass each individual assessment station or whether compensation should be allowed;
- the assessment of advocacy.

Methods of stakeholder engagement included conferences, on-line surveys, meetings of the SQE reference group, posting to the SQE LinkedIn group, meetings with BAME representatives; round-table discussions/workshops with the profession, and engagement with candidates of the SQE2 pilot via an on-line survey. Further stakeholder engagement will be held on the post-pilot recommendations.

Feedback from stakeholders presented a complex picture. As regards the extent of specialisation which should be allowed in SQE2, stakeholders could see the advantages and disadvantages of each model. Some stakeholders, particularly academics but also some practitioners, accept that there is a clear regulatory justification for a uniform exam and that it could provide a more rigorous and consistent assessment of high professional standards. However, stakeholders could also see that a uniform assessment would not necessarily fit the profession’s current approach to training or the preference of some stakeholders for candidates to be assessed in practice areas that their firm practises in.

A key concern of stakeholders was that candidates should not be able to compensate for poor performance in some areas with higher scores in others. As regards assessment of advocacy, some stakeholders felt all candidates should do at least one civil and one criminal advocacy assessment. Others felt one assessment of advocacy in a courtroom context was sufficient.

4. Results of the pilot

4.1 Candidate Performance

Table 1: Candidate Performance on the Skills Tests (law and skills 50:50)							
Test and Candidate Group	N Candidates	N Stations	Mean %	SD %	Median %	Min (%)	Max (%)
Common Core	167	7	56.48	15.79	57.43	11.46	91.14
Common Core (Business Only)	97	7	58.80	15.31	59.38	23.77	91.14
Common Core (Criminal Only)	70	7	53.26	15.99	53.68	11.46	86.49
Business Specialty	97	7	56.39	16.59	58.71	18.98	94.82
Criminal Specialty	70	7	53.50	18.41	54.51	12.48	97.50
Core + Business	97	14	57.59	15.11	59.12	23.84	92.44
Core + Criminal	70	14	53.38	16.33	53.29	11.97	91.99

Table 1 sets out the overview candidate performance. The average score for all candidates on their 14 stations was 55.8%, with a range of 12% - 92%, standard deviation (SD) 15.1%. For the seven common core stations, the respective figures were 58.8%, 14% - 92%, and 15.8%.

Differences in performance between the specialist groups are summarised in Table 2.

Table 2: Business and Criminal Candidates' Performance Compared on the Skills Test (law and skills 50:50)						
Comparison	Business		Criminal		Significance of difference	Significant difference favours?
	Mean %	SD %	Mean %	SD %		
7 Common Core Stations	58.80	15.31	53.26	15.99	F = 5.140 p = 0.025	Business
7 Specialist Stations	56.39	16.59	53.50	18.41	NS	-
All 14 Stations	57.59	15.11	53.38	16.33	NS	-

It should be noted that while the difference in performance between the two groups of candidates on the specialism is not statistically significant, it is statistically significant on the common core. This will be referred to later.

As with SQE1 it was not considered appropriate to set a pass mark for the pilot. However we did estimate the pass mark using the borderline regression method³. On these pilot questions the pass mark would have been 62% for both the candidates taking the criminal specialism and those taking the business specialism.

4.2 Test quality

Test quality statistics are set out in Table 3.

Table 3: Test Quality Statistics					
Candidates	Test Stations	Cand's (n)	Stations (n)	Alpha	SEm %
All candidates	Common Core Stations	167	7	0.84	6.10
Business	Business Stations	97	7	0.82	6.63
Criminal	Criminal Stations	70	7	0.89	6.01
Business	C Core + Business Stations	97	14	0.90	4.60
Criminal	C Core + Criminal Stations	70	14	0.93	4.24

Key quality statistics are the alpha and SEm. Alpha measures reliability and SEm measures accuracy or precision. Reliability is about how reproducible the results would be on another, similar test. It is crucial that candidates who pass deserve to pass, and candidates that fail deserve to fail, and the exam will not provide that assurance if the results do not reach a high

³ Naveed Yousuf, Claudio Violato & Rukhsana W. Zuberi (2015) Standard Setting Methods for Pass/Fail Decisions on High-Stakes Objective Structured Clinical Examinations: A Validity Study, *Teaching and Learning in Medicine*, 27:3, 280-291, DOI:10.1080/10401334.2015.1044749.

level of reproducibility on a subsequent sitting with a different paper. All measurements contain some error, and accuracy is about how large this error is.

Common targets for a high stakes professional skills assessment are an alpha of over 0.8 with an SEM of less than 5%. The alpha value in these tables is higher (better) than we would expect given the number of stations. However, our view is that caution should be exercised in extrapolating from these results. Detailed inspection of the data indicates that the high alphas arise from extreme candidate heterogeneity, an aspect which is quite common in a pilot, but which would be unlikely to be reproduced in a live exam.

5. Analysis

5.1 Number of stations for adequate reliability and precision

As explained above, alpha was inflated in the pilot due to the extreme heterogeneity of the sample. In these circumstances, where the pilot population was unusual, it is also helpful to refer to QLTS experience. The QLTS prediction of alpha for 14 stations is 0.81 and SEM for 14 stations is 5.5%. The alpha on 14 stations is therefore very close to the boundary of what is acceptable and the SEM is too high to be acceptable.

Our conclusion from all the available evidence ⁴ is that while 18 stations would be ideal, 15/16 would be adequate to achieve sufficiently precise outcomes, given stations and marking of a similar quality to that in the pilot and in QLTS. In any event there should not be less than 14 stations.

5.2 Compensation between skills

A key concern of stakeholders is the extent to which candidates can compensate for poor performance in some areas with higher scores in others. We therefore looked at the extent to which compensation was occurring between skills.

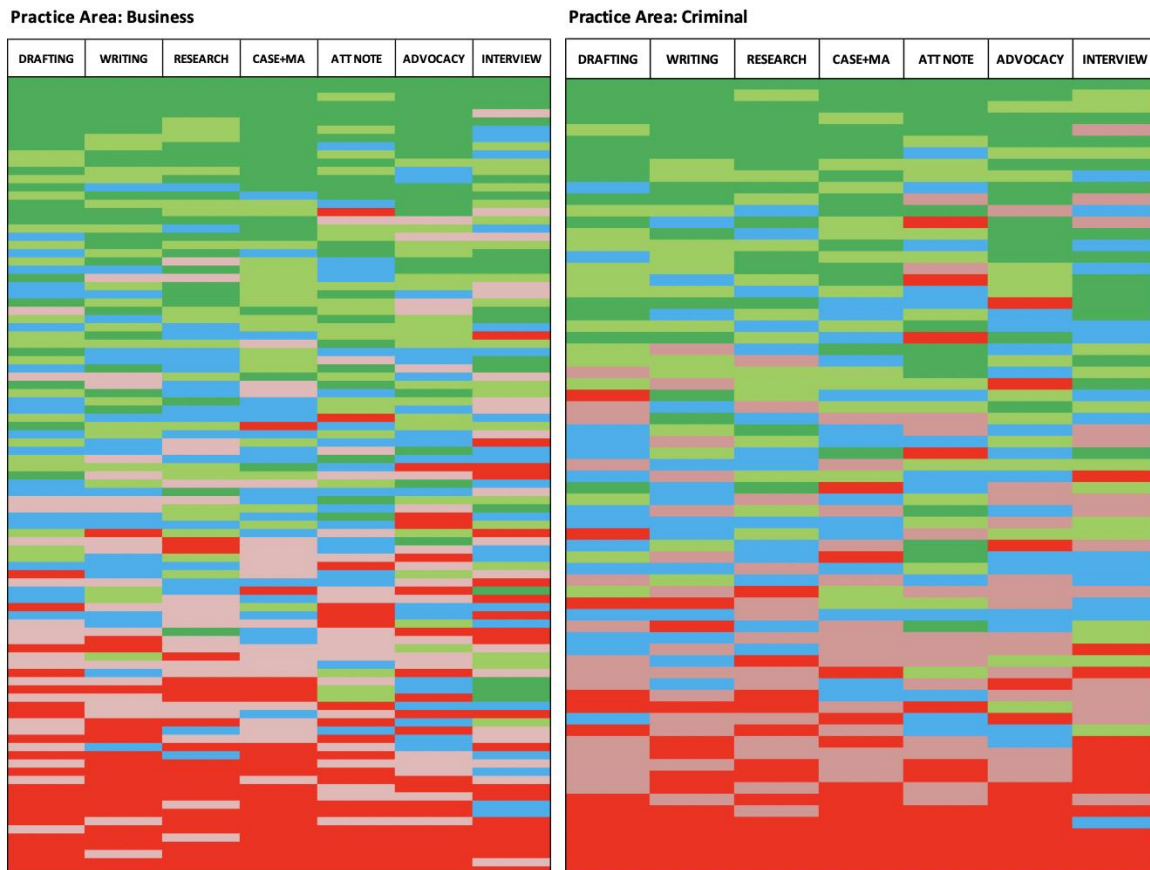
The following charts give an analysis of compensation between skills. For each skill, candidates' scores were averaged across the two occasions (common core and specialism) on which the skill was assessed. All candidates were arranged vertically top-to-bottom by their total SQE2 skills score and thus each is represented by a row in the chart. Their score on each skill is then shown divided into quintiles:

- top = bright green
- next down = light green
- middle quintile = blue
- next down = pink
- bottom quintile = red.

⁴ The conclusion is based on the pilot data, QLTS data covering over 5,000 legal skills test candidate attempts, collected across 18 sittings, and the expertise of our Advisory Board in interpreting pilot data which often results in inflated reliability because the range of scores is likely to be reduced in a live setting. Inflated reliability was also present in the SQE1 pilot.

So the appearance of red or pink quintiles near the top of the chart shows a broadly successful candidate with a deficiency in the indicated skill. The first chart shows compensation for candidates who selected the business specialism; the second for those who selected the criminal specialism.

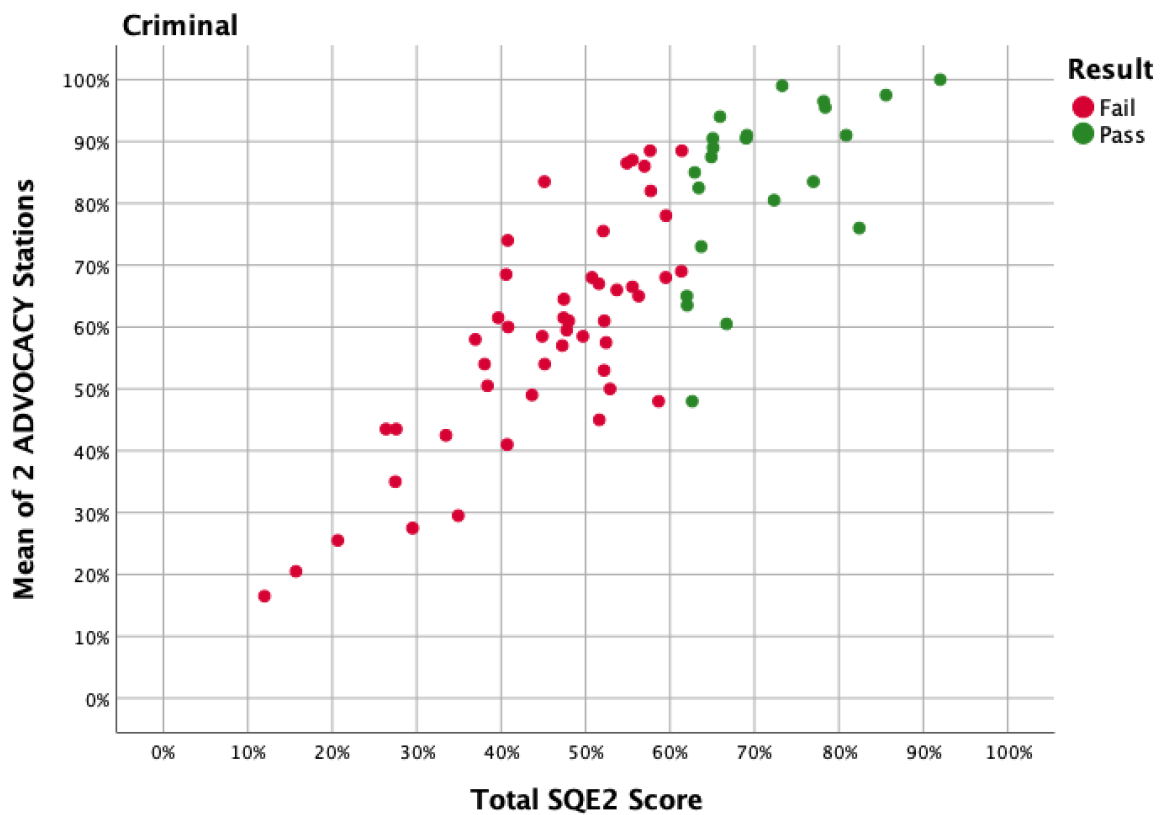
Overall it appears that while some compensation occurs, it is not very common and/or marked. On the whole, good candidates tend to do well in all sections and weak candidates tend to do poorly.



5.3 Compensation between skills (advocacy)

Compensation for poor performance in advocacy is of particular concern to stakeholders because qualification as a solicitor entails rights of audience in the courts.

The following two graphs show candidates' advocacy scores (the average of the common core and specialism marks) against their total SQE2 score by specialism and candidate outcome (result: red dot = fail and green dot = pass) with the pass mark of 62% for both groups. In summary, no candidates taking the business specialism passed the assessment and scored less than 50% on advocacy. One candidate taking the criminal specialism passed the assessment with less than 50% on advocacy but the mark was only just under at 48%.



Compensation between skills, including advocacy, did occur in the pilot. However, it was neither frequent nor marked. We would recommend that no minimum pass marks are set for separate skills, including advocacy, because of the potential lack of reliability of those pass

marks and the low confidence which can be placed on candidate sub-scores generally⁵. However this should be kept under review once the exam is live.

5.4 Fairness: Equality issues

Kaplan undertook a series of statistical analyses that aimed to provide insight into any differential performance between demographic groups within the pilot study. These analyses should be viewed with caution given the small numbers of candidates in some demographic groups, potentially confounding⁶ variables due to the fact that some categories had a very significant overlap with other categories, characteristics being self-declared by candidates, and that motivation and subsequent behaviour in a pilot will be different from that in a live exam. For some analyses, particularly that of sources of score variance, the complexity of statistical modelling is also a factor in the reliability and validity of its interpretation.

Appendix 3 shows the percentage of candidates with key protected characteristics who would have passed at different pass marks.

Sex

Appendix 3 shows differential performance by sex is present at lower pass marks, especially in the range 40 – 50%. However at 62% (the pass mark calculated provisionally for the pilot questions) and for the range where the pass mark is more likely to lie, this difference is negligible. Exploratory statistical analysis to try to determine the sources of score variance did not suggest sex was a significant factor.

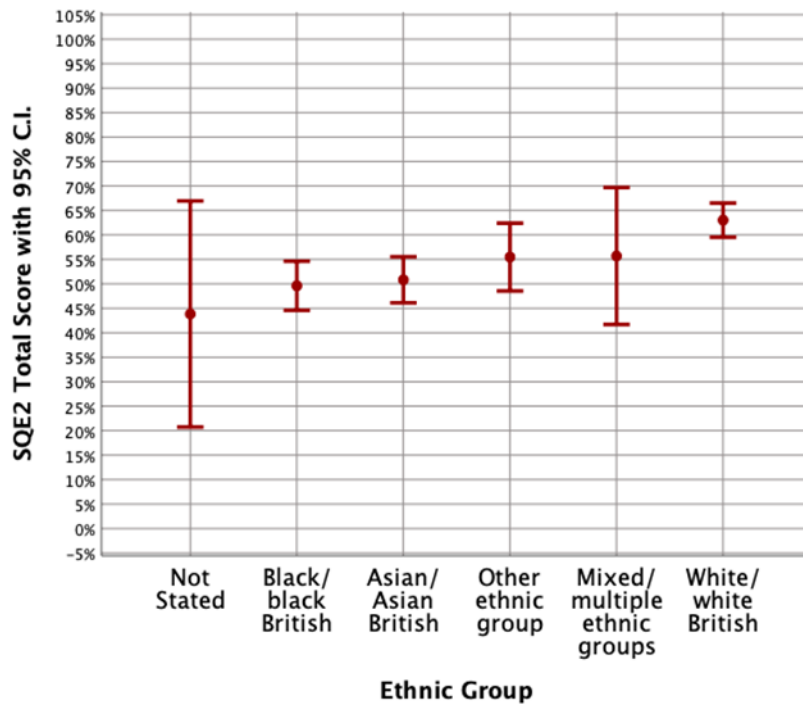
Ethnicity

Appendix 3 shows differential performance between BAME and white candidates persisted for any pass mark set between 35% - 85%.

In addition we looked further at the breakdown of scores for groups within this wide categorisation. Of course this meant that the numbers in some of the sub-categories were considerably smaller, which is a reason to view the analysis with additional caution. The following error bar chart shows the actual mean scores for each of the groups with the T-bars indicating, with 95% confidence, the range within which the population mean is likely to lie if the population behaved like this group. A rule of thumb is that if the error bars overlap vertically the difference between the groups is unlikely to be statistically significant. If there is no vertical overlap then the difference is likely to be statistically significant. There are statistically significant differences between white candidates and both black/black British and Asian/Asian British candidates.

⁵ Sub-scores and pass-marks derived from them are less reliable and less confidence can be placed in them because they are based on too few assessment points to achieve acceptable levels of reliability and precision for decision making in a high stakes licensing exam.

⁶ Confounding variables are variables which appear to be causing a predictive effect when in reality the true cause is something else. For instance it might appear that the “true” predictor of score variance was ethnicity when in reality it is completion of the GDL.



Again exploratory statistical analysis was conducted to try to determine the “true” sources of score variance in a situation where there were confounding variables. It was done for the two candidate groups separately according to which specialism candidates selected. At 2.1 we mentioned that candidates were also asked to sit a single best answer multiple choice test to enrich the data. The exploratory statistical analysis of the sources of score variance was performed separately with and without the multiple choice question (MCQ) score being entered as a predictor variable. Other predictors included in the analysis were:

- Age (Older or Younger b 1992+/-)
- Binary Ethnicity BAME/white
- English First Language
- GDL successfully completed
- LPC Compulsory units completed or LPC passed
- Sex
- U/grad degree in Law from outside the UK
- U/grad Law Degree from UK Russell Group University
- University U/grad degree in Law from UK
- And, as appropriate for each candidate group, 3/12+ work experience in criminal practice, and 3/12+ work experience in business practice.

The results are shown in the tables below:

Speciality	Predictor Added	R	R ²	Adjusted R ²	Change Statistics			
					R ² Change	Change in Variance explained	F Change	Sig. F Change
Business	Higher MCQ Score	0.790	0.624	0.620	0.624	62%	152.992	0.000
	White ethnicity	0.805	0.648	0.640	0.023	2%	5.981	0.016
Criminal	Higher MCQ Score	0.756	0.572	0.566	0.572	57%	86.935	0.000
	In younger Age Group	0.817	0.667	0.657	0.095	10%	18.314	0.000
	Undergraduate degree in law from a UK law school	0.836	0.699	0.684	0.031	3%	6.562	0.013
	White ethnicity	0.847	0.718	0.700	0.019	2%	4.284	0.043

Speciality	Predictor Added	R	R ²	Adjusted R ²	Change Statistics			
					R ² Change	Change in Variance explained	F Change	Sig. F Change
Business	LPC Compulsory Units (or LPC) Completed	0.388	0.150	0.141	0.150	15%	16.259	0.000
	White ethnicity	0.459	0.211	0.193	0.060	6%	6.961	0.010
Criminal	White ethnicity	0.368	0.136	0.122	0.136	14%	10.211	0.002
	LPC Compulsory Units (or LPC) Completed	0.460	0.212	0.187	0.076	8%	6.187	0.015

Educational factors are clearly significant predictors, particularly the MCQ score, but ethnicity is also present as a predictor.

A final analysis was to look at differential item functioning (DIF) on the 14 legal skills stations. We used DIF to look at differential performance on stations by key protected characteristics, controlling for overall group differences. In other words DIF looks at whether individual questions are disadvantaging protected groups over and above any general differences in performance between those groups. The results were encouraging in that none of the oral stations (where unconscious bias might play a role) showed DIF. One station (the common core attendance note) showed significant DIF for binary ethnicity when analysis was limited to just the seven stations in the common core, but not when analysis included all 14 stations taken by either the business or the criminal group. The comparison with 14 stations bears more weight than that with only seven but nevertheless we will look further at this station to see if we can establish any reason for its DIF.

Disability

There were no differences in performance by disability in SQE2 skills. However only 12 candidates declared a disability under the Equality Act 2010 which was too few for any conclusions to be drawn. Numbers will be larger following implementation when these analyses will be performed again.

Conclusions and Equality issues

In summary there was differential performance by binary ethnicity in the SQE2 pilot. This is substantially confounded by educational differences, and the DIF analysis was encouraging, but nevertheless there is cause for concern. Kaplan will continue to work with the SRA to ensure that candidates from minority groups protected under the Equality Act 2010 are not unfairly

disadvantaged while maintaining the standards of the assessment. Some of the measures we will be taking are listed below:

- We will work to recruit a diverse range of assessors, markers and question writers and will work with the SRA to reach out to solicitors from minority protected groups under the Equality Act to assist in encouraging solicitors from these groups to apply to become SQE assessors, markers and questions writers
- All assessors, markers and question writers will receive diversity training
- All assessors will receive unconscious bias training prior to each live assessment as part of their assessor training
- There will be statistical monitoring of the performance of live assessors for evidence of unconscious bias
- We will use DIF to look at whether individual questions are disadvantaging protected groups over and above any general differences in performance between those groups
- We will monitor performance of minority groups protected under the Equality Act 2010 and report this statistically to the Exam Board.

5.5 Work experience

An analysis of the predictive effect of work experience presents a mixed and complex picture with work experience having an inconsistent predictive effect on performance. In addition, the analyses presented are based on small numbers with many confounding variables.

The actual marks of candidates with and without work experience are presented in the following table:

Table 6: Candidates' Total Scores (common core + specialty) by Work Experience in any Practice Area and in the Practice Area of their Specialty						
Statistic	3 months+ Work Experience					
	All candidates: work experience in any practice area		Business candidates: work experience in business practice area		Criminal candidates: work experience in criminal practice area	
	Yes	No	Yes	No	Yes	No
Number of Candidates	127	40	52	45	19	51
Mean Score (%)	57.59	50.22	61.03	53.63	53.35	53.39
Median Score (%)	58.50	51.83	60.55	52.48	52.20	54.89
SD (%)	15.64	14.79	15.85	13.30	21.89	13.98
Minimum Score (%)	11.97	15.69	29.02	23.84	11.97	15.69
Maximum Score (%)	92.44	75.12	92.44	75.12	91.99	78.38
Range (%)	80.47	59.43	63.42	51.28	80.02	62.69

The highest marks were achieved by candidates with more than three months' work experience in the practice area of the specialism choice. Particularly notable was a group of high performing candidates who chose the business specialism and who had experience in the business specialism area. However, more than three months' work experience was neither a necessary nor a sufficient condition of doing well in the pilot. There were many examples of

candidates doing very well without any work experience, while others did very poorly despite specialism or any legal work experience.

The statistical analyses also produced a mixed picture. Univariate analysis, which shows if there are significant differences in performance by individual characteristics, suggested that more than three months' work experience in any practice area was a significant predictor of performance for candidates who took the criminal specialism, but not for candidates who took the business specialism. Work experience in the practice area of the specialism was a significant predictor for candidates who took the business specialism but not for those who took the criminal specialism.

Multivariate analysis explores which are the significant "true" predictors of candidate performance where there are confounding variables. Again it should be viewed with considerable caution. Multivariate analysis suggested that work experience was not a significant predictor of performance for either the business or the criminal candidates.

In conclusion then this is a mixed and complex picture with work experience having an inconsistent predictive effect on performance.

5.6 Specialisms/common core/uniform exam

As outlined at 2.1 the pilot was designed to evaluate three options for the design of SQE2.

5.6.1 Specialisms only

In this model candidates select two areas of legal practice out of the five mentioned above at 2.1 and are examined in each of the skills, only in those practice areas. Pass marks would be set by borderline regression⁷ accompanied by quality assurance measures.

Advantages:

This model would allow more candidates to be assessed only in their area(s) of current practice, and would be favourably viewed by some segments of the market. It might be less disruptive to the market and may also reduce training costs.

Disadvantages:

Ten different pass marks would be required per sitting to allow for the ten specialist combinations. With decisions about pass marks made using borderline regression and quality assurance methods, and without any direct method of comparing standards in the different

⁷ Naveed Yousuf, Claudio Violato & Rukhsana W. Zuberi (2015) Standard Setting Methods for Pass/Fail Decisions on High-Stakes Objective Structured Clinical Examinations: A Validity Study, *Teaching and Learning in Medicine*, 27:3, 280-291, DOI:10.1080/10401334.2015.1044749.

specialist areas, it is difficult to be sure that candidates are being assessed fairly and consistently across different specialisms.

Results from the pilot (discussed below) suggest that legal skills were not sufficiently generic that a candidate's performance in one area would translate into similar performance in another area. This would further undermine the idea that borderline regression is sufficient to achieve a common standard of entry into the profession. It also suggests that the model may not be suitable from a regulatory perspective in that it may not adequately protect the public by certifying competence across the areas of a solicitors' practice.

5.6.2 Common core: Putting the business and criminal specialisms on a common scale via the common core

As explained at 2.1, candidates in the pilot took a test of skills within a universal part of SQE2 that all candidates took (the common core) and a test of skills in a specialism which for the purposes of the pilot was either business practice or criminal practice. We wanted to investigate the process and outcomes of attempting to put candidate marks in the criminal and business specialisms on a common scale via the common core.

Advantages of putting the criminal and business specialisms on a common scale via the common core:

Our aim was to use the common core as a yardstick of the difficulty of the specialisms so that different specialisms can be put on a common scale. If successful this would allow room for specialisation while retaining a universal standard for admission.

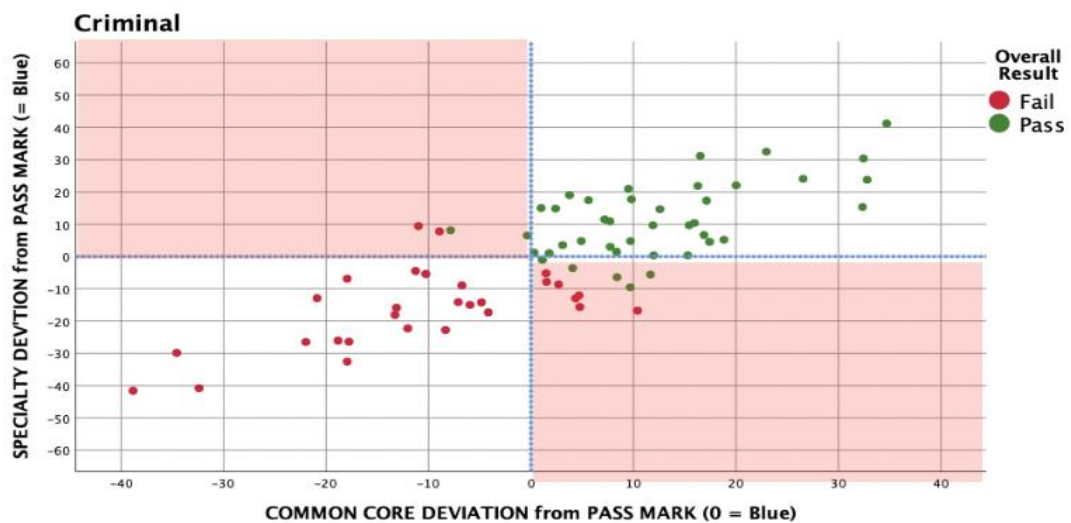
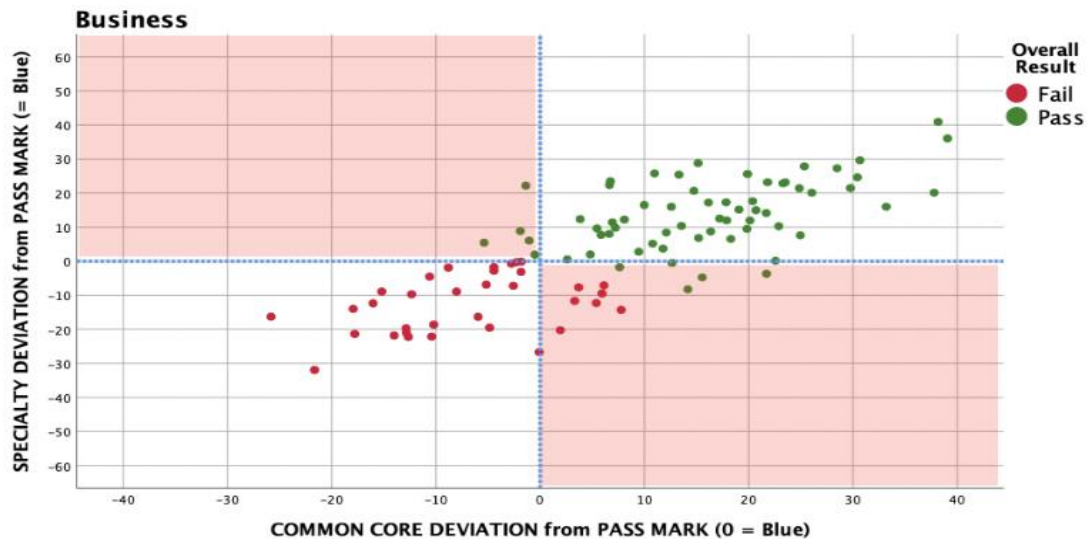
Disadvantages of putting the criminal and business specialisms on a common scale via the common core:

Mathematically we were able to perform the exercise of establishing, via the common core stations, what a mark in the business specialism (and the pass mark) would convert into in the criminal specialism and vice versa. This exercise is known technically as coarse projection, with the pilot mirroring a Non Equivalent Anchor Test (NEAT) equating design with internal anchor. However, what we wanted to establish was whether these mathematical adjustments were justified.

The fundamental question here is whether the common core did act as a true yardstick of performance in the specialisms. The key to this will be the extent to which legal skills are generic and therefore transferable from one context to another.

Previously it was noted that demographically the two groups of candidates were very different. This is significant because a prerequisite for this kind of exercise is that the groups are broadly similar. If groups taking different specialisms are diverse, confidence in the outcomes must decrease as it becomes harder to separate factors to do with differences between the groups and factors to do with the difficulty of the assessment. We also showed that while difference in performance between the two groups of candidates on the specialism was not statistically significant, it was statistically significant on the common core. This must also throw doubt on the extent to which the common core was acting as a yardstick of performance in the specialisms.

In order to investigate the issue further we looked at the relationship between marks on the common core and marks on the specialisms. The two charts which follow show, separately for business and criminal candidates, each candidate's scaled mark on common core stations on the X-axis against their scaled mark on their specialism skills stations on the Y-axis. Being scaled to the pass mark, the pass mark here is zero. The colour shows their exam outcome – pass (green) and fail (red).



Candidates in the plain areas of each chart were above (or below) each pass mark (specialism and common core). Candidates in the areas shaded red would have received two different outcomes from the two sets of stations. These outcomes are not consistent. Some candidates would have passed on the common core but failed on the specialism (whichever specialism it was); and some would have failed on the common core but passed on the specialism (again whichever specialism it was). Further, this discrepancy was not a matter of a few marks. Some candidates who passed one area failed the other by as much as 20%.

Correlation coefficients between skills across the two stations (common core and specialism) in which they were assessed complete the picture. These show the correlations in the range of 0.311 at the lowest to 0.565 at the highest. The statistics indicate that only between 9.7% of

the score variance (in interview stations) and 31.9% of the variance (in case and matter analysis) was shared between the two instances in which the skill is assessed.

The conclusion must be that the common core did not act as a sufficiently good yardstick to measure the difficulty of the specialisms and that legal skills are not sufficiently generic to justify doing so.

5.6.3 Uniform exam

Various designs of a uniform exam are possible. However the underlying principle would be that the exam would take a sample from all the skills in all the practice areas and that all candidates would take the same exam.

Advantages:

The advantages of a uniform exam are that it is the most defensible alternative as a single entry point into the solicitors' profession. It guarantees a universal standard with a single pass mark for all sitting the exam at the same time. It is in line with other credentialing exams⁸. By sampling more widely from all the practice areas and skills it affords the best protection of public safety. Because of its simplicity and lack of logistical complications it is also significantly cheaper than the alternatives.

Disadvantages:

The disadvantages are that some stakeholders favour specialisation. In addition there is a concern that it would be disruptive to the market and would lead to an increased need for training.

As regards disruption and an increased need for training this can be mitigated to some extent by ensuring that only fundamental legal principles, not detail, are examined. Candidates should only be required to have sufficient knowledge to make them safe and competent to practise on the basis that they could look up detail later. In a case where detail was required to answer a question, candidates would be supplied with materials which a Day One Solicitor would look up. In addition, the tentative results of the pilot suggest these issues may be less of a concern than envisaged because work experience is neither a necessary nor a sufficient condition of doing well in the exam.

Kaplan would recommend the SRA adopt a uniform design for SQE2.

⁸ Examples of uniform credentialing exams in the UK are the transfer exam for overseas doctors and the exams of the various medical Royal Colleges (for example, for general practitioners); veterinary qualification exams; the qualification accountancy and finance exams of ICAEW, CIPFA, CIMA, chartered tax accountants and chartered financial analysts; and the Qualified Lawyers Transfer Scheme. In the United States, the Uniform Bar Examination, Multistate Professional Responsibility Examination, and medical licensing exams (the United States Medical Licensing Exam and its component parts) are examples of uniform credentialing exams.

6. Recommendations

As a result of the strong indicative evidence from the SQE2 pilot; expert psychometric analysis including from an Advisory Board with world leading expertise in professional assessment for licensure, including for licensure in law; and QLTS data covering over 5,000 legal skills test candidates collected across 18 sittings, Kaplan recommends the following:

18 stations would be ideal for SQE2 as regards reliability and precision. 15/16 would be adequate to achieve sufficiently precise outcomes given stations and marking of a similar quality to that in the pilot and in QLTS. In any event there should be not less than 14 stations.

No minimum pass marks should be set for separate skills including advocacy. This should be kept under review once the exam is live.

The SRA should adopt a uniform design for SQE2. Candidates should only be required to have sufficient knowledge to make them safe and competent to practise on the basis that they could look up detail later. In a case where detail was required to answer a question, candidates would be supplied with materials which a Day One Solicitor would look up.

Appendix 1

SQE 2 pilot blueprint (Business Specialisation)		Context				
		1 – Criminal Litigation: Criminal liability	2 – Dispute resolution: Contract and tort	3 – Property: Land Law	4 – Wills & intestacy, probate: Trusts	5 –Business: Contract
Skills	1 – Client Interviewing			*		+
	2 – Completion of Attendance note/Case Analysis			*		+
	3 – Advocacy/oral presentation		*			+
	4 – Case and matter analysis				*	+
	5 – Legal Research		*			+
	6 – Legal Writing	*				+
	7 – Legal Drafting					*+
*	Core					
+	Specialisation					
	No station					

SQE 2 pilot blueprint (Criminal Specialisation)		Context				
		1 – Criminal Litigation: Criminal liability	2 – Dispute resolution: Contract and tort	3 – Property: Land Law	4 – Wills & Intestacy, probate: Trusts	5 –Business: Contract
Skills	1 – Client Interviewing	+		*		
	2 – Completion of Attendance note/Case Analysis	+		*		
	3 – Advocacy/oral presentation	+	*			
	4 – Case and matter analysis	+			*	
	5 – Legal Research	+	*			
	6 – Legal Writing	+*				
	7 – Legal Drafting	+				*
*	Core					
+	Specialisation					
	No station					

Appendix 2

Candidate Demographics

Variable	Value	All Candidates		Business		Criminal		Sig Diff B->C by χ^2 / Fisher ?
		N	%	N	%	N	%	
Level of parental education when 18	1+ degree level qualification	66	39.5%	41	42.3%	25	35.7%	NS
	No formal qualifications	27	16.2%	14	14.4%	13	18.6%	
	Not stated	14	8.4%	9	9.3%	5	7.1%	
	Qualifications below degree level	60	35.9%	33	34.0%	27	38.6%	
School type classified	Independent	16	9.6%	13	13.4%	3	4.3%	NS
	Other / not stated	7	4.2%	4	4.1%	3	4.3%	
	Outside UK	58	34.7%	32	33.0%	26	37.1%	
	State: non-selective	69	41.3%	38	39.2%	31	44.3%	
	State: selective	17	10.2%	10	10.3%	7	10.0%	
Experience as CILEX, trainee or apprentice solicitor, or paralegal? (In any area)	No experience	40	24.0%	21	21.6%	19	27.1%	NS
	Experience	127	76.0%	76	78.4%	51	72.9%	
Experience as CILEX, trainee or apprentice solicitor, or paralegal In own practice area?	No experience	96	57.5%	45	46.4%	51	72.9%	SIG Fisher p < 001
	Experience	71	42.5%	52	53.6%	19	27.1%	

Variable	Value	All Candidates		Business		Criminal		Sig Diff B->C by χ^2 / Fisher ?
		N	%	N	%	N	%	
Have you successfully completed the Graduate Diploma in Law ?	Not completed	132	79.0%	71	73.2%	61	87.1%	SIG Fisher p = .034
	Completed	35	21.0%	26	26.8%	9	12.9%	
Have you or will you have completed the compulsory modules of the LPC by the time of the pilot? (Or completed the LPC)	Not completed	87	52.1%	45	46.4%	42	60.0%	NS
	Completed	80	47.9%	52	53.6%	28	40.0%	
If LPC completed, what was your overall grade/award?*	Commendation	14	20.9%	8	19.5%	6	23.1%	SIG χ^2 p = .002
	Distinction	34	50.7%	27	65.9%	7	26.9%	
	Pass	19	28.4%	6	14.6%	13	50.0%	
Do you have a UK university undergraduate degree in law ?	No	74	44.3%	44	45.4%	30	42.9%	NS
	Yes	93	55.7%	53	54.6%	40	57.1%	
Do you have a UK University law degree from a UK Russell group University?	No	144	86.2%	82	84.5%	62	88.6%	NS
	Yes	23	13.8%	15	15.5%	8	11.4%	
Do you have an overseas University law degree?	No	122	73.1%	73	75.3%	49	70.0%	NS
	Yes	45	26.9%	24	24.7%	21	30.0%	
What class of University law degree were you awarded?*	Other	9	6.5%	7	9.1%	2	3.3%	NS
	Third	5	3.6%	2	2.6%	3	4.9%	
	Two two	41	29.7%	23	29.9%	18	29.5%	
	Two one	64	46.4%	36	46.8%	28	45.9%	
	First	19	13.8%	9	11.7%	10	16.4%	

* = variable not applicable to any missing candidates

Table 3: Candidates' Personal Demographics								
Variable	Value	All Candidates		Business		Criminal		Sig Diff B->C by χ^2 / Fisher ?
		N	%	N	%	N	%	
Sex	Female	109	65.7%	63	65.6%	46	65.7%	NS
	Male	57	34.3%	33	34.4%	24	34.3%	
English 1st Language?	Yes	124	74.3%	68	70.1%	56	80.0%	NS
	No	43	25.7%	29	29.9%	14	20.0%	
Religion classified*	Christian	66	41.8%	36	39.1%	30	45.5%	NS overall, but C had Sig more Muslims than B χ^2 p = 0.05
	Muslim	23	14.6%	9	9.8%	14	21.2%	
	None	54	34.2%	36	39.1%	18	27.3%	
	Other (<10/religion)	15	9.5%	11	12.0%	4	6.1%	
Sexual orientation*	Bisexual	6	3.8%	1	1.1%	5	7.7%	NS
	Gay/lesbian	2	1.3%	1	1.1%	1	1.5%	
	Heterosexual/straight	149	94.9%	90	97.8%	59	90.8%	
Classified Ethnicity*	Asian/Asian British	41	25.5%	24	25.5%	17	25.4%	NS
	Black/Black British	32	19.9%	15	16.0%	17	25.4%	
	Mixed/Multiple Ethnic Groups	8	5.0%	6	6.4%	2	3.0%	
	Other Ethnic Group	13	8.1%	7	7.4%	6	9.0%	
	White/White British	67	41.6%	42	44.7%	25	37.3%	
Binary Ethnicity^	Non-white	94	58.4%	52	55.3%	42	62.7%	NS
	White	67	41.6%	42	44.7%	25	37.3%	
Gender ID different from birth?*	No	166	100.0%	96	100.0%	70	100.0%	NS
	Yes	0	0.0%	0	0.0%	0	0.0%	
Age classified	Older <=1991	78	46.7%	42	43.3%	36	51.4%	NS
	Younger >=1992	89	53.3%	55	56.7%	34	48.6%	
Disability as per Equality Act 2010?*	No disability	151	92.6%	88	92.6%	63	92.6%	NS
	Yes, a disability	12	7.4%	7	7.4%	5	7.4%	
Reasonable adjustments requested?	No	157	94.0%	91	93.8%	66	94.3%	NS
	Yes	10	6.0%	6	6.2%	4	5.7%	

* a few candidates did not respond

Appendix 3

Percentage of candidates with key protected characteristics passing at different pass marks

