

Review of first sitting of SQE1

20 January 2022

Independent Reviewer for SQE – Geoff Coombe

Executive Summary

The first sitting of the first part of a single, nationally recognised solicitor qualifying exam, for candidates wishing to practice in England and Wales was held in November 2021. The first candidates for the Solicitors Qualifying Examination sat part one (SQE1) exams on 8 and 11 November 2021.

As the Independent Reviewer for the SQE, I provide external assurance as to whether the exam is likely to deliver outcomes which are fair, defensible and will command public confidence. This report is based on my observations of the first live SQE1 exam. I observed key aspects of the planning and preparation for the exam as well as observing one of the exam venues when candidates were present and the subsequent operational processes and decisions when determining which candidates should pass. These exams are regulated by the SRA, following approval by the LSB, and Kaplan are their appointed examination services supplier.

Overall, the initial SQE1 exam appears to have successfully delivered valid fair, reliable and defensible outcomes. Each of the stages of preparation; delivery and processing outcomes for the exam demonstrated significant evidence of good practice. The operational and logistical processes to set up and deliver the exam proved effective.

I observed good evidence of a robust lessons learned process being implemented which will make, mainly minor, improvements for future sittings. The technical analyses, which evaluates the questions set and the examination overall, was thorough and provided a wealth of performance information about candidates that was previously not available nationally before the SQE was set up.

The process for determining the pass boundaries for each SQE1 exam was well considered and effective. This demonstrated many aspects of best practice and Kaplan showed evidence of having built on learning and experience of prior Qualified Lawyer Transfer Scheme (QLTS) exams they have operated previously. This in addition to the SQE pilot and used worldwide evidence of best practice in similar high stakes professional qualifying examinations contexts. The academic literature supports the methods used to process and determine the results of the exam and the pass boundary. From a technical perspective the SQE1 exams appear to deliver an effective assessment.



Differences in candidate performance outcomes were observed, these were typical of those seen in other professional exam contexts. For example, candidates declaring themselves to be of white ethnicity achieved a greater percentage of passes than all other ethnic categories.

I am satisfied that the setting and editing of the questions and mark schemes, the wholly objective marking, and the very detailed post results review of performance by different candidate groups showed good practice (for example to avoid unintentional bias) and indicated every effort had been made to ensure fair outcomes for all candidates.

The new data generated offers both the SRA and Kaplan new opportunities to continue to investigate issues causing this difference. Not surprisingly, and reassuringly, factors such as prior ability of candidates, for example those achieving a top grade at university, and prior work experience were indicators of a greater likelihood to pass.

[Open all \[#\]](#)

Method of review and sources of evidence

The review of the November 2021 SQE1 sitting was achieved in three main stages. The preparation of the questions and make-up of the SQE1 exams, the operational and logistical activities associated with preparing and conducting the exam, and the post exam results processes.

My review was conducted through interviews with key people at Kaplan and SRA and observation of key meetings and activities, supported by documentation provided for these meetings and subsequently requested to provide evidence of actual performance. This included:

- observing training meetings for new exam question writers
- reviewing processes and training materials for all preparatory processes
- reports from those participating in Angoff standard setting meetings and outcome data
- interviews with key Kaplan and SRA staff before and after the exam
- observing candidates sitting the exam at a regional exam venue
- summary management information about the candidate services contact centre, the technical exam performance and advice provided to Pearson VUE exam centres on the day of the exam
- candidate survey response information
- Kaplan lessons learned report on Nov 2021 SQE1

Preparation for the SQE1 exam product

My earlier SQE1 readiness report, published in April 2021, summarises much of the preparatory activity for the SQE1 exam. I observed good practice in the training of new question writers, including the quality

assurance process that is used to commission, review, edit and, if necessary, amend questions before they are approved as suitable to go into the question bank for use in the SQE1 exams.

This observation included training provided to question writers to be aware of risks of unconscious bias and ensuring questions are written to be as precise and simple as possible. So that, for example, the language used in questions should not have any unnecessary complexity, and avoids cultural, gender, religious and ethnic bias or stereotypes.

A 'blueprint' for each SQE1 exam was drawn up to ensure appropriate coverage of the Functioning Legal Knowledge (FLK), as well as a suitable number of ethics-based questions.

By mid-summer 2021 Kaplan had produced a significant number of approved SQE1 exam questions to populate their question bank. The bank was of a suitable scale to manage risks around the security of the exam prior to it being taken.

By September 2021 each of the two SQE1 exams had been compiled, meeting the expectation of the blueprint for each exam and the process of 'rendering' the exam ready to display on computer screens in the Pearson VUE exam centres was well advanced. The SQE1 exam is split into two parts assessing different topics of functioning legal knowledge. Each exam comprises 180 questions.

Overall, the preparation of the exams demonstrated much good practice and a suitable 'review, do and improve' cycle is operating so that learning from each exam cycle can be fed into future question preparation activity.

SQE1 Angoff standard setting meetings

Angoff is a process whereby qualified solicitors act as judges and assess the likelihood of a just competent newly qualified solicitor correctly answering each question in the SQE1 exam. The collective judgement of the Angoff panel members is therefore a key determining factor on how, and where, the pass boundary is drawn when reviewing the candidate results data post exam.

The Angoff panels were convened to record an Angoff score for all questions in both the main and reserve papers for FLK1 and 2.

Given that each question requires careful judgement the process took nine working days, these were completed between 20 September and 8 October 2021.

Full training was provided to panel members on 20 September by Kaplan staff responsible who were also present throughout all nine days of panel sittings. The training was observed by SQE1 subject heads, Kaplan's

Head of Quality Assurance and a subject matter expert appointed by SRA. All panel members are solicitors of England and Wales who are newly qualified solicitors (up to two years PQE) or more experienced solicitors who can demonstrate an understanding of the standard expected of a newly qualified solicitor at the start of their practice as a solicitor. The panel members were reassuringly diverse in their backgrounds. Before each panel, the Kaplan Academic Resources Team compiled a spreadsheet with all the questions to be presented to each panel. Questions were presented to panel members and scores were entered 'live' and the Angoff judgements were automatically populated and shared on screen for monitoring purposes. At the end of each day, final checks were made to ensure that all scores were recorded and back-ups were saved for safe keeping.

Once the Angoff process was complete, the Kaplan Director of Psychometrics and Assessment Development completed a final check to ensure the Angoff scores were correct and that no amendments were necessary. The process included significant quality assurance activity to check all steps had been taken appropriately, for example, a full reconciliation exercise was carried out to ensure each score was recorded against the correct question, showing the correct Panel date and displayed to three decimal places.

Overall, the conduct and outcomes of the Angoff panels delivered judgements in which I have confidence. This is because those participating appeared to provide a secure representation of solicitors with requisite knowledge and experience of expectations of newly qualified solicitors and of the functioning legal knowledge being assessed.

Reasonable adjustments

I observed that reasonable adjustments are treated with high importance at Kaplan and SRA. Providing reasonable adjustments is a requirement under the Equality Act 2010. In 2020, policies and processes for responding to requests for reasonable adjustments from candidates were agreed, and the process was updated in 2021 following engagement and collaboration with stakeholders representing disability groups. The opening of candidate registrations and SQE1 bookings was the first time Kaplan were able to put these processes into action and 76 were received. Kaplan have set up an Equality and Quality (E&Q) team which takes responsibility for agreeing and arranging reasonable adjustments. Implementation has identified some areas for improvement which have been quickly addressed by the E&Q team, and others which require further work or future enhancements.

Overall, given the confidentiality, complexity and importance of these processes and the sensitivities involved when making arrangements for

candidates, the processes worked at least satisfactorily and were often good.

The conduct of the SQE1 exams

The first half of the SQE1 exam was held on Monday 8 November and the second on Thursday 11 November 2021. Exams took place across England and Wales and internationally at 139 different Pearson VUE test centres.

I attended the Exeter-based test centre to observe the first examinations taking place. I observed the method by which candidates were: welcomed; booked into the venue, security checked and briefed prior to the exam. I then observed candidates taking the exam from the proctor observation room. The process by which candidates were received was efficient and business-like, which included a friendly receptionist. The briefing for each candidate was individually delivered and consistent.

Security checks of both identity and to demonstrate proof that each candidate was not carrying any aide or device to cheat was thorough and consistent. The examination administration appeared to be smooth and well organised, candidates appeared well informed about what to expect and all started the exam on time at the same time.

After my visit to one test centre, I sought assurance as to the performance of other centres nationally. There were a small number of relatively minor issues affecting candidates on the day of the exam in other venues, however, with one exception, all issues were resolved in time for candidates to successfully take and complete the exam. The unfortunate exception was where a candidate had completed a booking but this booking did not show at the test centre, meaning the candidate was unable to sit the first SQE1 (FLK1) exam.

A thorough investigation into the issue is being conducted and this will prove a key lesson to learn to avoid a similar future occurrence. In a small number of other cases candidates were directed to the wrong (or old) address of an assessment centre, while these cases did not stop the candidates affected from taking the exam, it led to some candidates reporting a more stressful start to the exam. I understand this issue has already been raised by Kaplan with Pearson VUE so that additional checks take place prior to the next live exam dates.

If issues do arise which affect candidates' performance on the day of the exam, a process for raising mitigating circumstances exists. Kaplan carefully record and consider each issue raised by a candidate. I reviewed the mitigating circumstances log and the Assessment Board (see later) reviews and makes decisions on all applications in relation to mitigating circumstances.

My conclusion was that a thorough and robust process exists which has the right balance of investigating, reviewing and implementing action where a candidate has suffered a genuine misfortune or problem that was outside their control, without undermining the integrity of the exam by giving candidates any 'allowance' where they do not deserve it. It is important a consistent line continues to be taken to applying mitigating circumstances outcomes throughout the operation of the SQE, because there is evidence from other exam contexts that such processes can be vulnerable to exploitation for dubious purposes, or are applied inconsistently, as candidates (and sometimes those that advise or teach/train them) wish to receive some benefit to their exam outcomes.

In addition to the mitigating circumstances process, a complaints process for candidates exists. I reviewed the action taken on the basis of each of the small number of complaints received and was satisfied that the outcomes appeared to be fair and reasonable.

Once again, as the SQE continues, as a case log is built over time, it will be important to refer to outcomes of previous complaints to ensure consistency of decision making over time, unless there are grounds to take a different view, for example, due to technological developments over time in the delivery of the exams.

The IT systems used to manage and deliver the processes relating to candidate registrations and booking, candidate support team, exam product creation and rendering ready for assessment, Pearson VUE test centres, test delivery and results processing appeared to have delivered at least satisfactorily. With the exception of the candidate who was unable to take the exam, as mentioned above.

The functionality of the IT systems was a mixture of building upon Kaplan exam platforms used in other (mainly closely related) professional exam contexts and new capability designed for SQE. Given the complexity of the new data storage and processing requirements, the need to have satisfactory user experiences and new processes to support, this was a significant achievement.

Kaplan maintained a comprehensive lessons learned log and created a well-considered management summary report, demonstrating a robust process in place for following up lessons learned, both positive and negative. This is a critical aspect of learning and improving whenever a new exam is set up. I was impressed by the detail and care that went into this process and am convinced that future candidates will benefit from the (relatively few and minor) issues being acted upon in time for future SQE1 sittings.

Candidate survey

Shortly after the exams were sat Kaplan issued a candidate survey to seek feedback on the candidate experience to gain insight to improve

further sittings and understand better any issues arising from this sitting. Approximately two thirds of all candidates sitting completed the survey, which is a remarkably high response. A summary of the survey results was shared with me and the SRA.

Overall, the survey responses demonstrated the effectiveness of the processes deployed and while candidates offered a range of constructive views the overriding impression was that the exam had worked well.

The survey itself was exemplary, not only were quantitative views captured but a wealth of qualitative information was supplied by candidates. This included lengthy and highly articulate written responses, some of which were followed up, via in depth focus groups or one to ones held with smaller sub sets of candidates. Kaplan have demonstrated an excellent commitment to hear from and learn from any issues arising as well as receiving feedback about what worked well for, and was appreciated by, candidates as they experienced the exams.

Given this was the first live cohort the outputs derived by Kaplan should provide confidence that any changes will further enhance the candidate experience while maintaining the quality of features of this exam that were appreciated by candidates. Although some candidates commented on perceived differences in the style or nature of questions in the live exam compared to practice questions made available in advance, I could find no evidence to support this concern.

Setting the pass mark and other post-exam processes

The candidate responses from the FLK1 and FLK2 exams are marked automatically to generate question (or item) level information about their performance. This information is compiled and carefully analysed by Kaplan's psychometric expert, it is then checked and assured by the SRA's independent psychometrician. The outcomes of these analyses were provided to the Assessment Board (where the final pass mark is decided).

Overall, the process for creating the statistical analyses and quality assuring the results data was thorough and comprehensive. As this was the first sitting of a new exam a wide range of different psychometric based statistical analyses were investigated so that many aspects of the performance of each item, and overall candidate performance, could be reviewed.

The great benefit of having the opportunity to create different statistical analyses is that it allows a thorough psychometric review of the efficacy of the exam to take place. This included detailed demographic analyses, using the candidates self-declared demographic information. All the initial high-level outcomes of the exam indicated the questions set had effectively discriminated candidate performance and had been set at an

appropriate level and range of difficulty and the summary indicators are of a well-functioning assessment.

Prior to the Assessment Board, a meeting was held between senior staff in Kaplan and the SRA to review these analyses. I observed this meeting, the purpose of which was to ensure that all the information needed was available to enable the Assessment Board to make an appropriately informed decision about the pass mark. A very good discussion reviewing the performance of the exam took place. While the meeting, rightly, emphasised how well the assessments had performed, there were two issues which received particular focus: the difference in performance outcomes between FLK1 and FLK2 and across different ethnic groups.

It was noted that candidates generally had achieved a slightly higher performance on FLK1 than FLK2. The judges, involved in the earlier Angoff meeting, who reviewed the difficulty of all the items had expected the FLK2 exam to be slightly more demanding than FLK1 and the summary of their judgements predicted a slightly lower pass mark, which is what happened. However, this did not wholly account for the actual difference in performance observed. We can speculate as to why this might be the case. It is possible that as candidates take FLK2 just three days after FLK1 they were becoming fatigued and/or had less time to prepare as well for FLK2.

It was also observed that candidates generally did less well on the practical FLK topics of which there are three topic areas assessed in FLK2 and only one in FLK1. There is no reason to suspect that the difference observed was for any other reason than candidates performing (and therefore preparing) differently across the two papers, less well generally for FLK2. This was tested through an array of analyses available about candidate performance on each item, for example, there was no evidence that FLK2 items took a longer time to read than FLK1 items. These analyses support a conclusion that questions appear appropriately set.

I therefore recommend that Kaplan and the SRA continue to monitor the potential for differential performance across FLK1 and FLK2 in future sittings. With a view to monitoring trends and potentially offering support to future candidates and training providers about improving preparation and to see if any mitigations, such as spacing the exams further apart in the future, might be worthy of consideration. I am not recommending any immediate action, rather to see how candidate performance settles over time and seek understanding of this. It is likely that this first cohort will prove to be atypical compared to future cohorts as the source of candidates becomes more stable over time.

The performance outcomes of candidates declaring themselves to be of white ethnicity compared to all other ethnic groups generally showed higher pass rates in the former group. Unfortunately, this pattern is

consistent with other comparable professional exams leading to licencing in England and Wales.

It is important that all possible safeguards to prevent potential unfairness for one or more demographic (including ethnic) groups are taken. For SQE1 exams I observed, or received evidence of, the following safeguards:

- Kaplan training was provided to question writers with advice to make sure of fairness to all candidates including careful use of language to make questions as easy to comprehend as possible and use of neutral terms eg describing people by role and ensuring context of questions is culturally neutral
- Editing of questions by those experienced academic assessment experts that provided the training to the question writers and therefore were well versed in avoiding potential problems with unfairness in the wording of questions and mark schemes
- Subject matter experts (SMEs) appointed by the SRA reviewed a sample of questions prior to the exam with an instruction to raise any concerns about context of wording of questions which may disadvantage any demographic group
- All candidate responses are objectively marked, SQE1 exams are processed automatically, therefore there cannot be any human bias in the marking process
- Post-exam review of items that performed atypically. These reviews included reviewing items which showed a significant difference in performance by different demographic groups, which could positively or negatively bias a particular demographic group, including ethnicity. The review provided a back stop opportunity to ensure there was no potential evidence of bias in the way the question had been presented to the candidate.

These activities appear appropriate to assure that all demographic groups are treated fairly and equally in the assessment process. As there was a difference in performance across different ethnic groups this received particular attention post-exam. This included an SRA subject matter expert review, who is external to the Kaplan team and was not involved in the editing of the questions and has experience of recognising the risks of unintended bias.

Clearly different performance by ethnic group is an undesirable outcome, however it does not mean it is an incorrect or unfair outcome of the exam. The SQE1 exam can only measure achievement and functioning legal knowledge as presented on the day of the exam by each candidate. Each candidate will have received different levels of support, for example some benefiting from external training provider support, others not, and all building on prior experience, including differing levels of support received in their individual educational and socio-economic settings.

Whilst I am satisfied that all reasonable steps have been taken to avoid any form of bias or unfairness, I recommend Kaplan and the SRA keep investigating the issue of differential performance by ethnicity. Specifically, by continuing to:

- unlock the new found power of the large-scale data set which summarises the performance of questions and candidates' outcomes, including by different demographic groups
- seek understanding and explore new ideas for reducing significant outcome differences linked to any demographic
- build a team responsible for producing, editing and reviewing questions who are representative of the wider community and maintaining focus on this issue as the preparations for the first SQE2 exams are finalised, where human marking is used.

In summary, I must emphasise that overall, the way in which the individual questions and overall tests performed was very good from a technical assessment perspective and lessons learned from the SQE pilot have been applied very well.