Glass Houses: how might we decide on a ‘good enough’ assessment to become a solicitor?[[1]](#footnote-1)

Elaine Hall[[2]](#footnote-2), School of Law, Northumbria University, UK.

1. ***Introduction***

The development of the Solicitor’s Qualification Examination (SQE) is often attributed to the Legal Education Training Review (LETR), specifically to the section in the Executive Summary which says “*The current system of [legal services education and training] does not consistently ensure that desired levels of competence are reliably and demonstrably achieved.*”[[3]](#footnote-3). Other papers in this Special Issue will unpack the intention of the LETR team in making that statement, this one will focus on the SQE – what people hope and fear about it, what has been claimed for it and how it sits within our understanding of assessment and of assessment as evidence of legal education.

Last year I wrote a short paper for the Law Teacher about the proposed SQE, reporting on the consultation responses to the Solicitor’s Regulation Authority in response to some significant inconsistencies between the SRA press releases and the data[[4]](#footnote-4). In Table 1, I reproduce the elements from the consultation that relate directly to the SQE as ‘an independent professional assessment’[[5]](#footnote-5), ‘a robust and effective measure of competence’[[6]](#footnote-6) and ‘a suitable test of the requirements needed to become a solicitor’[[7]](#footnote-7) and to the people who support or resist the introduction of the SQE.

**Table 1: Summary of findings from the SRA Consultation data analysis (Hall, 2017).**

|  |  |  |
| --- | --- | --- |
| SRA Claim | Fact check | Indicative findings |
| *Support for an independent professional assessment in principle* | In principle question not asked in second consultation.First consultation question 1 is problematic.Even so, responses indicate that **only 20%** of respondents support the SQE. | There is no evidence from the consultation for the SRA’s claim that there is ‘in principle’ support from more than **one in** **five** respondents. |
| *‘Robust and effective’* | **68%** of respondents disagreed. They were more likely to strongly disagree (46%) than to disagree (22%). Only 22% agreed. | Strong rejection of the idea that a test ***yet to be designed*** can be described as robust and effective. |
| *‘Suitable test’* | **66%** of respondents disagreed. They were more likely to strongly disagree (43%) than to disagree (23%). 24% agreed. | Strong rejection of the idea that a test ***yet to be designed*** can be described as a suitable test of the requirements to become a solicitor. |
| *General support from individual solicitors and those about to qualify* | Around half of ‘aspiring solicitors’ and the majority of individual practitioners oppose the SQE. | When this is added to the majority of law firms rejecting the SQE, support from those in and about to join the profession is not evidenced. |
| *Clear resistance to the SQE, most consistently from academic institutions* | Clear resistance to the SQE is demonstrated by members of all groups of respondents apart from the 3 non-university training providers | Resistance may be strong amongst universities but as other groups are also resistant, it would be difficult to attribute resistance just to protectionism. |

It is clear therefore that most respondents to the consultation, not just Universities and the learned societies, share a belief that the format and focus of the SQE assessments may not be fit for purpose. It’s important to emphasise ‘may’: the SQE does not yet exist, we cannot test it until it comes into being.

1. **Exploring educational systems**[[8]](#footnote-8) **- is something wrong?**

The SQE is presented as a solution to a number of problems with legal education, graduates and employability. The support for and resistance to it are adjacent to but not entirely immersed in scientific enquiry and data. This section is an attempt to explain why that is without denigrating the intelligence or motivation of supporters or opponents and to hopefully introduce a frame for more positive discussion of this issue.

Charles Peirce, in his paper ‘The Fixation of Belief’ (1877), describes how a question emerges from a sense of dissatisfaction: an awareness either that we do not know something or that our existing belief about how the world is no longer a ‘good enough’ fit with our experience. This stimulus for discomfort could come from a singular event or from a gradual realisation. Peirce recognises that while real, flawed individuals do their best to navigate a complex universe in good faith, we naturally wish to hold to our existing and functional beliefs and will only be moved to doubt by genuine discomfort. When a belief becomes unsatisfactory, provoking discomfort and doubt then we want to return to a state of comfort and certainty as quickly and effortlessly as possible (in Figure 1). Peirce presents each of these methods of inquiry as necessary developmental steps: tenacity and authority are not stupid pre-steps to ‘proper inquiry’, to be skipped over if possible, they are critical indicators of the context in which these questions are being asked, as such they are as worthy of attention as any other stage.



Figure 1: Fixation of Belief and Methods of Inquiry - overview

In Figure 2, below[[9]](#footnote-9) the dimensions of competence as delineated by the Legal Education Training Review are set out. Although this was a systematic and rigorous synthesis of previous literature, it should be noted that it was consciously labelled by the LETR team as “non-exhaustive” and does not cover all of the elements of the ‘practice-ready’ discourse[[10]](#footnote-10). The hierarchy of importance might change according to context, nevertheless, it is a piece of common ground for legal educators, practitioners and regulators from which we can begin our investigations.



**Figure 2: The LETR non-exhaustive dimension s of competence**

Table 2 highlights some of the problems with law graduates (sometimes in general, sometimes specifically graduates of the LPC) identified by the SRA and by members of the profession[[11]](#footnote-11) and seeks to specify what kind of deficits these are under the LETR framework. In the discussion that follows I will be linking the cognitive and integrative deficits to the method and format of SQE1.

**Table 2: Summary of problems with the current system**

|  |  |
| --- | --- |
| Perceived risk of QLD as evidence for readiness to enter the profession(arguments put forward in support of centralised assessment) | Form of deficit |
| They don’t know the law | Cognitive |
| They don’t know how to apply the law | Cognitive/ Integrative |
| They are unprepared for the commercial and practical realities of practice | Context |
| They will provide a lower quality service to the public  | Affective/ Moral |
| They are more likely to breach professional regulations | Affective/Moral |

The method of tenacity might reject the premise entirely: there’s nothing wrong with the graduates, their preparation or the method of assessment, people who say so are mis-remembering their own neophyte levels of competence and charm in a pattern as old a written records[[12]](#footnote-12). However, this may not be entirely satisfactory since the complaints about graduates persist and not everyone who makes complaints can be dismissed so easily. The enquirer shifts to the method of authority, the appeal to institutional wisdom. The dominating question “*What’s wrong with?*” will inevitably be applied in turn to students (and their families and communities), individual law teachers, universities or training providers. Crucially, the question is reliant on established beliefs about each of these individuals or groups, not an investigation of them. This is where systems of belief about convergence and divergence begin to compete: providers of legal education for a qualifying award have historically had relative constraint over subject matter and relative freedom over pedagogy and assessment. This is simultaneously assessed by the convergence team as not rigorous (because not reliable) and by the divergence team as rigorous (because ecologically valid). As we’ll see in Section 4, it’s not really an either/or situation.

If these unsupported explanations fail to satisfy, an a priori approach could help us to separate what is *supposed to happen* from what *might be happening* and to make explicit the logical links we *think* *are* *there* between the elements in the situation and thus develop and interrogate possible hypotheses. This is normally the point where logic demands that we consider all the elements, especially the modes of assessment.

*Note: in this hypothesis, everything is good enough.*

Figure 3: Base Hypothesis (how the world is supposed to work logically)

If there is a lurking doubt, we take each element as potentially faulty in turn so *either*

Alternative Hypothesis A

*Or* Alternative HypothesisB

*Or* Alternative Hypothesis C

 *Note: this excludes Hypotheses D, E and F where two or more things go wrong at once.*

Figure 4: Alternate Hypotheses

The a priori method prepares the ground for the scientific enquiry, making it clear both what the underlying belief is and the range of possible explanations for the observed phenomenon. A robust enquiry would need to investigate

a) the past performance and qualities of the students,

b) the curriculum and the learning experience and

c) the nature of the assessment, its reliability and validity.

By subjecting cohort, pedagogy and assessment data to the same scrutiny as if it were research data, it becomes clear that all parties to this debate, including and perhaps especially university law teachers, live in heavily glazed structures. We simply have not done the scientific investigations; or if we have, we have not woven them into a coherent narrative. While this does not set us apart from other disciplinary academics, we are not justified in any rejection of proposals using the methods of tenacity or authority. The sections that follow will unpack some philosophical, technical and pedagogic frameworks that can help us to organise our thinking.

1. ***What is the philosophical threshold for evidence about a ‘good enough assessment’?***

Lawyers (particularly those in common law jurisdictions) have a disciplinary comprehension of the term ‘evidence’, that to the outside observer, appears grounded in the role that evidence plays in the adversarial legal encounter. The weight of evidence is not absolute but relative both to the evidence presented by opponents and by context: the level of certainty demanded by the court (balance of probability/ reasonable doubt). In this sense, evidence used in legal processes could be, in Achinstein’s[[13]](#footnote-13) terms (table 3 below), epistemic situation evidence or subjective evidence, without that being a threat to society or justice as we understand it in our own time. Reflecting on these categories does, however, account for our difficulty in fully appreciating the logic of actors in different epistemic situations or those who hold different subjective beliefs.

Table 3: Summary of Achinstein’s forms of evidence (H=hypothesis, E=evidence)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Epistemic situation evidence** | **Subjective evidence** | **Veridical evidence** | **Potential evidence** |
| **Description** | *Evidence that is understood within a particular cultural, historical or knowledge context* | *Evidence that is part of an individual or group belief structure* | *Evidence that transcends situations and beliefs* | *Evidence that is strongly related to experience, present and future* |
| **Requirement** | *That the inquirer could construct or maintain H based on the E within the limitations of their context* | *That the links between E and H are held to be true and consistent by the inquirer(s)* | *The data supporting E need to be objective, although not necessarily complete (conclusive)* | *The data supporting E must be objective and rigorous and are understood not to be conclusive* |
| **Limitation** | *This belief is not challenged by ideas from beyond the epistemic context* | *It is not necessary for any empirical elements to come into this inquiry* | *Both E and H need to be true (very hard to establish)* | *H may be false even where there is good E to support it* |
| **Link to beliefs** | *The inquirer was justified in believing H on this E, in context.* | *The inquirer(s) believe that E is evidence for H, that H is true, E does not have to be empirically true, provided that it is believed.* | *E is evidence for H and provides a good reason to believe H, since both E and H are true.* | *E is evidence for H and provides a good reason to believe H until other E emerges to challenge it.* |

Philosophers of science and those concerned with the public understanding of science have noted that our ‘lay understanding’ of evidence, particularly in relation to research, tends to align with the veridical definition where everything is true, objective and proven[[14]](#footnote-14). This makes good psychological sense from the perspective of Fixation of Belief[[15]](#footnote-15) but as Achinstein and Kvernbekk both discuss, veridical evidence is a very difficult standard, perhaps even an ideal state or benchmark, towards which potential evidence might take us. For this reason, it is helpful to consider John Dewey’s concept of ‘warranted assertibility’[[16]](#footnote-16), often referred to simply as ‘warrant’. Dewey is consciously positioning the enquirer as in a state of manageable uncertainty. In this state one has sufficient knowledge, grounded in enquiry and probability, to act without being trapped by the decision to act into rejecting all evidence against the idea:

*all knowledge, or warranted assertion, depends upon inquiry and that inquiry is, truistically, connected with what is questionable (and questioned) involves a sceptical element, or what Peirce called "fallibilism." But it also provides for probability, and for determination of degrees of probability in rejecting all intrinsically dogmatic statements, where "dogmatic" applies to any statement asserted to possess inherent self-evident truth.[[17]](#footnote-17)*

The concept of warrant as judgement resting upon an aggregate of all the available evidence takes the psychological pressure from each piece of evidence. The separation warrant from evidence enables each piece of evidence to be examined using the criteria from Achinstein’s other categories, considering its’ fallibility and its context, whereas elision of warrant with evidence requires evidence to be at least weakly veridical. If each piece of evidence has to be this strong, to achieve such high standards, what kinds of data might be privileged? By seeking veridical evidence, are we likely to seek convergence and centralisation? In the next section, I return to the present issue of the introduction of the SQE and to a discussion of whether what the SRA hope to achieve is possible.

1. ***Tools for measurement – what might give a ‘super-exam’ some ‘super’ qualities?[[18]](#footnote-18)***

If the problem as identified by the SRA is not-good-enough graduates and we would like something simple, clean and rigorous to solve this problem in one action[[19]](#footnote-19), then having a centralised super-exam is posited as an appropriate solution. In this section I will consider what the SQE 1 would actually have to be like in order for it to fulfil the implied remit, to safeguard the public from graduates who lack the cognitive and integrative skills required by the LETR (Figure 2, above). The currently available information[[20]](#footnote-20) suggests that SQE1 will be a gateway test to other, more complex and nuanced assessments in SQE2. This means that in that first stage of examination threshold competence is being assessed in order to winnow out those who are not able to demonstrate the knowledge, skills and other attributes tested, so it is reasonable to expect that the SQE1 will conform to standard conventions of test construction. Test design consists of different forms of validity and reliability and the technical distinctions between them are more than semantic. In tables 4 and 5 the purpose of each kind of validity and reliability is explained[[21]](#footnote-21).

**Table 4: Foundational elements of test construction**

|  |
| --- |
| **Foundational elements**(necessary in order to support basic claims to be a ***test***) |
| construct validity | how far test scores can be interpreted as measuring only what they are intended to measure | *Tests of specific knowledge, for example, should not allow success via other attributes, such as general knowledge or test-taking technique* |
| ecological validity | the quality of being well grounded in the reality of a particular context | *Tests of professional knowledge should be realistic and reflect the practice in the field* |
| reliability | An umbrella term for the coherence (internal consistency) of a set of test items, and/or the stability (test–retest) of a set of test scores over time |
| internal consistency (reliability) | the degree to which the items in a test measure the same thing, measured by the average correlation between each item and the other items | *For example, all the questions which contribute to the civil process scores are about civil process, not about a particular civil case. Someone with a good enough knowledge of civil process would be likely to get most of them right* |
| test–retest reliability | the stability of test scores as indicated by retesting the same group and calculating a correlation coefficient using the two sets of scores | *Performance by a good enough student would not significantly vary on different versions of the test: that student would do about as well in 2022 or 2023.* |
| ***Although these elements are separate, they are mutually interdependent to meet the quality threshold.*** |

We need to be clear that the foundational elements all have to be considered, robustly tested and a discussion of the extent to which they are considered to be ‘good enough’ must be in place before any claims about the quality(ies) of a test can be made – indeed for it to be correctly identified as a test. The two foundational validities, together with reliability must all be considered.[[22]](#footnote-22) Only then can additional or specialist elements be considered (table 5 below).

**Table 5: Additional and Specialist elements of test construction**

|  |
| --- |
| **Additional elements**(necessary in order to support ***comparative*** claims about the ***test***) |
| concurrent validity | support for the meaning of a construct or the value of a test, based on correlational evidence from another set of measurements taken at the same time (also *external validity* where two tests intended to capture the same attribute are used concurrently) | *In order to make claims that something is a ‘better test’, comparative research has to be done on both measures and triangulated with an independent phenomenon.[[23]](#footnote-23)* |
|  |
| **Specialist elements**(necessary in order to support claims to ***specific utility***) |
| catalytic validity | the extent to which those involved in testing become motivated to understand and transform the situations in which they operate | *Since taking any test impacts on the learner’s perception of the world, assessments can promote or retard particular behaviours (cf.”I don’t want to know it if it’s not on the exam”)* |
| predictive validity | the extent to which a set of scores predicts an expected outcome or criterion assessed through future performance | *Baseline assessments of whole cohorts can be analysed against data from future assessments of whole cohorts (e.g. core modules). Disruptions to the cohort by choice (choosing electives) or by winnowing (gateway assessments) mean that only certain members of the original cohort are available for comparison. It might be possible to excel at SQE2 or in practice without passing SQE1 but that data would never be available.* |

Of course, we don’t yet know what SQE1 will consist of but Julie Brennan from the SRA has written a blog in support of multiple choice questions (MCQs)[[24]](#footnote-24). She acknowledges that MCQs may only test recall and advances the virtues of clear vignettes to set the parameters of the knowledge and application being tested, using as her benchmark:

 …*the type of questions used in the US Multi-State. What is striking how well these map on to the competences in the Solicitors Competence Statement which we might assess through the proposed SQE Part 1.*

She does not specify exactly how this mapping works but helpfully includes Table 6 below[[25]](#footnote-25) from which we can work out our own connections.

**Table 6: Juxtaposition of Multi-State questions with Solicitors’ Competence Statements**

|  |  |
| --- | --- |
| **MCT questions which have appeared in the US Multistate Bar Exam** | **Competences in the Solicitors' Competence Statement tested** |
| Gathering information* What issue must the lawyer resolve before advising his client?
* What information does the lawyer need, in order to proceed?
* What question should the lawyer ask next?
* Which source of information will be the most helpful?

  | A4: identification and application of legal principlesA5: Applying understanding, critical thinking and analysis to solve problems, including recognizing gaps in information and evaluating the quality and reliability of information.B1 (c): recognizing when additional information is needed.B3: developing and advising on relevant options, strategies and solutionsB7(a): applying relevant processes and procedures to progress the matter |
| Identifying issues and formulating strategy:* Which of the following causes of action is most likely to be successful?
* What is the best defence?
* Whom should your client sue?
* What test should the judge apply?
* Which law governs the claim?
 |

MCQs of the Multi-State Bar exam model have four possible answers to choose from. They don’t really test simple recall, as the answer is somewhere on the page. The skill of identification from A4 is therefore limited to discrimination amongst what is offered, rather than the ability to identify and apply a legal principle without prompts. Similarly, recognising gaps (A5) and devising strategies (B3) are only tested in this limited way. These are all tests of recognition and as such, vulnerable to particular patterns of error in recording competence: a study of a similar test found that when the questions were excluded and students made a ‘best guess’ based only the answers available they were correct 52% of the time (instead of the 25% of the time that would be expected statistically)[[26]](#footnote-26). Therefore, although the internal consistency of the items of the Multi-State appear to be robust[[27]](#footnote-27), it is possible that there is a construct validity problem and what is being tested is not legal knowledge but also the ability to spot the format of the correct answer. After all, the answers to most realistic (and therefore ecologically valid) legal questions have degrees of ambiguity; there are fuzzy boundaries of interpretation about strategy, client’s best interests and the iterative development of the law. It is therefore very difficult to construct an MCQ that is more complex than the recall of technical facts and simultaneously has:

* robust construct validity (is about the thing that it aims to test)
* is rigorous (not too easy to guess) and
* fair (the difference between the correct answer and the red herrings is discernible through the use of the knowledge/ skill being tested).

In the pursuit of fairness, rigour is likely to suffer. The MCQ is a good tool for some things but as a single threshold measure is arguably carrying too much weight. There is a broad consensus for a diverse assessment for professional education and SQE as a whole promises a mixed diet but we might heed the warnings from medical education: consecutive and consequential assessment regimes risk an atomisation of competencies which has the capacity to “trivialise content and threaten validity”[[28]](#footnote-28) and given the limitations of every single measure we would be wiser to focus on the relationship between assessments[[29]](#footnote-29). This is the focus of the next section.

1. ***A discussion of purpose and assessment***

*“Assessment is the senior partner in learning and teaching. Get it wrong and the rest collapses.”****[[30]](#footnote-30)***

Assessment drives what is taught and how in all forms of education[[31]](#footnote-31) and in professional education the framing and validation of assessment is the responsibility of the regulator. Again, the message from medical education is clear:

*A very basic question in assessment is the principle of fitness for purpose. So what is the purpose of assessment for regulators? Naturally that purpose is good health care delivered by a competent workforce. But how do we achieve a competent workforce? It can happen only if the individuals in the workforce keep learning. In my view, the assurance of lifelong learning is the prime aim for which a regulator should strive. So the issue here is to develop assessment strategies that help learning. The next purpose for the regulator is to guarantee patient safety by safeguarding the public from incompetent individuals in the workforce. These two purposes should be separated, even firewalled, and treated differently in developing an assessment strategy.[[32]](#footnote-32)*

Thus, the responsibility of the regulator is not all-encompassing[[33]](#footnote-33) and we should not be engaged in a pitched battle for overall dominance of legal education. Arguably, the ideal is a situation in which all interested parties (universities, other legal trainers, employers, regulators and individual professionals) take on elements of disciplinary preparation and lifelong learning that are appropriate to their interests and resources. We can have specific debates about which elements are fundamental to client protection and standards in the profession and so must form part of consequential assessments. We can offer collegial support and challenge to the evidence each of us presents that we have transferred and robustly assessed each element. This kind of collaboration rests upon a framework of learning as a process involving the iterative use of knowledge and skills (figure 5) in which replication underpins application which underpins interpretation and then association.



**Figure 5: Broudy’s four stage model of knowledge use** [[34]](#footnote-34)

When a student first meets foundational knowledge, it is important to assess their basic mastery of it. It is unfair to expect me to correctly apply a civil procedure rule or a perform basic skill like reading and analysing a case if I have not internalised its existence (knowing-that A) and its plain text meaning (knowing-that B). A replication test might address one or both of those aspects: depending on the format of the assessment it might provide evidence of my ability to pick the rule out of an MCQ line-up or it might provide evidence of my ability to provide in oral or written form a rich description of the rule from which a recognisable facial composite picture is generated. An application test can also be differentiated: a singular solution applying a dominant pattern as most MCQs require (knowing-how A) or the more complex delineation of competing possibilities within a bounded example, ‘showing the working’ of the student’s decision to choose a particular route (knowing-how B). Interpretation tests should provide students with the opportunity to richly describe the intended and actual use of rules (whether from a doctrinal or socio-legal perspective), the factors guiding change, variety or evolution (knowing-how C). Association tests need to give students space to make the cultural disciplinary links, whether these are recognition of overarching philosophical principles (knowing-how-we A); patterns of similarity and difference in sub-disciplines (knowing-how-we B: for example the importance/unimportance of intent in crime and contract law) or key disciplinary skills (knowing-how-we C: for example why lawyers use the case method and what claims to knowledge this allows us). As, this unpacking of the various aspects reveals, for the purposes of assessment item design, it is best to focus on one aspect at a time.

However, the progression through Broudy’s model is not simply linear, as Eraut identifies when discussing professional expertise *“The process of using knowledge transforms that knowledge so that it is no longer the same knowledge”*[[35]](#footnote-35)***.*** Thus, at each stage it is possible to critically and reflexively examine the knowledge gained at the previous iteration. In learning to apply the rule in the classroom, I may critique its utility, on returning to the application of the rule in practice I may notice the variability of its application and this may then inform my interpretation of decisions and precedents. Over time, I may adopt a more complex strategic understanding of the rule or set out on a process of reform to replace it with less ambiguous one. In terms of programmatic assessment and curriculum design, we need to give students multiple opportunities to alight on each aspect, recognising that we all forget foundational knowledge and need to be reminded when next it is needed[[36]](#footnote-36) and that the return, if framed as an opportunity for interpretation and association is an enriching experience for the learner.

1. ***Working together to understand practice readiness, legal education and the warrant of all our assessments***

In this paper, I have focused on the methods that can be used for producing consistent, reliable and demonstrable measures of competence and consider what we already know about assessment in legal education and in other forms of professional education. I am writing from the perspective of an academic with a career-long interest in measures and assessment and though I am critical of the SQE proposals I am also not blind to the limitations of assessment practices in Universities. More broadly, I am concerned that in reifying or denigrating a ‘super-exam’, we are all losing sight of the limitations of all assessments: since each one is a tool which is well-adapted to a particular purpose, not a universal panacea. As I have been preparing this paper, Kaplan have been appointed to develop the SQE[[37]](#footnote-37) and I would like to offer this paper, its language and frameworks as part of a collegial conversation about what we can all reasonably expect that the SQE can and cannot do. Our shared goal is a competent, flexible and robust legal profession and we can be clear about our different contributions to the development and measurement of this, provided that we look critically at what ‘fit for purpose’ would look like and what the threshold for evidence of fitness would be. The final table (table 7 below) is by way of an invitation – what contribution could each of us make to populating the fields and producing an evidence-informed picture of legal education?

**Table 7: Evidence about legal education**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Perceived risk of QLD as evidence for readiness(arguments put forward in support of centralised assessment) | *Empirical evidence to support these claims* | *Ways in which University assessments are supposed to address this problem* | *Logical proposition that the current assessments are capable of producing robust evidence of addressing the problem* | *Empirical evidence to support this proposition*  | *Ways in which SQE1 and 2 are supposed to address this* | *Logical proposition that the tools proposed are capable of producing robust evidence of addressing the problem* | *Empirical evidence to support this proposition*  |
| They don’t know the law |  |  |  |  |  |  |  |
| They don’t know how to apply the law |  |  |  |  |  |  |  |
| They are unprepared for practice |  |  |  |  |  |  |  |
| They will provide a lower quality service to the public  |  |  |  |  |  |  |  |
| They are more likely to breach professional regulations |  |  |  |  |  |  |  |

1. A version of this paper, developed with Cath Sylvester- “Hey, greenhouse, you’re looking really glassy” says Crystal Palace: justifying our robust alternatives to simplistic testing- was presented at the ‘LETR- Five Years On’ Conference hosted by Leeds Beckett University in July 2018. My thanks to participants in the session for their feedback and questions which have contributed to the areas of clarity in the final piece. Readers will note my profound debt to Cath’s paper *Through a glass darkly: Assessment of a real client, compulsory clinic in an undergraduate law programme.* International Journal of Clinical Legal Education, 2016, 23(1), 32. [↑](#footnote-ref-1)
2. Elaine Hall is Professor of Legal Education at Northumbria University. [↑](#footnote-ref-2)
3. <http://letr.org.uk/the-report/executive-summary/executive-summary-english/index.html> [↑](#footnote-ref-3)
4. Hall, E. (2017) Notes on the SRA report of the consultation on the Solicitors Qualifying Exam: ‘Comment is free but facts are sacred’ *The Law Teacher, 41, 1.;* <http://www.sra.org.uk/sra/consultations/solicitors-qualifying-examination.page> *Consultation response and our decision on next steps* [↑](#footnote-ref-4)
5. Question 1, Consultation 1- ‘Training for Tomorrow’ (published in October 2016) [↑](#footnote-ref-5)
6. Question 1, Consultation 2- ‘A New Route to Qualification’ (published in March 2017) [↑](#footnote-ref-6)
7. Question 4, Consultation 2- ‘A New Route to Qualification’ (published in March 2017) [↑](#footnote-ref-7)
8. This section is adapted from a longer discussion of the fixation of belief and the construction of educational research questions which appears in Chapter 2 of Hall, E and Wall, K. (2019, in press) *Research Methods for Understanding Professional Learning.* London, Bloomsbury [↑](#footnote-ref-8)
9. Webb, J., et al., Setting Standards: The Future of Legal Services Education and Training Regulation in

England and Wales p.118 (The Legal Education and Training Review (LETR) (2013)). Available at:

<http://letr.org.uk/the-report/index.html> [↑](#footnote-ref-9)
10. A thorough review of the academic, regulatory and employer discourse on practice readiness is available in Rachel Dunn’s 2018 doctoral thesis The Knowledge, Skills And Attributes Considered Necessary To Start Day One Training Competently And Whether Live Client Clinics Develop Them, available through <http://nrl.northumbria.ac.uk/> [↑](#footnote-ref-10)
11. Drawn from a non-systematic scoping review of press reports, online comments and blogs. [↑](#footnote-ref-11)
12. *The children now love luxury; they have bad manners, contempt for authority; they show disrespect for elders and love chatter in place of exercise. Children are now tyrants, not the servants of their households. They no longer rise when elders enter the room. They contradict their parents, chatter before company, gobble up dainties at the table, cross their legs, and tyrannize their teachers*. Attributed to Socrates [↑](#footnote-ref-12)
13. Peter Achinstein, The Book of Evidence (Oxford: Oxford University Press, 2001) This table is a summary of Tone Kvernbekk’s 2011 discussion of Achinstein in ‘The concept of evidence in evidence-based practice’ *Educational Theory* 61, 5, pp.515-532 developed by Elaine Hall for the Association of Law Teachers 2016 conference keynote: *‘The Use of Evidence in Writing About Legal Education’* [↑](#footnote-ref-13)
14. Twining, W. *Rethinking Legal Education*. Lord Upjohn Lecture 2017, published online in the Law Teacher <https://doi.org/10.1080/03069400.2018.1497260> [↑](#footnote-ref-14)
15. Peirce, op. cit. [↑](#footnote-ref-15)
16. Dewey, J. (1941) Propositions, Warranted Assertibility, and Truth. The Journal of Philosophy, 38, 7, pp. 169-186 [↑](#footnote-ref-16)
17. Ibid., p174 [↑](#footnote-ref-17)
18. This section is a response to a difference of opinion with Julie Brennan at the LETR 5 Years On event. She claimed that the SQE would have predictive validity. I said that it would not. She replied that it would, because the SRA had promised that it would. I said that it doesn’t work like that – this is where I unpack what I meant. [↑](#footnote-ref-18)
19. Even if we don’t agree about the nature of the problem, or that there is robust evidence that the problem exists as stated, or that it can be solved with one intervention. [↑](#footnote-ref-19)
20. <http://www.sra.org.uk/sra/consultations/solicitors-qualifying-examination.page> *Consultation response and our decision on next steps* [↑](#footnote-ref-20)
21. Definitions for both tables are drawn from the glossary of Coffield, F., Moseley, D., Hall, E and Ecclestone, K. (2007*) Learning styles and pedagogy in post-16 learning: a systematic and critical review* London: Learning and Skills Research Centre. [↑](#footnote-ref-21)
22. Despite Meat Loaf’s belief that two out of three ain’t bad. [↑](#footnote-ref-22)
23. Extreme caution is urged here as there are very few successful examples. Cees van der Vleuten remarks *“The seminal discovery in assessment was that performance on one problem-solving exercise – the Patient Management Problem of the 1960s – was hardly predictive of performance on another exercise, a persistent finding that turned out to be shared across assessment methods”* Medical Education 2014; 48: 234 [↑](#footnote-ref-23)
24. <http://www.sra.org.uk/sra/policy/training-for-tomorrow/sqe-blog/using-multiple-choice-questions-in-legal-education.page> [↑](#footnote-ref-24)
25. Ibid, n supra. [↑](#footnote-ref-25)
26. Albright, D.L. and Thyer, B.A., 2010. A Test of the Validity of the LCSW Examination: Quis Custodiet Ipsos Custodes? Social work research, 34(4), pp. 229-234. [↑](#footnote-ref-26)
27. Yin, P. (2005). A Multivariate Generalizability Analysis of the Multistate Bar Examination. Educational and Psychological Measurement, 65(4), 668-686. [↑](#footnote-ref-27)
28. C. Van Der Vleuten, L W T Schuwirth, *Assessing professional competence: from methods to programmes* 2005, Medical Education p39 A full discussion of the relevance of programmatic assessment to the law degree can be found in Sylvester, C. (2016). Through a glass darkly: Assessment of a real client, compulsory clinic in an undergraduate law programme. International Journal of Clinical Legal Education, 23(1), 32. [↑](#footnote-ref-28)
29. *“Any assessment, old or new, objective or subjective, standardised or unstandardised, requires at least 3–4 hours of testing time to achieve minimal reliability. Even with a reliability criterion of 0.80, we should realise that 20% of the pass/fail decisions we make may be false positives and negatives”* C. Van Der Vleuten *Revisiting ‘Assessing professional competence: from methods to programmes’* Medical Education 2016 50: 885 [↑](#footnote-ref-29)
30. John Biggs and Catherine Tang, *Teaching for Quality Learning at University*, 4th Edition, Society for Research into Higher Education and Open University Press, 2011 [↑](#footnote-ref-30)
31. Bernstein, B. (1996) Pedagogy, symbolic control and identity, theory, research, critique London: Taylor and Francis [↑](#footnote-ref-31)
32. C. van der Vleuten, p.216 in *Dialogue: A conversation about the role of medical regulators between Lesley Southgate & Cees P M van der Vleuten* Medical Education 2014; 48: 215–218 [↑](#footnote-ref-32)
33. As Simon Bullock implied in 2014 <http://www.sra.org.uk/sra/policy/training-for-tomorrow/sqe-blog/bonfire-of-the-regulations.page> [↑](#footnote-ref-33)
34. Broudy, H.S. personal communication, in Eraut, M. (1994) *Developing Professional Knowledge and Competence* London, Falmer Press [↑](#footnote-ref-34)
35. Eraut, M. op. cit. p25 [↑](#footnote-ref-35)
36. Any complicated technical task done infrequently – taxes, expenses claims, remembering where the manual for something that has never broken before was filed – is vulnerable to this. [↑](#footnote-ref-36)
37. <http://www.sra.org.uk/sra/news/press/sqe-assessment-organisation-appointed.page> [↑](#footnote-ref-37)