The Case for Integrating Multi-Source Data for a Fairer and Holistic Judgement of Competence in Medical Education & Training

Tackling Differential Attainment in Medical Professions - Bridging the Gap Workshop Series 2020

Abstract

Being a doctor in the 21st Century requires a diverse range of skills, a broad base of knowledge and a suite of professional values and attitudes that enable clinical practice to be safe, effective and caring. Doctors, irrespective of their specialty, need to be knowledgeable and skilful not just in their area of expertise, but also need a range of generic skills and capabilities such as communication, leadership, academic scholarship and research, teaching, quality improvement, advocacy, digital literacy to name a few. These capabilities, all relevant to clinical practice, are assessed routinely in clinical settings. This rich information about trainees, available from their formative assessments, does not inform high-stakes judgements about progression. Instead, these judgements are usually made on the basis of summative examinations conducted in simulated settings.

Unfortunately, these summative assessments have consistently delivered results with a large magnitude of differential between the outcomes of candidates, based on factors such as ethnicity, gender, other protected characteristics and also country of primary medical qualification. Formative assessment during training, however is individualised and tends not show this level of difference; leading to a situation where failure in summative examinations comes as a surprise to both trainees and to training programme directors.

There is evidence that periodic assessment of trainees' acquisition of core capabilities can help make balanced, informed judgements about readiness for progression. The move from a pass/fail categorisation to a yet/not yet categorisation when coupled with appropriate remedial measures can improve, both the validity, as well as fairness of assessments.

The large magnitude of differential in outcomes of high-stakes assessments cannot be fixed by tweaking current assessment systems. Instead, there needs to be a recognition that high-level of capabilities consistently demonstrated in the workplace need to play a role in judgements about progression. Failure to do so is unfair, wasteful of public finances, and in breach of the trust places by the public, in training safe and competent clinicians.

Keywords Differential attainment, medical professionals, summative assessments, formative assessments

Subodh Dave, Roshelle Ramkisson, Chelliah Selvasekar, Indranil Chakravorty
Bapio Institute for Health Research, Bedford, UK

Correspondence to
subodhdave@gmail.com

Introduction

Assessments in medical education and training, from the perspective of the public, have a critical role in identifying doctors that are capable of delivering safe, effective care in a wide range of settings. There is a public expectation, implemented by the regulator (e.g. General Medical Council) embedded in the appraisal and revalidation process, that doctors will demonstrate continuous and life-long learning to develop new and/ or maintain their professional knowledge, skills, behaviours and attitudes. (1) Assessment of these diverse skills is difficult if not impossible in a simulated setting. (2)

Currently, the decision about an individual doctor's capability for either entering specialist training or completion of training thus entering independent practice as a senior, is usually based on success in summative examinations. Getting this decision correct, is important not only for the individual, but also in retaining public confidence in the healthcare education and training system. The lifetime cost (to the taxpayer) of training a doctor in the UK in 2017 is estimated as GB £230,000, and each wrong decision can cost up to £80,000/year (3) in additional costs, which comprise extra-time to retrain and extended-remedial skills training. Section 35C(2) of the Medical Act 1983 as amended, states that a doctor's fitness to practise can be impaired by deficient professional performance. (4). If summative assessments are a reliable and valid measure of capability/ competence/ performance, then allowing a doctor (who is unsuccessful in summative assessments), thus unable to demonstrate appropriate capability to practice, in non-training roles with limited supervision, may theoretically pose an increased risk to patients.

Despite the high stakes, consistent differences in attainment have been reported affecting some groups of doctors in medical examinations based on factors outside of their capability viz. factors such as race, demographics, social, cultural or deprivation. The country of primary medical qualification (PMQ) and ethnicity have significant effect sizes in analyses of outcomes. International medical graduates (IMG), doctors from Black or minority ethnic origin, face a double whammy of the negative effect on performance. (5,6)

The personal impact of DA is high for learners resulting in distress, mental and physical ill-health and in some cases with tragic consequences. It is worth bearing in mind, that for many, the attainment gap amplifies the microaggressions that some doctors face in clinical and educational settings.(7) This has been expressed in several personal narratives received by professional organisations, such as the British Association of Physicians of Indian Origin (BAPIO), where exam related stress has been specifically linked to personal and professional challenges. A consistent theme that emerges from the testimonies of both trainees and trainers is the "surprise element" of the results of the final examinations. Trainees who have been receiving good feedback from their clinical supervisors, often who are rated as being 'excellent clinicians' by team members and by patients, seem to fail the exams much to everyone's surprise.

The reasons for DA in summative examinations are multifarious and these have been explored in detail in our paper on the issue.(8) The focus of this paper is on two interrelated issues, in relation to being a safe and competent doctor; (i) making valid judgements about competence, and (ii) addressing the differential impact of the reliance on snapshot summative assessments of competence.

This treatise on DA, will argue that assessment strategies that rely solely on single, high stakes, summative assessment, undertaken in artificial or simulated clinical environments such as in written or clinical examinations (e.g. OSCEs), which have consistently demonstrated an unfair phenomenon of DA, may fail doctors who are otherwise judged by their supervisors and being competent in the real clinical world and entrusted with the care of patients. The paper will also argue that relying on summative assessment alone to make high-stakes progression decisions, goes against the fundamental principles of equality. Such reliance on snapshot assessments lacks validity and goes against the principles of lifelong learning. Thus arguing in favour of integrating either continuous or multiple low stakes assessments, over the length of training and clinical practice. This exercise is but one component of the comprehensive review of DA across the many domains in the lifecycle of a medical professional.

Formative Assessment

Assessment is the use of a set of procedures to collect information about learning. The primary aim of assessment should be *for learning* (9) and indeed assessment should be central to the learning experience (10). However, in practice, assessment is often equated with examination as a process that *follows* learning. Historically, summative assessments have been viewed as a process that provides information to judge the educational value and success of a training program. Formative assessments in contrast, have been viewed as a process that provides information to facilitate improvements in the training programme.

Formative assessment is defined as 'all activities are undertaken by teachers, and/or their students to modify teaching and learning activities in which they [the students] are engaged'. Focusing on the learner, summative assessments focus on what the learners have learnt (*assessment of learning)* while formative assessments focus on what learners need to learn (*assessment for learning*; (11). The categorical distinction between these approaches is now being replaced by a blended or hybrid approach that views both formative and summative assessments as part of a more comprehensive system to facilitate learning. Specific and meaningful feedback to learners following summative examinations can be helpful in improving and altering professional practice, while formative assessments can help both educators and learners gauge the progress made in achieving desired learning outcomes(12). Indeed, this approach of utilising multiple assessment points (13) to inform learning progression through the training programme, has now been adopted by several medical schools (14). This not only helps learners monitor their progress but also helps educators elicit the educational impact of the training programme.

This is of particular importance where programmes are known to be vulnerable to differential outcomes. Regular monitoring of progress is then not merely desirable but essential to demonstrate that training programmes are; a) adding value and more crucially, b) not systematically disadvantaging specific groups of trainees. The finding that learners from Black and ethnic minority backgrounds are significantly less likely than their White peers (50% vs 70%), to obtain a higher second or a first-class degree, having started with comparable pre-course attainment, is a reminder of the importance of careful monitoring of progress to enable early remedial action. (15)

Continuous Learning

Educationally, there is the recognition that learning is a lifelong exercise and continuous professional development (CPD) is essential in the sphere of professional/clinical practice and thus upholding evidence-based care for patients. The UK General Medical Council's (GMC) Generic Professional Capabilities (GPC) Framework (16) has signalled a change from competency-based learning to continuous learning, centred around a framework of high-level outcomes (HLOs). The move, from learning focused on acquiring discrete competencies, to capabilities, is a significant shift to the acquisition of knowledge and skills, through real-world *authentic* tasks, preparing for independent clinical practice. The shift in emphasis in the GPC framework from "*shows how*" to "*does*", also indicates the need for a shift in assessment framework from mere demonstration of skills and knowledge in an artificial (albeit simulated) setting, to the demonstration of the *capability* of applying such knowledge and skills in real-world settings.

Black and ethnic minority trainees often experience the negative experience of learning and rather more likely to experience cultural discrimination that may contribute to the DA. (17) IMGs coming to work in the NHS may have little or no acculturation or avenues for adaptation to working in the UK. Moreover, supervisors do not yet co-design individualised learning plans with their trainees. Early formative assessments are vital to shaping an individualised learning plan taking into account the trainee's strengths and learning needs.

Competence and Capability

The primary purpose of assessment in medical training is to discriminate those who are capable of delivering clinical care, that meets defined professional standards (18), from those who are not. Regulators like the UK GMC (19) (16) have the responsibility to define standards, professional practice, and of the learning environment as well as the culture that supports learning. GPCs in particular transform

responsibilities identified in GMP, to learning outcomes that can be incorporated in specialty curricula. Capability is more than competence; Competence—what individuals know or are able to do in terms of knowledge, skills, attitude and Capability—the extent to which individuals can adapt to change, generate new knowledge, and continue to improve their performance. (20)

Clearly, doctors practising clinically in a healthcare sector/ country (i.e. UK) need to do so in consonance with the principles enshrined in the code of professional standards (i.e. UK GMC's Good Medical Practice). There is the awareness that the "recognition of the ethical, legal and cultural context of (UK) health care does not actually happen until doctors are working in practice".(21) Having moved from one medical jurisdiction to another, there is often little support for IMGs in terms of their professional practice apart from a copy of the Good Medical Practice (GMP) that they are given on registration. Since the publication of the report, Welcome to UK Practice, has been introduced by the GMC and this training seems to have improved awareness and understanding. (22) However, there appears to be a decay in trainees' understanding of the application of GMP guidance in clinical practice.

Assessment of Capabilities = Assessment of Continuous Learning

Professional expertise in the post-GPC curricula clearly includes technical knowledge and skills needed to assess and examine patients, to make appropriate diagnoses and to deliver evidence-based patient care. They also include a range of professional knowledge, skills, attitudes and behaviours concerning a diverse range of capabilities including health promotion, advocacy, safeguarding, digital literacy, leadership skills, communication and interpersonal skills, team working, quality improvement, academic research, training and education. The requirement that those completing training must demonstrate the acquisition of capabilities in these diverse domains, poses a challenge for assessments.

Even a cursory examination of the domains affirms that evaluating learner proficiency in such a wide range of capabilities necessitates the use of more than a single assessment tool. The introduction of formative assessment tools such as workplace-based assessments and the move in summative examinations (such as MRCP and

MRCPsych) to domain-based assessments (from solely competency-based assessments) is demonstrative of this paradigm shift.

This move towards improving the *validity* of postgraduate assessments follows a similar reform in the assessments used for selecting candidates for medical training. In recognition of the fact that '*good*' doctors need a diverse range of qualities, assessments for selection have moved from knowledge-based tools such as Medical College Admission Test to the inclusion of tools such as Situational Judgement Test (SJT) and more recently to the use of Multiple Mini Interviews (MMIs).(23) This format allows for the assessment of verbal and non-verbal skills, attitudes, aptitude and even personality traits, that many medical schools find more reliable and valid in screening students for admission to medical schools. (24)

In training programmes at both undergraduate and postgraduate levels, the application of the principle of lifelong learning leads to the acknowledgement that knowledge and skills demonstrate incremental growth with training. Competency-based spiral curricula are now widely used in training programmes to support (and to demonstrate) the acquisition of the same competency at a higher level of proficiency. However, the measurement of such growth necessitates frequent point-in-time assessments. Such "progress testing" allows the evaluation of improvement in trainees' capabilities over time (25). Blueprinting specific capabilities to particular assessment methods allows for "tracking" of progress against defined curricular outcomes (13,26).

Assessment in this context then shifts from a pass/fail categorisation to a *yet/not yet* categorisation. These developments are not new. Competency-based medical education (and assessment) was introduced in Northern America earlier this century. In the UK, ARCP- Annual Review of Competency Progression (ARCP) replaced the Record of In-Teaching Assessment (RITA) with the notion that reporting and outcome of formative assessments would inform judgements about progression.

However, concerns about standardisation and reliability of assessment tools have led to the continuation of reliance on the assessment of these competencies in controlled environments such as OSCEs (Objective Structured Clinical Examinations) where standardisation and

reliability can be assured. But, performance in such controlled environments does not predict performance in the real clinical world, raising questions about the *validity* of these tests.(27) There is also recognition that clinical performance is contextual and that performance in a controlled environment, does not accurately predict performance in the complex clinical environment, and it is, therefore, important to assess trainees in the clinical environment while they are conducting their daily activities. (28) Workplace-based Assessments (WPBAs) were born out of the need to assess trainees, while they conduct their routine clinical practice.

What implications does this have for sub-groups of trainees that are known to be subject to differential outcomes?

Evidence from studies of DA in higher education tell us that BAME students are less likely to feel supported by their trainers and report less faith in assessment tools. (15)These factors need to be taken into account when selecting formative assessment tools and considering their implementation in the workplace. IMGs may be unfamiliar with formative assessment tools and indeed with the skills such as reflective practice needed to engage with such assessments. (29) Acculturation to such assessment tools should be an iterative on-going process rather than through a one-off induction.

Current Formative Assessment Tools

A range of tools for workplace based formative assessment have been developed- many of these have been adopted in the UK and are still in use - viz. Mini-CEX (mini Clinical Examination), DOPS (Direct Observation of Procedural Skills), ACE (Assessment of Clinical Expertise) etc.
These WPBAs (WorkPlace Based Assessments), based on Dreyfuss' model of development (30) offer the promise of progressive assessments that could track progressive educational milestones leading to the acquisition of a greater level of expertise. Dreyfuss' model provides a helpful framework especially for IMGs, who are more likely to have high level of expertise in some curricular areas on account of their experience abroad, and yet might be at novice level in many others, on account of the move to a different medical jurisdiction. The model also offers the opportunity to apply a strengths focused approach rather than the deficit model, which can lead to a self-fulfilling prophecy of failure. (29)

Challenges with current Formative Assessment tools
A critique of WPBAs is beyond the scope of this paper but a few difficulties have emerged.

WPBA assessor factors

Data of mean WPBA scores suggests clustering of scores around scores of 4, 5 and 6 on a 6-point scale, creating a halo effect. This suggests a tendency on the part of the rater to award a constant rating across all items reducing the discriminant quality of the tool. (31) A well-known criticism of all forms of formative feedback tools is educators' awkwardness in giving negative feedback especially to those personally known to them, which may lead to 'grade inflation'. Given that WPBAs, a formative tool, have been used in the UK to inform the ARCP outcome, (a summative process) it is not surprising that both supervisors and supervisees tend to do a series of WPBAs in a batch, to meet the threshold needed. Poor interrater reliability has also been reported.

The lack of protected supervision time for clinicians, means that service demands intrude upon and often trump educational activity. One of the casualties is the lack of direct observation of trainee's clinical activity.

WPBA assessee factors

Trainees vary in their willingness to engage with the learning process, but as adult learners, ultimately, trainees are responsible for their own learning. Trainees in difficulty are more likely to choose a non-consultant supervisor, and are less likely to seek active feedback. Supervised Learning Events (SLE), a trainee-led reflection-based formative assessment tool, has been helpful in identifying "trainees in difficulty", which may be a reference to those trainees who are more at the novice end of the spectrum, with regards to certain learning outcomes.
Clearly, there are systemic barriers to the realisation of the full potential of formative assessment tools in the workplace - lack of appropriate training for trainers, lack of buy-in from trainees and trainers, lack of buy-in from employers and local education providers with a consequent lack of structures to embed formative assessment in practice and most importantly a culture of "one size fits all" to WPBAs, rather than an approach tailored to individualising learning.

These issues particularly disadvantage IMGs who may, as relative novices in a new system, need more intensive supervision. Systems that are not designed to enable such supervision for e.g. fixed out-patient clinic slots without Consultant supervision time factored in or without the flexibility to enhance the level of supervision for trainees that need it, fail to provide the learning environment needed for continuing growth. Detrimental as this is for any trainee, it serves to worsen the trajectory of DA for IMGs and BAME trainees.

DA in Continuous/Formative Assessment

So far, we have presented the case for the important role of formative assessment in supporting IMG and BAME doctors' professional growth trajectory, in training. However, a valid question emerges. Are formative assessment processes and tools not subject to the same biases and potential discrimination as summative tools?

Unfortunately, equality impact data for formative assessments is hard to come by. There is data from ARCP outcomes (summative in nature), demonstrating differential attainment based on ethnicity and gender with black trainees suffering the most adverse outcomes. (32) IMGs, BAME doctors and older doctors are more likely to have unsatisfactory outcomes at ARCP. (33) There is concern that current ARCP processes, reliant as they often are on a narrow range of sources of evidence; for e.g. significant reliance on a single educational supervisor's report, can adversely impact on the validity and reliability of the summative decision reached. While ARCP decision aids have been introduced to reduce variability, the lack of robust, consistent and reliable processes reduce the faith that both trainers and trainees have in the ARCP process.

The ARCPs have been criticised for setting the bar very low, being sensitive only to identify the '*worst performing*' trainees and for failing to encourage excellence. For the small number of doctors who are given an outcome 4 (exiting the training programme), much like the doctors who fail postgraduate summative examinations, there is no standardised process, to offer an alternative career path, that ensures ongoing development and thus patient safety. In fact, given that the trainees have failed to achieve competency of looking after patients safely and effectively, the outcome more often than not, is the *withdrawal*

of supervision rather than the offer of additional supervision.

For IMGs and BAME trainees therefore, the process may seem particularly punitive. ARCP is a summative process based on data from formative assessments but in contrast to the principles of formative processes, environmental and contextual learning factors are not taken into account. However, according to a comprehensive review of ARCPs commissioned by the HEE (Health Education England), the main criticism of the ARCP process relates to the absence of the standard of scrutiny through psychometric and equality impact evaluation, that other high stakes summative assessments are subject to. On the one hand, this reduces the validity and fidelity of the ARCP process, a serious concern given its recognised impact on safeguarding patient care and safety; but equally, the lack of standardisation creates the risk and/or the perception of bias, unconscious or otherwise.

It is important to note that while the review criticises the ARCP process on procedural grounds, it not only defends but calls for further strengthening of the *educational* viz. formative and developmental aspects of ARCP. This review makes a series of recommendations to address the shortcomings detailed above. In conjunction with the arguments made in this paper, these recommendations offer a way forward to a more robust process of making judgements about progression.

The Way Forward

The central purpose of assessments in the medical context is to assist both trainee and trainer in identifying ways of effective acquisition of knowledge, behaviour and skills to perform as a safe and competent doctor. An additional responsibility is to detect deficiencies and offer specific training needed to develop. A few key principles merit reiteration before considering implementation issues.

1. *Assessment of capabilities*
   There is a shift in emphasis from assessment of narrow, artificial competencies that do not reflect real-world clinical scenarios, to the assessment of core capabilities that inspire confidence in assessors of a trainee's ability to deal with similar clinical presentations in a range of

contexts. There is a recognition that the clinical workplace offers a golden opportunity to assess the trainee '*in vivo*' and does not have the obvious limitations of assessments in '*simulated environments*'.

2. *Assessment for learning*
Formative assessments in the workplace offer the opportunity of immediate feedback. Accompanied by reflection and appropriate supervision, they enable a learning culture, allowing for in-depth learning that can transform real world practice. Learning plans can then be tailored based on individual needs and strengths. This tiered instruction allows for second (or multiple chances) at reaching a particular milestone. When mapped to curricular outcomes, these assessments can be used to map growth in knowledge, skills and attitudes.

3. *Assessment of growth*
Doctors in training will acquire new knowledge and skills at varying pace depending on a range of factors, and based on their individual curiosity, intellectual engagement and clinical opportunity as well as supervision. The new GPC framework is used as a scaffolding for designing the new specialty curricula. Extending beyond training, revalidation criteria are mapped onto Good Medical Practice guidance. GPC capabilities offer a thread that stretches spirally across the training years and beyond. Demonstrating growth in this context necessitates continuous assessment to establish progress from novice to expert across a range of domains. When these capabilities are blueprinted to specialty curricular outcomes and to assessment strategies, current summative assessments are but one staging post on a continuum. Programmatic assessments have used this methodology with success in progress testing.(26)

4. *Assessment of training programmes*
A feature of formative assessment often forgotten, is that milestone assessments are as much a progress guide for the training programme, as for the trainees. Failure of a trainee to reach a particular milestone should trigger corrective action from the training programme, to address the learning need. Regular reviews of trainees' progress against set curricular outcomes then allows for training programmes to alter their training offer. Indeed, this is not an optional extra that training programmes might offer, but is embedded in the requirements set out in the Gold Guide.(34) This is particularly relevant for trainees with a different set of learning environments and contexts as may be true for IMGs for example.

5. *Embedding Equality in Assessment Structures*
Good assessments ought to be a) valid i.e. measure what they are supposed to measure; b) reliable i.e. do so consistently in various contexts independent of raters; and c) fair - discriminate fairly based on ability and not based on any other characteristic. Addressing differential outcomes in candidates then is primarily an *educational* task. Poor assessment processes discriminate based on ethnicity, gender etc. Equality impact assessments are not an optional extra, but are an integral part of designing a good assessment. This principle is vital as hitherto the discussion around DA has been framed around a deficit model - attributing differences in outcomes to communication skills, acculturation problems etc. on trainee characteristics. This fails to acknowledge the role of systemic failures in assessment systems and structures that entrench DA. The recognition of this fact has now led for shifting the nomenclature from Differential Attainment to Differential Awards.

Assessment vs Judgement

This paper has focused on two key questions. First, concerning the validity of current assessment processes, and second concerning the fairness of current assessment systems. Validity is reported to be the most important element of assessment quality. (33) Achieving 100% reliability on a measure that is invalid or false, is a wasted opportunity, and in the context of medical assessments, may even be dangerous. Validity is often defined narrowly as the ability of

a test to measure, what it purports to measure, but given the consequences of medical assessments on patients, employers, public finances the measure also needs to include the answer to the broader question - "is the assessment method fit for purpose"?.

The broadening of the learning outcomes framework to include a wide range of capabilities, emphasises findings from reviews that a wide range of assessment tools are needed to evaluate them. Evidence also suggests that a single assessment method, in a simulated environment, may not adequately inform what is a crucial judgement -viz. the competence of an individual to practice as a safe and competent clinician.

This principle has been accepted with ARCPs, now a well-established part of the overall assessment process. ARCPs are summative judgements about progression made on the basis of a series of formative assessments. This establishes the principle of continuum of assessments, informing progression decisions, and this has been further emphasised with Woolf's review recommending the institution of low stakes pre-ARCP reviews.(33) The adoption of progress testing in medical schools has added further weight and momentum to this change.

Equality perspectives have not been centre-stage in what is primarily viewed as educational discussions about the validity, reliability and robustness of assessments. However, a parallel discussion has been taking place in the educational world about the inherent unfairness and discrimination associated with DA.

This paper draws together these two strands of discussion. Unfair assessments are abhorrent because they are discriminatory but equally importantly, they are unacceptable being invalid assessments. Unequal assessments are inherently poor assessments. Unfair assessments are not only morally wrong, but are dangerous from a patient safety and public confidence perspective. Placing disproportionate reliance on a form of assessment that has been proven to be unequal, is wrong morally, educationally and violates the fundamental validity of the assessment process. The principle of using data from multiple sources to inform judgements about progression is established. Excluding certain elements of assessments from the assessment continuum, is internally inconsistent and raises important questions, not merely from an equality perspective but also from a core validity perspective.

This paper argues that methodological problems with the implementation of formative assessments have contributed to a misdirected focus on IMGs/BAME, factors to the exclusion of systemic issues related to assessment processes. Judgements about progression benefit from the richness of data available about trainees from their routine clinical practice.

A variety of problems have been identified that might have precluded formative assessment processes from playing a fuller role in informing overall progression decisions. A full discussion of remedial strategies is beyond the scope of this paper. However, none of the issues seem insurmountable. Indeed, tools such as Entrustable Professional Activity (EPAs) (35) seem more aligned to the capability framework and having been adopted in Australia and New Zealand, are now being explored in the UK.

Coproduction and co-designing formative tools and rater training involving diverse groups of trainers and trainees, can address issues around engagement of trainers. However, the key to embedding these changes will be ensuring a clear blueprint of learning outcomes matched to training methods, and to workplace-based assessments supported by continuous testing. Crucially, introducing to formative assessments, the gravitas and robust standardisation associated with traditional summative testing, will be required to ensure the validity, reliability, utility and fairness these assessment strategies deserve. Ideas such as the use of external assessors, introduction of low stakes assessments and the use of technology to record, assess and in some cases provide formative real-time feedback all need consideration. Selection of appropriate low stakes formative tools when mapped to the curricular outcomes may also provide a form of adaptive testing, preparing trainees for summative tests. This would help address one of the key problems related to DA viz. that the results of summative assessments come as a "surprise".

It is worth noting that using diverse sources of data in making summative judgements about progression or qualification is not a new process. Certificate of Eligibility for Specialist Registration (CESR) is a process by which doctors who have acquired practice capabilities outside GMC approved training programmes are able to demonstrate their specialist training, qualifications and experience. Core and higher training recruitment processes routinely blend

summative assessments with an evaluation of work experience through portfolios.

Conclusion:

The primary aim of all the assessments should be to produce a safe, competent and professional clinician. It is important that assessments are valid and fair as well as being reliable and defensible. Unfair assessments may lead to false positives i.e. trainees poor in clinical practice in the real-world, but good at passing exams and also false negatives i.e. trainees good at clinical practice in the real world, but poor at exam performance thus not able to pass exams.

Formative assessments have the potential to complement summative assessments and the overall judgements related to progression. The use of a more blended approach will be necessitated with the diversity of high-level outcomes that comprise the General Professional Capabilities Framework; but should also, in line with the recommendations made by Roe et al, lead to improvements in differential attainment (36) Our doctors need to be judged by a regime of assessments that can take into account their specialist expertise as also the broader more general skill set (37) and do so fairly, equitably, consistently and reliably. Failure to do so urgently risks damaging the confidence that the public have invested in the fidelity of our assessment systems.

A broadened perspective on the types of construct, assessment tries to capture, the way information from various sources is collected and collated, the role of human judgement and the variety of psychometric methods to determine the quality of the assessment. Research into the quality of assessment programmes, how assessment influences learning and teaching, new psychometric models and the role of human judgement is much needed.

References

1. What is revalidation? [Internet]. [cited 2020 Nov 1]. Available from: https://www.gmc-uk.org/registration-and-licensing/managing-your-registration/revalidation/introduction-to-revalidation

2. Improving Assessment: Further Guidance and Recommendations [Internet]. Academy of Medical Royal Colleges. 2016 [cited 2020 Nov 1]. Available from: https://www.aomrc.org.uk/reports-guidance/improving-assessment-guidance-recommendations/

3. Cost of Training 2017 [Internet]. Academy of Medical Royal Colleges. 2017 [cited 2020 Nov 1]. Available from: https://www.aomrc.org.uk/reports-guidance/cost-training-2017/

4. Participation E. Medical Act 1983 [Internet]. Statute Law Database; [cited 2020 Nov 1]. Available from: https://www.legislation.gov.uk/ukpga/1983/54/contents

5. Woolf K. Fair Training Pathways for All: Understanding Experiences of Progression - Final Report. :69.

6. Woolf K, Potts HWW, McManus IC. Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis. BMJ. 2011 Mar 8;342:d901.

7. NHS Staff Surveys - NHS Staff Survey Results [Internet]. [cited 2020 Jul 8]. Available from: https://www.nhsstaffsurveys.com/Page/1085/Latest-Results/NHS-Staff-Survey-Results/

8. Dave S, Chakravorty I, Menon G, Sidhu K, Bamrah JS, Mehta R. Differential Attainment in Summative Assessments within Postgraduate Medical Education & Training: Sushruta J Health Policy Opin [Internet]. 2020 Aug 9 [cited 2020 Sep 28];13(3). Available from: https://sushrutajnl.net/index.php/sushruta/article/view/89

9. Gibbs G, Simpson C. Conditions Under Which Assessment Supports Students' Learning. :30.

10. Denholm-Price J, Page N, Williams N, Dourado L. Exploring differential attainment by assessment type in mathematics, chemistry and life sciences [Internet]. Presentation presented at: Horizons in STEM Higher Education Conference : Making Connections and Sharing Pedagogy; 2019 Jul 3 [cited 2020 Oct 23]; Kingston-upon-Thames, U.K. Available from: https://eprints.kingston.ac.uk/id/eprint/43637/

11. Bloom BS, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, Raths J, et al. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Longman; 2001. 342 p.

12. Bennett RE. Formative assessment: a critical review. Assess Educ Princ Policy Pract. 2011 Feb 1;18(1):5–25.

13. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: From assessment of learning to assessment for learning. Med Teach. 2011;33(6):478–85.

14. Holmboe ES, Call S, Ficalora RD. Milestones and Competency-Based Medical Education in Internal Medicine. JAMA Intern Med. 2016 Nov 1;176(11):1601–2.

15. Mountford-Zimdars A, Sanders J, Moore J, Sabri D, Jones S, Higham L. What can universities do to support all their students to progress successfully throughout their time at university? Perspect Policy Pract High Educ. 2017 Jul 3;21(2–3):101–10.

16. Generic professional capabilities framework [Internet]. [cited 2020 Jul 8]. Available from:

https://www.gmc-uk.org/education/standards-guidance-and-curricula/standards-and-outcomes/generic-professional-capabilities-framework

17. Differential attainment in medical education and training | The BMJ [Internet]. [cited 2020 Nov 2]. Available from: https://www.bmj.com/content/368/bmj.m339

18. Good medical practice [Internet]. [cited 2020 Oct 27]. Available from: https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/good-medical-practice

19. Promoting excellence: standards for medical education and training. :51.

20. Coping with complexity: educating for capability | The BMJ [Internet]. [cited 2020 Oct 27]. Available from: https://www.bmj.com/content/323/7316/799

21. Slowther A, Lewando Hundt GA, Purkis J, Taylor R. Experiences of non-UK-qualified doctors working within the UK regulatory framework: a qualitative study. J R Soc Med. 2012 Apr;105(4):157–65.

22. evaluation-of-gmc-welcome-to-uk-practice---january-2019_pdf-79429900.pdf [Internet]. [cited 2020 Nov 2]. Available from: https://www.gmc-uk.org/-/media/documents/evaluation-of-gmc-welcome-to-uk-practice---january-2019_pdf-79429900.pdf

23. Hawkins RE, Margolis MJ, Durning SJ, Norcini JJ. Constructing a validity argument for the mini-Clinical Evaluation Exercise: a review of the research. Acad Med J Assoc Am Med Coll. 2010 Sep;85(9):1453–61.

24. Wolcott MD, Zeeman JM, Cox WC, McLaughlin JE. Using the multiple mini interview as an assessment strategy within the first year of a health professions curriculum. BMC Med Educ [Internet]. 2018 May 3 [cited 2020 Nov 2];18. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5934879/

25. Schuwirth LWT, van der Vleuten CPM. The use of progress testing. Perspect Med Educ. 2012 Mar 1;1(1):24–30.

26. Van Der Vleuten CPM, Schuwirth LWT, Driessen EW, Govaerts MJB, Heeneman S. Twelve Tips for programmatic assessment. Med Teach. 2015 Jul 3;37(7):641–6.

27. Rethans JJ, Sturmans F, Drop R, van der Vleuten C, Hobus P. Does competence of general practitioners predict their performance? Comparison between examination setting and actual practice. BMJ. 1991 Nov 30;303(6814):1377–80.

28. Epstein RM, Hundert EM. Defining and Assessing Professional Competence. JAMA. 2002 Jan 9;287(2):226–35.

29. Warwick C. How international medical graduates view their learning needs for UK GP training. Educ Prim Care. 2014 Jan 1;25(2):84–90.

30. Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. Health Aff Proj Hope. 2002 Oct;21(5):103–11.

31. Brittlebank A, Archer J, Longson D, Malik A, Bhugra DK. Workplace-Based Assessments in Psychiatry: Evaluation of a Whole Assessment System. Acad Psychiatry. 2013 Sep 1;37(5):301–7.

32. The State of physicianly training in the UK - 2019 report [Internet]. JRCPTB. 2019 [cited 2020 Nov 7]. Available from: https://www.jrcptb.org.uk/documents/state-physicianly-training-uk-2019-report

33. Academic support for the Assessment and Appraisal workstream ofHealth Education Englands review of the ARCP.pdf [Internet]. [cited 2020 Nov 7]. Available from: https://www.hee.nhs.uk/sites/default/files/documents/Academic%20support%20for%20the%20Assessment%20and%20Appraisal%20workstream%20ofHealth%20Education%20Englands%20review%20of%20the%20ARCP.pdf

34. Gold Guide - 8th Edition - Conference Of Postgraduate Medical Deans [Internet]. [cited 2020 Nov 7]. Available from: https://www.copmed.org.uk/gold-guide-8th-edition/

35. a_guide_to_entrustable_professional_activities.pdf [Internet]. [cited 2020 Nov 7]. Available from: https://www.rcr.ac.uk/sites/default/files/a_guide_to_entrustable_professional_activities.pdf

36. Roe, V., Patterson, F., Kerrin, M., Edwards, H. (2019) "What supported your success in training?" A qualitative exploration of the factors associated with an absence of an ethnic attainment gap in post-graduate specialty training. https://www.gmc-uk.org/-/media/documents/final-report-success-factors-in-training-211119-pdf_pdf-81634780.pdf [Internet] [cited 2020 Nov 10]

37. Future Doctor [Internet]. Health Education England. 2019 [cited 2020 Nov 7]. Available from: https://www.hee.nhs.uk/our-work/future-doctor