

# Bridging the GAP

ALLIANCE FOR EQUALITY IN HEALTHCARE PROFESSIONS  
2020 Thematic Series on Differential Attainment

*Theme II Assessment*

*Roundtable 2.1 17 September 2020*

## **Summative Assessments in PG Medical Education & Training**

**PRE-WORKSHOP SCOPING PAPER – II.1 (v5 -09.08.2020)**

**Authors**

**Subodh Dave, Geeta Menon, Kamal Sidhu, Indranil Chakravorty, Ramesh Mehta & JS Bamrah**

### **Executive Summary**

- This discussion paper has been prepared for the expert roundtable exploring the ‘*Differential Attainment in PG Medical Education and Training*’ planned for 17 September 2020. This will be the first engagement exercise launching the 2020 Thematic series on *Tackling differential attainment in Healthcare professions*, bringing together an interdisciplinary *Alliance on equality in healthcare professions*.
- This paper presents a preliminary outline of the current evidence on differential attainment in high stakes postgraduate summative assessment, explores its impact, deliberates on known causes and discusses a number of potential solutions.
- This paper is written with a view to present the case for tackling DA in PG summative assessments and will be accompanied by a prioritised selection of ‘focussed questions and solutions’ to be discussed at the roundtable with subject experts.
- This paper and roundtable will form part of, and contribute to the thematic synthesis in the section on ‘**Assessment - formative and summative**’. Therefore, as described in the ‘protocol’, will be followed by a focussed systematic review, engagement with priority setting partnerships (via questionnaires, focus groups and workshops) and culminate in an expert consensus.
- The final outcome will be presented as synthesized recommendations, solutions, policy enablers and areas for further research.

### What is Differential Attainment?

Differential attainment (DA) is a term used to describe the variations in levels of educational achievement that occur between different demographic groups undertaking the same assessment. UK doctors from Black and Minority Ethnic (BAME) groups, and International Medical Graduates (IMG) i.e. doctors whose Primary Medical Qualification (PMQ) is from a medical school outside of the UK have, consistently, poorer outcomes in assessments and recruitment compared to white doctors and UK medical school graduates. <sup>1 2</sup> Differential attainment has been recognised as a challenge for medical professionals and educators since the 1990s.

### How big is the problem?

Ethnic minority medical graduates in the UK have 2.5 times higher odds of failing high-stakes exams. <sup>3</sup> Summative assessments for the membership of the Royal Colleges of Physicians (MRCP), General Practitioners (MRCGP) and Psychiatrists (MRCPsych), amongst others have shown a consistent medium sized ethnicity effect and a larger country of PMQ effect. This translates to a 10-15% gap in pass rates for UK BAME candidates and a larger approximately 30-50% gap in pass rates for IMGs. The CSA – Clinical Skills Assessment exam of the Royal College of General Practitioners has a number of specific issues which makes the issue of DA particularly problematic.

### Impact of COVID-19

COVID-19 pandemic led to cancellation of both applied knowledge test (AKT) <sup>4</sup> and Clinical Skills Assessment (CSA)<sup>5</sup> and alternative solutions being considered. After concerns from General Practice registrars (GPRs) and various organisations including British Association of Physicians of Indian Origin (BAPIO), the RCGP has provided an interim alternative to the CSA in form of Recorded Consultation Assessments (RCA).<sup>5</sup> This format involves recording thirteen consultations i.e. same number as CSA of in audio, video or face-to-face format and submitting it to the panel of examiners,

who will carry out objective assessments using same criteria as used in the CSA.<sup>6</sup> The advantage is that the GPRs can select from the consultations carried out in their own surgery environment rather than in an artificial environment that involved actors.

Although understandably, this is posing some logistical challenges for the trainees, especially those working remotely due to personal risks such as pregnancy or other health conditions, this format may well give a basis or a 'trial run' of an alternative option. There is also concern that the CSA may well be an outdated method of assessment and not reflective of the changing nature of general practice.<sup>7,8</sup>

### Why is Differential Attainment a problem?

#### Moral and Ethical Impact

Clearly, the significant attainment gap based on ethnicity (and country of origin) poses a significant social justice issue. The fact that these attainment gaps have persisted for decades with no institutional redressal, compounds the ethical and moral problem and makes the case for urgent remediation.

For IMGs, whose visas or permission to remain in the UK may be dependent on exam success, this creates uncertainty, economic instability, anxiety and undue distress. In practice, the attainment gap serves to multiply the microaggressions that BAME students, trainees and staff face in clinical and educational settings. <sup>9</sup> BAPIO has received testimonies from a large number of individuals where exam related stress has been specifically identified as a source of great personal and professional difficulties. <sup>10</sup>

#### Workforce and Financial Impact

Around a third of UK medical students (n ~ 11000) and graduates (who are not Consultants or GPs) are of BAME origin (n ~ 28000).<sup>11</sup> IMGs also constitute a very large part of the workforce and especially so in some specialities such as Psychiatry and General Practice where they constitute >35% of the workforce.<sup>12</sup> In 2019, the number of IMGs entering

the General Medical Council (GMC) register exceeded the number of UK graduates.<sup>13</sup> These numbers illustrate the scale and extent of the impact of DA.

The inevitable necessity of the UK National Health Service (NHS) in depending on IMGs to deliver patient care is evident also in the high number of vacancy rates across the country in many clinical specialties, in various geographical locations and in the high cost of providing locum cover to run essential services.<sup>14</sup> If clinical examinations prove an unfair barrier to career progression, this may represent a significant workforce challenge with direct adverse impact on patient care.<sup>15</sup> Furthermore, the costs of failure in high stakes examinations costs (approximately £65,000 per failure) pose huge economic burden in further education and ancillary costs and organisational level.<sup>16, 17</sup>

#### *Impact on Patient Care*

A sense of equality among health workers translates to better team working which inevitably leads to better patient outcomes and satisfaction for the organisation. It is known that the proportion of staff believing the employing organisation provides equal opportunities for career progression or promotion “was a very important predictor of patient satisfaction.”<sup>9</sup> Unfortunately, BAME staff routinely report microaggressions at work.<sup>18</sup>

However, there is currently little evidence linking success or failure in high stakes exams with a direct or indirect impact on patient care and safety<sup>19</sup> there may even be evidence to the contrary, demonstrating that overseas trained IMGs delivered improved patient outcomes.<sup>20</sup> Moreover, there are concerns on the OSCE (Objective Structured Clinical Examination) as a valid assessment reflecting clinical reality particularly in certain specialties.<sup>21</sup>

Given the multicultural and diverse population in the UK, it is important to address inequalities in medical education and training to ensure patients can benefit from an ethnically diverse medical workforce.<sup>22</sup>

#### *Legal Impact*

Mr Justice Mitting’s ruling in the BAPIO vs. RCGP legal action has clearly indicated that providers and standard setters of education and training viz. Health Education England, Deaneries, Health Boards and Royal Colleges in the UK are subject to the Public Sector Equality Duty and hence have a legal and regulatory obligation to monitor and tackle inequalities.<sup>23</sup>

#### *Causes of DA in PG Medical Assessment*

Several factors have been implicated as causative or contributory in DA. Prior educational attainment generally predicts future academic attainment, but multivariate analysis of data shows that DA in medical school finals persist even after accounting for prior educational attainment. DA persists even after accounting for socio-economic deprivation. In fact, ethnic differences in attainment persist even after controlling for type of school, personality, motivation, study habits and mental health of candidates as well as linguistic ability, often cited as a cause for DA. Ethnic differences in attainment persist after controlling for one’s own first language and parents’ first language.<sup>24</sup>

There are a range of factors related to either the examination itself or to the training environment leading up to the examination that may explain DA. IMGs often face additional difficulties which impede examination success due to differences in educational experience, content familiarity and language, some of which may be potentially amenable to modification or additional support.<sup>25</sup>

Apart from the factors that have been ruled out (see above), possible candidate factors that have been implicated include relationship with peers, relationship with educators, the presence of undiagnosed and undetected learning disability such as dyslexia and undue pressure from expectations of passing/failure.<sup>24</sup>

Factors relating to examinations may include unconscious or conscious bias in examiners, in the recruitment of examiners, in the choice of exam questions or case selection for OSCE stations or in standard setting and/or applying the set standards in the exam.<sup>26, 27</sup>

## Are summative exams unfair?

Esmail and Roberts' study analysing the data of academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between 2010 and 2012 showed that, even after controlling for performance on the machine-marked AKT, ethnic minority UK graduates were nearly four times and international medical graduates 14 times as likely to fail their first CSA attempt as white candidates. The authors concluded that "subjective bias due to racial discrimination in the CSA may be a cause of failure for UK trained candidates and IMGs. 28, 29

However, in the courts the examination was judged lawful. Others too, have argued that DA is indicative of a true attainment gap based on consistent and correlated DA seen in candidates taking both MRCGP and MRCP (UK) exams 30 31 lack of proven ethnicity or gender bias in examiners in MRCP exams on two-examiner stations 32 or the lack of proven role player bias in CSA exams. 33 It is indeed worth noting that gender or ethnicity bias have not been disproven in single examiner stations. Unconscious bias training often provided to examiners and role players to mitigate against DA has proved to be ineffective 34 and while systematic review evidence suggests that discrimination is unlikely to be the sole cause of DA, 3 the current evidence clearly does not rule out covert or overt discrimination as a cause of DA.

Assessment oversight committees and annual programmatic evaluations, while recommended, will not guarantee fairness within postgraduate medical education programs, but they can provide a window into 'hidden' threats to fairness, as everything from training experiences to assessment practices may be open to scrutiny. 35

Ensuring Fairness in Clinical Training and Assessment: Principles and examples of good practice, was recommended by the BMA outlined a few principles that need to be considered with respect to assessment methods.

## Current Difficulties with Objective Structured Clinical Examinations (OSCE)

When evaluated against the standard criteria, independent of its ethnicity effect, a few problems emerge with the current traditional OSCE format.

Firstly, the artifice of OSCEs makes validity a significant concern. Rating scales and checklist assessment tools used to improve reliability ends up rewarding mechanistic "performance" from candidates. A striking example of this problem is the paradoxical third person rating of empathy often used in OSCEs assessing communication skills. OSCEs that reward feigning empathy rather than actual empathy have been blamed for the striking reduction in empathy seen in medical students as they progress through their medical training.36 Validity depends on high levels of fidelity but that is usually lacking as OSCEs usually test isolated skills in a fragmented fashion. 37 38

OSCEs improve on their reliability coefficients by increasing the duration of the exam but these remain susceptible to biases in sampling of stations. Standard setting in high-stakes exams is done variably for different cohorts and while this could be improved, there remains the variability in examiners. All exams do review the "hawks and doves" in their examiner pool but again this categorical distinction may mask granular details for e.g. the finding that IMG examiners may be more hawkish. 39

Another interesting finding relates to the finding that performance at the MRCGP clinical skills assessment in IMGs was better predicted by scores on a situational judgment test, evaluating interpersonal skills, than by achievement on a knowledge-based test. 17 This finding is also supported by previous reports that GMCs Professional and Linguistic Assessment Board examination (PLAB) part 2 scores, rather than those for part 1, predicted performance in the clinical components of MRCP and MRCGP CSA exams. 31 This is of concern particularly given the known ethnicity discriminatory effect (against BAME candidates) that is a consistent feature of the Situational Judgement Test. 40

Assessment does drive learning and clearly summative examinations have a role in not merely quality assurance but in also promoting essential learning and practice that delivers high quality and safe care for patients. However, this does depend on high quality, specific and credible feedback being delivered to failed candidates with tailored remediation. Currently, the feedback given to failed candidates fails to meet any of these criteria. Pertinently, there is no evidence to link success or failures in OSCE-style exams with patient safety or patient outcomes.

### **Alternatives to OSCEs - Programmatic Assessment; multiple low stakes assessments**

There is some shift in focus within medical education, from learning discrete skills and knowledge to continuous learning with authentic tasks focused on transfer to clinical practice. GMC's Generic Professional Capabilities Framework signals this direction very clearly and is now leading to changes in postgraduate curricula across the board. <sup>41</sup> The underlying message is clear – we need to move from “shows how” to “does”.

The public expect their doctors to be capable of working in a range of different situations and settings and there is wide understanding that no single assessment method can capture it all. Current assessment strategy focusing as it does, on summative assessment at a single point of time, provides little weightage for longitudinal assessments.

Narrative feedback embedded in a dialogue (rather than one-way provision of feedback) is significantly more impactful in developing complex clinical skills than scores. Longitudinal and more diverse programmatic assessment can address the inherent difficulties in relying on a single data point viz. the summative OSCE examination. Moving from a sum of a few summative/formative assessments to a programme of multiple low-stakes assessment would provide multiple data points which can be optimised for learning. The format of assessments can be varied at various data points which would improve the validity of assessment.

Current summative examinations are focused on delivering a categorical pass/fail distinction and considerable effort is expended in designing exams that are defensible- the main focus of the assessment is this decision rather than on the primary function of assessment, which is to drive patient-centred learning.

Switching from decision-oriented to feedback-oriented multiple assessments with varying degrees of stakes at each data point would generate feedback focused on improving the quality of care for patients, something that current assessment strategies do not emphasise. Crucially, such longitudinal assessment delivers non-surprising results in the final stages of the assessment. The fact that the failure in high-stakes assessment comes as a surprise to both trainers and trainees has been a significant problem with current summative exams. Those likely to fail should be identified earlier on in their learning trajectory and remedial action instituted.

Such programmatic assessments are being used in many centres across the world including the USA, Canada and Holland. Within the UK setting, the current system of Workplace Based Assessments, Annual Review of Competency Progression and summative paper exams including OSCEs should be adapted relatively easily to create a more longitudinal systematic and programmatic assessment. This will empower trainers to use their professional judgement (rather than relying on standard setting or on narrow checklists which have been associated with reduced validity). Increasing the number of data points will increase the diversity of the assessment sample, potentially increase the diversity in the examiner pool and aided by procedural bias reduction methods should deliver an exam that puts person-centred care and learning rather than pass/fail decisions at the heart of assessment.

### **Initiatives so far**

- Following the legal challenge, the GMC and some Royal Colleges have had regular discussions with BAPIO and have produced examination preparation resources as well as enhanced guidance for trainers.

- RCGP has introduced an exceptional 5<sup>th</sup> attempt for some candidates in the CSA.
- A Health Education North West Pilot programme for enhanced training has been shown to improve outcomes of CSA resits.
- Address the conscious and unconscious biases that exist amongst tutors as well as examiners

### Recommendations

- Use real patients rather than role players.
- **Two** examiners may mark rather than one at every station or virtual examiners as employed in some USA systems may reduce undue stress
- Video of the assessment should be made available to failing candidates
- **Number of attempts** may be increased or made unlimited as long as the doctor is continuing in active medical practice.
- **Culvert Scoring:** The Education Supervisor provides a 'culvert score' to the trainee about 6 months prior to proposed finishing date of training. This score ranges from 0-3 depending on the overall performance of the candidate during the whole period of training and will be influenced by overall knowledge, communication skills, quality of the WPBA and several other factors. This score is not disclosed to the trainee but is available to the examining body. If a candidate is marginally falling short of CSA pass score, this culvert score may be added to the marks obtained in the CSA examination. If the candidate has already scored the pass marks, there is no need to use a culvert score.
- **Weight allocation:** "Weights" may be provided to the current three parts of the assessments (i.e. WPBA, AKT and CSA). Weighted scores from all three assessments then may be combined to provide the accreditation score. The accreditation score may be fixed beforehand again based on the survey results, for example 65% or 70%. Actual weights may be decided following a survey conducted from the trainees, trainers and examiners.
- Promoting cultural safety, cultural humility and decolonization of the curriculum and content

## References

1. The state of medical education and practice in the UK. <https://www.gmc-uk.org/about/what-we-do-and-why/data-and-research/the-state-of-medical-education-and-practice-in-the-uk>.
2. Woolf, K. Fair Training Pathways for All: Understanding Experiences of Progression - Final Report. 69.
3. Woolf, K., Potts, H. W. W. & McManus, I. C. Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis. *BMJ* **342**, d901 (2011).
4. MRCGP | Applied Knowledge Test (AKT). <https://www.rcgp.org.uk/training-exams/mrcgp-exam-overview/mrcgp-applied-knowledge-test-akt.aspx>.
5. MRCGP: Clinical Skills Assessment (CSA). <https://www.rcgp.org.uk/training-exams/mrcgp-exam-overview/mrcgp-clinical-skills-assessment-csa.aspx>.
6. MRCGP: Recorded Consultation Assessment (RCA). <https://www.rcgp.org.uk/training-exams/mrcgp-exam-overview/mrcgp-recorded-consultation-assessment.aspx>.
7. March 2016, 16. Former RCGP chair suggests CSA exam should be scrapped. Pulse Today <http://www.pulsetoday.co.uk/news/gp-topics/education/former-rcgp-chair-suggests-csa-exam-should-be-scrapped/20031376.article>.
8. Covid-19: changes to GP assessment should only be the beginning. *The BMJ* <https://blogs.bmj.com/bmj/2020/06/05/covid-19-changes-to-gp-assessment-should-only-be-the-beginning/> (2020).
9. NHS Staff Surveys - NHS Staff Survey Results. <https://www.nhsstaffsurveys.com/Page/1085/Latest-Results/NHS-Staff-Survey-Results/>.
10. Woolf, K., Viney, R., Rich, A., Jayaweera, H. & Griffin, A. Organisational perspectives on addressing differential attainment in postgraduate medical education: a qualitative study in the UK. *BMJ Open* **8**, (2018).
11. NHS workforce. <https://www.ethnicity-facts-figures.service.gov.uk/workforce-and-business/workforce-diversity/nhs-workforce/latest#by-ethnicity-and-grade-medical-staff>.
12. Valero-Sanchez, I., McKimm, J. & Green, R. A helping hand for international medical graduates. *BMJ* **359**, (2017).
13. O'Dowd, A. More non-UK graduates than home grown clinicians joined medical register in past year. *BMJ* **367**, (2019).
14. Rimmer, A. Locum pay since the cap. *BMJ* **357**, (2017).
15. Background | Falling short: the NHS workforce challenge. Health Foundation <https://reader.health.org.uk/falling-short/background>.
16. Entry Through the Narrow Door.
17. Fielding, S. et al. Do changing medical admissions practices in the UK impact on who is admitted? An interrupted time series analysis. *BMJ Open* **8**, (2018). <https://www.bmj.com/lookup/doi/10.1136/bmjopen-2017-022222>.
18. [inclusion-report-aug-2017.pdf](https://www.bmj.com/lookup/doi/10.1136/bmjopen-2017-022222).
19. Archer, J. et al. The impact of large scale licensing examinations in highly developed countries: a systematic review. *BMC Med. Educ.* **16**, 212 (2016).
20. Norcini, J. J. et al. Evaluating the quality of care provided by graduates of international medical schools. *Health Aff. Proj. Hope* **29**, 1461–1468 (2010).
21. Hughes, N. S. et al. Exit examination: a survey of UK psychiatrists' views. *BJPsych Bull.* **39**, 254–259 (2015).
22. Krishnan, A., Rabinowitz, M., Ziminsky, A., Scott, S. M. & Chretien, K. C. Addressing Race, Culture, and Structural Inequality in Medical Education: A Guide for Revising Teaching Cases. *Acad. Med. J. Assoc. Am. Med. Coll.* **94**, 550–555 (2019).
23. Dyer, C. RCGP is cleared of ethnic discrimination in clinical skills assessment. *BMJ* **348**, (2014).
24. Woolf, K., Rich, A., Viney, R., Needleman, S. & Griffin, A. Perceived causes of differential attainment in UK postgraduate medical training: a national qualitative study. *BMJ Open* **6**, e013429 (2016).
25. Pattinson, J., Blow, C., Sinha, B. & Siriwardena, A. Exploring reasons for differences in performance between UK and international medical graduates in the Membership of the Royal College of General Practitioners Applied Knowledge Test: a cognitive interview study. *BMJ Open* **9**, e030341 (2019).
26. Stone, J. & Moskowitz, G. B. Non-conscious bias in medical decision making: what can be done to reduce it? *Med. Educ.* **45**, 768–776 (2011).
27. Epstein, R. M. Assessment in Medical Education. *N. Engl. J. Med.* **356**, 387–396 (2007).
28. Academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between 2010 and 2012: analysis of data | *The BMJ*. <https://www.bmj.com/content/347/bmj.f5662>.
29. Linton, S. Taking the difference out of attainment. *BMJ* **368**, (2020).
30. Wakeford, R., Denney, M., Ludka-Stempien, K., Dacre, J. & McManus, I. C. Cross-comparison of MRCGP & MRCP(UK) in a database linkage study of 2,284 candidates taking both examinations: assessment of validity and differential performance by ethnicity. *BMC Med. Educ.* **15**, (2015).
31. McManus, I. C. & Wakeford, R. PLAB and UK graduates' performance on MRCP(UK) and MRCGP examinations: data linkage study. *BMJ* **348**, (2014).
32. McManus, I. C., Elder, A. T. & Dacre, J. Investigating possible ethnicity and sex bias in clinical examiners: an analysis of data from the MRCP(UK) PACES and nPACES examinations. *BMC Med. Educ.* **13**, (2013).
33. Denney, M. & Wakeford, R. Do role-players affect the outcome of a high-stakes postgraduate OSCE, in terms of candidate sex or ethnicity? Results from an analysis of the 52,702 anonymised case scores from one year of the MRCGP clinical skills assessment. *Educ. Prim.*

- 
- Care Off. Publ. Assoc. Course Organ. Natl. Assoc. GP Tutors World Organ. Fam. Dr. **27**, 39–43 (2016).
34. Atewologun, D. & Tresh, F. Unconscious bias training: An assessment of the evidence for effectiveness. [/paper/Unconscious-bias-training-%3A-An-assessment-of-the-Atewologun-Tresh/509a580783bb5df10f3af42e8df720d8ad57bd0](#) (2018).
35. Colbert, C. Y., French, J. C., Herring, M. E. & Dannefer, E. F. Fairness: the hidden challenge for competency-based postgraduate medical education programs. *Perspect. Med. Educ.* **6**, 347–355 (2017).
36. Gillett, G. Communication skills and the problem with fake patients. *BMJ* **357**, (2017).
37. Schuwirth, L. W. T. & van der Vleuten, C. P. M. How 'Testing' Has Become 'Programmatic Assessment for Learning'. *Health Prof. Educ.* **5**, 177–184 (2019).
38. Vleuten, C. P. M. van der & Swanson, D. B. Assessment of clinical skills with standardized patients: State of the art. *Teach. Learn. Med.* **2**, 58–76 (1990).
39. Denney, M. L., Freeman, A. & Wakeford, R. MRCGP CSA: are the examiners biased, favouring their own by sex, ethnicity, and degree source? *Br. J. Gen. Pract.* **63**, e718–e725 (2013).
40. de Leng, W. E., Stegers-Jager, K. M., Born, M. P. & Themmen, A. P. N. Integrity situational judgement test for medical school selection: judging 'what to do' versus 'what not to do'. *Med. Educ.* **52**, 427–437 (2018).
41. Generic professional capabilities framework. <https://www.gmc-uk.org/education/standards-guidance-and-curricula/standards-and-outcomes/generic-professional-capabilities-framework>.