The Journal of Medical Education Research 2025 Vol 5 No 1 pp 45-58, DOI: 10.5750/jmer.v5i1.2680

RACE AND ETHNICITY IN UK MEDICAL LICENSING ASSESSMENT SINGLE BEST ANSWER QUESTIONS: ETHICAL, PSYCHOMETRIC AND METHODOLOGICAL CONSIDERATIONS

*Imran Jawaid**
Associate Professor of Professional Practice and Clinical Assessment Lead, Faculty of Medicine and Heath Sciences and Editor-in-Chief, Journal of Medical Education Research, The University of Buckingham, United Kingdom

ABSTRACT

Single best answer (SBA) questions dominate UK medical licensing assessments, valued for their standardisation, reliability, and psychometric rigour. Yet the inclusion of demographic descriptors, particularly race and ethnicity, raises ethical and methodological challenges. Evidence indicates that presenting race or ethnicity in the opening sentence of an SBA stem may trigger cognitive biases, reinforce stereotypes, and introduce construct-irrelevant variance, potentially compromising fairness and validity. This perspective critically examines the use of race and ethnicity in UK licensing SBAs, emphasising that such descriptors should be included only when clinically relevant and embedded within the clinical observation or context rather than as initial identifiers. We propose a framework to guide item writers, aligning with international best practice and psychometric principles, to ensure ethical, unbiased assessment. By implementing these recommendations, the UK Medical Licensing Assessment can enhance construct validity, reduce bias, and promote equity, setting a precedent for ethically sound and psychometrically robust medical examinations.

Keywords: medical licensing assessment, single best answer questions, race and ethnicity, assessment fairness, psychometrics, equity, diversity, and inclusion, cognitive bias, construct validity

INTRODUCTION

Assessment lies at the heart of medical education, shaping not only what students learn but how they approach clinical reasoning. From early formative

*Corresponding author: e-mail: imran.jawaid@buckingham.ac.uk

exercises to high-stakes summative examinations, the methods we choose signal the skills and knowledge we value. Multiple-choice questions (MCQs), particularly single best answer (SBA) formats, have become a ubiquitous tool in this landscape. Beneath their apparent simplicity lies complexity; crafting questions that probe reasoning rather than recall, that reflect authentic clinical decision-making, and that avoid unintentional cues or bias, requires careful attention to both educational principles and psychometric rigour. MCQs are not merely a test of memory, but in fact a lens into the cognitive processes we hope to cultivate in future doctors.

SBA QUESTIONS IN MEDICAL EDUCATION

SBAs, long used in undergraduate final examinations, are recognised to be an effective assessment tool, reliably testing applied knowledge and decision-making at the "knows how" level of Miller's pyramid, while offering a standardised and objective method of assessment (Miller, 1990).

In a conventional SBA question, candidates are presented with a focused clinical scenario and a set of plausible options, from which they must select the single most appropriate or correct response (see Box 1). The information included in the stem should allow a student reading the question to differentiate between the correct answer option and the remaining incorrect options/ distractors.

Box 1. Example of an SBA question

Stem:
A 35 year old woman has bilateral intermittent tingling, pain and numbness in her thumbs, index and middle fingers for 3 weeks. Her symptoms are worse at night. She is 32 weeks pregnant.

Phalen's test and Tinel's test are positive.

Lead in:
Which nerve is most likely compressed?

Options:
Axillary
Median
Musculocutaneous
Radial
Ulnar

Correct answer: Median

Unlike traditional MCQs, SBAs are designed so that all options may appear partially correct, requiring candidates to apply knowledge, interpret information, and exercise clinical judgement rather than recall isolated facts (Schuwirth and van der Vleuten, 2011).

The use of SBAs has expanded rapidly over the last decade due to demands for scalability, standardisation, reliability and alignment with psychometric best practice (Zegugu et al., 2024; Sam et al., 2019; van der Vleuten, 2000). SBAs are now the default format across UK medical schools, postgraduate examinations, and licensing structures. The UK Medical Licensing Assessment (MLA) represents a significant shift in the national approach to medical assessment, with the Applied Knowledge Test (AKT) composed entirely of MCQs in the SBA format, cementing SBAs as the UK's primary high-stakes knowledge-assessment modality (General Medical Council, 2024).

This shift reflects three converging trends:

1) **Standardisation pressures:** ensuring consistent graduate readiness across UK medical schools (General Medical Council, 2024).
2) **Psychometric favourability:** SBAs offer strong internal reliability and can be statistically equated across sittings (Okubuiro, Ebirim, and Okoli, 2019).
3) **Administrative scalability:** computer-marked SBAs minimise cost and examiner burden relative to essays or orals (Cantillon, Irish and Sales, 2004).

However, SBAs have inherent limitations as an assessment modality. Their fixed, forced-choice format restricts assessment primarily to the cognitive domain and limits the evaluation of affective and psychomotor learning outcomes. In addition, the presence of predefined response options may introduce cueing effects, potentially facilitating recognition rather than independent reasoning and thereby inflating student performance in ways that are not reflective of authentic clinical decision-making (Parekh and Vikesh Bahadoor, 2024; Norcini et al., 1985). Sam et al. (2019) demonstrated students scored 21 percentage points higher on SBAs than equivalent Very Short Answer (VSA) items, suggesting cue-driven overestimation of competence.

Within this evolving assessment landscape, questions have been raised about the ethical and methodological considerations associated with including demographic characteristics, such as race, ethnicity, and other protected attributes, in SBA question stems. This perspective article explores whether race or ethnicity should be included routinely in the opening sentence of an SBA stem alongside other demographic details or only included within the clinical description when directly relevant to the assessment purpose. Consideration is given to the potential implications of positioning race or ethnicity in the opening sentence of an SBA stem, including the possibility of introducing unintended bias, cognitive cueing, or construct-irrelevant variance

that may affect assessment validity. An alternative approach discussed is the inclusion of demographic information, where required, later in the stem and within the context of clinical observations, rather than as an initial identifying characteristic.

USE OF DEMOGRAPHIC DESCRIPTORS IN MEDICAL EDUCATION

Demographic descriptors in medical education, such as age, sex, gender, race, ethnicity, and socioeconomic background, are routinely included in teaching materials, assessment items, and clinical case vignettes (Carey-Ewend et al., 2023). These descriptors serve multiple purposes, providing context for clinical reasoning, reflecting population-level disease patterns, and helping students practice considering patient diversity when making diagnostic or management decisions. When used appropriately, demographic information can enhance realism and support the development of culturally competent care (Rukadikar et al., 2022). However, the relevance of each descriptor should be carefully considered to avoid reinforcing stereotypes or implying causal relationships where none exist. Best practice emphasises using demographic details that are directly pertinent to the learning or assessment objective, embedding them within clinical context rather than presenting them as identity labels in isolation (Carey-Ewend et al., 2023; Nieblas-Bedolla et al., 2020).

RACE AND ETHNICITY AS DEMOGRAPHIC DESCRIPTORS

Race and ethnicity descriptors are frequently used interchangeably, despite having distinct meanings. In the UK, *race* and *ethnicity* are recognised as social rather than biological categories. Race is understood as a socially constructed concept shaped by historical power structures and racism, not genetic difference (Nazroo, 2017). Ethnicity relates to shared cultural identity, ancestry, language, and social experience, and is the preferred term in UK public health and census classification (Office for National Statistics, 2016). Differences in health outcomes between ethnic groups arise primarily from social determinants, structural inequities, and experiences of racism, rather than inherent biological variation (Kapadia et al., 2022; Public Health England, 2020).

In medical education, the association between race and disease is often introduced as part of teaching epidemiology, risk factors, and clinical reasoning. Students are frequently presented with population-level differences in disease prevalence or outcomes linked to racial or ethnic groups, such as higher rates of sickle cell disease in individuals of African ancestry or type 2 diabetes in certain South Asian populations. While this information can aid understanding of population health trends, there is increasing recognition that presenting race as a biological determinant risk reinforcing misconceptions about innate

differences, rather than highlighting the complex interplay of social, environmental, and structural determinants of health (Yudell et al., 2016; Witzig, 1996).

REPRESENTATION OF RACE AND ETHNICITY DESCRIPTORS IN QUESTION BANKS

Race and ethnicity have been incorporated inconsistently in question banks and clinical case vignettes, with some demographic characteristics overrepresented while others remain underrepresented.

Overrepresentation can occur when certain groups or traits are disproportionately highlighted in clinical scenarios, which may unintentionally signal to learners that these characteristics are inherently linked to diseases or conditions, reinforcing stereotypes, and potentially introducing cognitive cueing. Ripp and Braun's (2017) analysis of a question bank for Step 1 of the United States Medical Licensing Examination (USMLE), found that race and ethnicity were mentioned in over 20% of questions, with White/Caucasian patients heavily overrepresented relative to population demographics and many references being routine rather than clinically justified.

Conversely, there is also a notable underrepresentation of demographic diversity in medical exam questions. A "diversity audit" of 3,566 pre-clinical and clinical multiple-choice and short-answer questions across multiple medical schools revealed that, while age and binary gender were commonly represented, other diversity characteristics, particularly ethnicity, were rarely included (only 7.2% of items included ethnicity) (Mohan et al., 2025). Within the broader context of medical education, limited representation of demographic diversity in clinical case vignettes may implicitly suggest to students that cultural factors are only relevant to healthcare delivery in situations where they are directly linked to risk stratification or diagnosis (Lee et al., 2022).

This imbalance in approach raises important concerns regarding the authenticity of patient portrayals, the fairness of representation, and the potential perpetuation of racialised or stereotyped interpretations in medicine. Question banks may both overemphasise some descriptors and neglect others, highlighting the need for careful, contextually appropriate inclusion of demographic information to reflect population diversity.

PROBLEMATIC USE OF RACE AND ETHNICITY DESCRIPTORS IN QUESTION BANKS

A growing body of empirical evidence highlights the problematic use of race and ethnicity as demographical descriptors in question banks and assessment items. The key issues surrounding the inclusion of demographic descriptors in SBA items and their impact on bias, validity, and fairness in medical assessments, fall broadly into three themes.

THEME 1: MISINTERPRETATION, FALSE ASSOCIATIONS, AND BIASES

The inclusion of race and ethnicity in question banks can create false associations and affirm biases among students. Evidence from US-based question banks and medical school assessments consistently shows that race and ethnicity are frequently included in question items even when not clinically justified, sometimes as routine descriptors, sometimes associated with disease predisposition or genetic assumptions (Sam et al., 2019; Ripp and Braun, 2017; Tsai et al., 2016). In their survey of 22 medical students, Mosley et al. (2021) observed that students interpreted race in clinical vignettes as a "clue" for choosing the correct answer, with non-white racial descriptors disproportionately perceived as significant, suggesting potential reinforcement of bias in clinical reasoning.

Such patterns of inclusion and selective relevance may encourage students to link particular patient groups with specific diseases. While these associations can aid performance on examinations that depend on pattern recognition, they risk reinforcing bias and limiting students' ability to evaluate which patient information is truly pertinent in real-world clinical practice. Furthermore, these biases potentially impact assessment reliability and validity (Tavakol, Stewart and Sharpe, 2024).

Contemporary medical education therefore emphasises using race and ethnicity only when clinically relevant and avoiding racialised or essentialist interpretations, and recognising that race and ethnicity should never be used as proxies for genetics or disease risk (Bhopal, 2018; Tsai et al., 2016). Furthermore, race and ethnicity should be framed as indicators of socio-environmental exposures and systemic inequities, with attention to context, ancestry, and social determinants of health, rather than innate biological predisposition (Lett et al., 2022). These principles underpin contemporary UK guidance and are essential for ensuring fairness, accuracy, and ethical integrity in clinical assessment design.

THEME 2: STRUCTURAL PLACEMENT MATTERS

Where the inclusion of race and ethnicity descriptors is considered appropriate and relevant in SBA questions, careful consideration should be given to how and where this information is positioned within the question stem.

A mixed-method analysis of US licensing-preparation question banks demonstrate that the location of race and ethnicity within a question, stem versus answer explanation, systematically influences student interpretation of disease associations (Cerdeña, Jaswaney and Plaisime, 2021). Items embedding demographic detail in the opening sentence of the stem may prime stereotypical associations, whereas placement later in the observation or examination paragraph may mitigate such effects.

In the context of clinical case presentations, Brett and Goodman (2021), stresses that "*immediately mentioning race or ethnicity may predispose clinicians to premature diagnostic closure, a cognitive error in clinical reasoning. The subliminal effect of classifying a patient by race or ethnicity before hearing or reading about the patient's illness history and physical findings may result in incorrect inclusion or exclusion of diagnostic hypotheses*".

This highlights that not only the inclusion, but the structural placement of demographic descriptors is critical, a factor not previously emphasised for UK SBAs or national licensing assessments.

THEME 3: DOWNSTREAM CONSEQUENCES OF ASSESSMENT BIAS AND DIFFERENTIAL ATTAINMENT

Evidence indicates that racial and ethnic minority students frequently receive lower grades or less favourable evaluations in clinical clerkships and licensing assessments, even after adjusting for objective performance metrics (Nguyen et al., 2025; Hanson et al., 2022). Beyond the context of question banks, studies have shown that racial and ethnic minority students often receive disproportionately lower honours grades or less favourable written evaluations in clinical clerkships, even after adjusting for objective performance measures, suggesting systemic bias in assessment (O'Sullivan et al., 2023).

These findings indicate multifactorial bias, with exam item design being a critical but not exclusive contributor. Given the dominance of MCQ/SBA formats in high-stakes licensing, careful item construction is essential to minimise inequitable outcomes.

POSITIONING RACE OR ETHNICITY DESCRIPTORS IN AN SBA STEM

SHOULD RACE OR ETHNICITY DESCRIPTORS BE POSITIONED IN THE OPENING SENTENCE OF AN SBA STEM?

Demographic descriptors are often misapplied, structural placement influences cognitive processing and bias, and systemic inequities may be perpetuated through exam design. The implications for UK medical assessment design are clear - demographic descriptors, such as race or ethnicity, should not appear in the first sentence of an SBA stem. Introducing such details early can contribute to bias and potentially lead to misdiagnosis or inappropriate associations. Instead, these descriptors should be confined to the observation paragraph, ensuring clinical relevance and minimising cognitive biases (see Box 2). Beyond ethical considerations, the psychometric implications of demographic descriptor placement are substantial.

Box 2. Position of race or ethnicity descriptors in an SBA stem

Why using race or ethnicity descriptors in the first sentence of an SBA stem is problematic

- Cognitive dominance: First impressions strongly influence diagnostic reasoning.
- Stereotype priming: Early mention may trigger heuristic-based reasoning rather than evidence-driven clinical thinking.
- Conflation risk: Social constructs may be misinterpreted as biological or genetic proxies, particularly where race is incorrectly linked to disease risk.
- Forced-choice distortion: In SBAs, such cues can alter item functioning, reduce discriminatory validity, and advantage or disadvantage candidates unpredictably.

Why embedding race or ethnicity descriptors in the observation paragraph of a SBA stem is preferable

- Clinical relevance: Demographic information is introduced only when directly relevant (e.g., skin findings, epidemiology, or social determinants of health).
- Reflects real practice: In clinical encounters, patient background emerges gradually through history and context, rather than as a label.
- Reduces bias: Embedding reduces the risk of stereotype activation and cognitive cueing, preserving construct validity and fairness.
- Aligns with international guidance: This approach is consistent with emerging reforms in US licensing examinations and global recommendations for ethical use of demographic descriptors in medical education.

While much of the empirical evidence originates from the US, similar concerns regarding race, ethnicity, and bias in medical education and assessments have emerged across other healthcare systems, including the UK, Australia, and Canada (Mouhab et al., 2024; Marjadi et al., 2023). These findings underscore the transnational relevance of the recommendations proposed in this perspective piece. Furthermore, the UK-specific nature of the Medical Licensing Assessment (MLA) and its evolving framework presents a timely opportunity for pre-emptive reform, guided by global best practices.

While these recommendations are grounded in global evidence and psychometric theory, empirical testing within the UK MLA context is essential to assess the real-world impact on assessment fairness and bias. Future research should aim to explore the effectiveness of how the positioning of race and ethnicity in SBAs influences key psychometric indicators, including item difficulty, discrimination indices, and reliability, to ensure validity is maintained.

GLOBAL MOVES TOWARD REFORM

Licensing bodies such as the USMLE, have recognised the problematic inclusion of race and ethnicity descriptors in question banks, and have publicly committed to reviewing and removing stereotyped or biased content, and to report demographic factors, including race and ethnicity, only when self-reported by patients (USMLE, 2019).

This supports the need for clear guidance on when and how to include demographic information in UK licensing assessments. A universally applicable framework that aligns with emerging global standards for ethical and equitable medical education.

IMPLICATIONS FOR THE MLA

As the MLA becomes the national standard, item writers and regulatory bodies must adopt rigorous methodologies and embed principles of fairness, EDI, and psychometric integrity. Given the international evidence of "misuse", as well as the structural risks posed by the early placement of demographic descriptors, our recommendations outlined below (see Table 1) should form part of item-writing and quality assurance guidelines for SBAs used in UK licensing assessments.

RECOMMENDATIONS FOR ASSESSMENT DESIGNERS

The following recommendations offer evidence-informed guidance for the inclusion and positioning of demographic descriptors, particularly race and ethnicity, in single best answer questions

Table 1. Recommendations for the placement of demographic descriptors in SBAs

Recommendation	Rationale	Reference
Avoid using race as a proxy for biology or genetics	Reduces risk of racialisation of disease and misinterpretation	Ross et al., 2020; Tsai et al., 2016
Include demographic descriptors only when necessary for learning outcome	Ensures relevance and construct validity	Cerdeña, Jaswaney and Plaisime, 2021
Do not include race or ethnicity in the first sentence of the stem	Prevents cognitive priming, stereotype activation, and cueing	Tsai et al., 2016
Place demographic details within the clinical observation paragraph	Mirrors real-world patient consultation and reduces bias	Ripp and Braun, 2017; Marjadi et al., 2023
Involve EDI review panels in item development	Enhances fairness, equity, and adherence to best practice	USMLE, 2024
Provide training on race, ethnicity, and bias	Improves awareness and reduces systemic bias in assessment	Nguyen et al., 2025; Hanson et al., 2022

SUPPORT FOR ITEM WRITERS

To aid item writers, we recommend developing a clear framework for operationalising race and ethnicity in SBA items. This framework should include criteria for determining when demographic factors are clinically relevant (e.g., for conditions influenced by race/ethnicity or where the patient's social determinants of health are crucial to diagnosis). Such guidance should be incorporated into ongoing training for item developers.

CONCLUSION

While prior work documents that race and ethnicity are used (or misused) in question banks, few, if any, have addressed the specific issue of where within the item demographic details are placed (first line of stem versus observation paragraph) and the structural consequences of this placement for bias, cueing, and validity. This perspective paper extends the debate to the UK licensing context, offering a methodological and ethical framework for item writers and policymakers, rather than merely describing prevalence.

The MLA era requires that SBA items be constructed with careful attention to fairness, clarity, and ethical integrity. Race and ethnicity, as complex social constructs, should not appear as opening descriptors within SBA stems. Such placement invites bias, encourages cue-driven reasoning, and undermines validity. When integrated appropriately within clinical observations, demographic information can be used responsibly, aligning with psychometric principles and ethical standards. This approach strengthens the validity of UK licensing assessments and promotes equity for future generations of medical graduates.

As the MLA becomes the national standard for medical licensing in the UK, it is essential that assessment designers and regulatory bodies implement these recommendations to ensure fairness, clarity, and equity in medical examinations. By adhering to these guidelines, the MLA has the potential to set a precedent for ethically sound and psychometrically rigorous medical assessments.

DECLARATIONS

SPONSORS

None

CONFLICTS OF INTEREST

The author holds multiple roles: Clinical Assessment Lead at the University of Buckingham Medical School; Chair of the GMC MLA AKT Question

Management Group; GMC MLA AKT Board member, EDI Lead for the RCGP South East Thames Faculty; and Editor-in-Chief of the *Journal of Medical Education and Research*.

REFERENCES

Bhopal, R. (2018). *Migration, Ethnicity, Race and Health in Multicultural Societies*. 2nd edn. Oxford: Oxford University Press.

Brett, A.S. and Goodman, C.W. (2021). First Impressions — Should We Include Race or Ethnicity at the Beginning of Clinical Case Presentations? New England Journal of Medicine, 385(27), pp. 2497–2499, doi: 10.1056/nejmp2112312.

Cantillon, P., Irish, B. and Sales, D. (2004). 'Using computers for assessment in medicine'. *BMJ*, 329(7466), pp. 606–609, doi: 10.1136/bmj.329.7466.606.

Carey-Ewend, K., Feinberg, A., Flen, A., Williamson, C., Gutierrez, C., Cykert, S., Beck, G.L. and Gilliland, K.O. (2023). Use of sociodemographic information in clinical vignettes of multiple-choice questions for preclinical medical students', *Medical Science Educator*, 33(3), pp. 659–667, doi: 10.1007/s40670-023-01778-z.

Cerdeña, J.P., Jaswaney, R. and Plaisime, M.V. (2021). 'Race, ethnicity, and immigration status in a medical licensing educational resource: A systematic, mixed-methods analysis'. *Journal of General Internal Medicine*, 37(5), pp. 1045–1051, doi: 10.1007/s11606-021-06843-0.

General Medical Council. (2024). UK medical schools guide to the MLA. [online] Available at: https://www.gmc-uk.org/education/medical-licensing-assessment/uk-medical-schools-guide-to-the-mla.

Hanson, J.L., Pérez, M., Mason, H.R.C., Aagaard, E.M., Jeffe, D.B., Teherani, A. and Colson, E.R. (2022). 'Racial/ethnic disparities in clerkship grading: Perspectives of students and teachers', *Academic Medicine*, 97(11S), pp. S35–S45, doi: 10.1097/acm.0000000000004914.

Kapadia, D. et al. (2022) *Ethnic Inequalities in Healthcare: A Rapid Evidence Review*. London: NHS Race and Health Observatory.

Lee, C.R., Gilliland, K.O., Beck, G.L. and Tolleson-Rinehart, S. (2022). 'Race, ethnicity, and gender representation in clinical case vignettes: A 20-year comparison between two institutions', *BMC Medical Education*, 22(1), doi: 10.1186/s12909-022-03665-4.

Lett, E., Asabor, E., Beltrán, S., Cannon, A.M. and Arah, O.A. (2022). 'Conceptualizing, contextualizing, and operationalizing race in quantitative health sciences research'. *The Annals of Family Medicine*, 20(2), pp. 157–163, doi: doi.org/10.1370/afm.2792.

Marjadi, B., Chiavaroli, N., Sorinola, O., Nymberg, V.M., Joyce, C., Parsons, C. and Ryan, A. (2023). 'Diversity audit of medical school examination questions', *Teaching and Learning in Medicine*, 36(5), pp. 557–565, doi: 10.1080/10401334.2023.2240776.

Miller, G.E. (1990). 'The assessment of clinical skills/competence/performance', *Academic Medicine*, 65(9), pp. S63–7, doi: 10.1097/00001888-199009000-00045.

Mohan, J., Pishia, M., Alim, D., McGown, P.J., Amin, A., Pinder, R. J., Halse, O., Brown, C. and Sam, A. H., (2025). 'The impact of ethnicity information in single best answer questions on student performance, item response time and psychometric properties', *Advances in Health Sciences Education: Theory and Practice*, doi: 10.1007/s10459-025-10489-6.

Mouhab, A., Radjack, R., Moro, M. R., & Lambert, M. (2024). 'Racial biases in clinical practice and medical education: a scoping review', *BMC Medical Education*, 24(1), doi: 10.1186/s12909-024-06119-1.

Nazroo, J.Y. (2003). 'The structuring of ethnic inequalities in health: Economic position, racial discrimination, and racism', *American Journal of Public Health*, 93(2), pp. 277–284, doi: 10.2105/AJPH.93.2.277.

Nieblas-Bedolla, E., Christophers, B., Nkinsi, N.T., Schumann, P.D. and Stein, E. (2020). 'Changing how race is portrayed in medical education: Recommendations from medical students', *Academic Medicine*, 95(12), doi: 10.1097/ACM.0000000000003496.

Norcini, J.J., Swanson, D.B., Grosso, L.J. and Webster, G.D. (1985). 'Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence', *Medical Education*, 19(3), pp. 238–247. doi: 10.1111/j.1365-2923.1985.tb01314.x.

Nguyen, M., Abrams, G., Hodgens, T., Curry, R.H., Kim, M.H., Betchkal, R., Low, C., Jain, S., Tarlanov, S., Eidtson, W.H., Coates, C., Weiner, C., Fancher, T. and Meeks, L.M. (2025). 'Race and disability characteristics and accommodation disparities on the USMLE Step 1'. *JAMA Network Open*, 8(9), p. e2534621, doi: 10.1001/jamanetworkopen.2025.34621.

Office for National Statistics. (2016). *Ethnic group, national identity and religion - Office for National Statistics. [online] Available at: https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion*.

Okubuiro, E. O., Longinus N. E., and Okoli, C. E. (2019). 'Utility of single-best-answer-questions as a summative assessment tool in medical education: A review'. *International Journal of Recent Innovations in Academic Research*, 3(1), pp. 1–12.

O'Sullivan, L., Kagabo, W., Prasad, N., Laporte, D. and Aiyer, A. (2023). 'Racial and ethnic bias in medical school clinical grading: A review'. *Journal of Surgical Education*, 80(6), pp. 806–816, doi: 10.1016/j.jsurg.2023.03.004.

Parekh, P. and Vikesh Bahadoor (2024). 'The utility of multiple-choice assessment in current medical education: A critical review', Curēus, 16(5), doi: 10.7759/cureus.59778.

Ross, P.T., Hart-Johnson, T., Santen, S.A. and Zaidi, N.L.B. (2020). Considerations for using race and ethnicity as quantitative variables in medical education research. *Perspectives on Medical Education*, 9(5), pp. 318–323, doi: 10.1007/s40037-020-00602-3.

Rukadikar, C., Mali, S., Bajpai, R., Rukadikar, A. and Singh, A. (2022). 'A review on cultural competency in medical education', *Journal of Family Medicine and Primary Care*, 11(8):p 4319-4329, doi: 10.4103/jfmpc.jfmpc_2503_21.

Public Health England (2020). *COVID-19: review of disparities in risks and outcomes. [online] Available at: https://www.gov.uk/government/publications/covid-19-review-of-disparities-in-risks-and-outcomes.*

Ripp, K. and Braun, L. (2017). 'Race/ethnicity in medical education: An analysis of a question bank for Step 1 of the United States Medical Licensing Examination', *Teaching and Learning in Medicine*, 29 (2), pp. 115–122, doi: 10.1080/10401334.2016.1268056.

Sam, A. H., Westacott, R., Gurnell, M., Wilson, R., Meeran, K., and Brown, C. (2019). 'Comparing single best-answer and very short answer questions for the assessment of applied medical knowledge in 20 UK medical schools: Cross-sectional study', *BMJ Open*, 9(9), pp. e032550, doi: 10.1136/bmjopen-2019-032550.

Schuwirth, L.W.T. and Van der Vleuten, C.P.M. (2011). 'Programmatic assessment: From assessment of learning to assessment for learnin', *Medical Teacher*, 33(6), pp. 478–485, doi: 10.3109/0142159x.2011.565828.

Tavakol, Stewart, C. and Sharpe, C.C. (2024). 'Ensuring fairness in assessment in health professions education: rapid analysis tools to detect differential item functioning across groups', *International Journal of Medical Education*, 15, pp. 80–83. doi: 10.5116/ijme.6694.de69.

Tsai, J., Ucik, L., Baldwin, N., Hasslinger, C. and George, P. (2016). 'Race matters? Examining and rethinking race portrayal in preclinical medical education', *Academic Medicine*, 91(7), pp. 916–920, doi: 10.1097/acm.0000000000001232.

USMLE (2019) *USMLE review focuses on patient individuality.* [online] Available at: https://www.usmle.org/usmle-review-focuses-patient-individuality.

USMLE (2024) *USMLE Program Updates: Diversity and Equity Initiatives*.

Van der Vleuten, C. (2000). 'Validity of final examinations in undergraduate medical training', *BMJ*, 321(7270), pp. 1217–1219, doi: 10.1136/bmj.321.7270.1217.

Witzig, R. (1996). 'The medicalization of race: Scientific legitimization of a flawed social construct', *Annals of Internal Medicine*, 125(8), p. 675, doi: 10.7326/0003-4819-125-8-199610150-00008.

Yudell, M., Roberts, D., DeSalle, R. and Tishkoff, S. (2016). 'Taking race out of human genetics', *Science*, 351(6273), pp. 564–565. doi: 10.1126/science.aac4951.

Zegugu, M., Chatoo, S.A., Choudhry, A., Metezai, H., Usman, D., Kamal, M. and El Badawey, A. (2024). 'Single Best Answer questions as a teaching tool in medical education: An international mixed-method comparative evaluation. Cureus, 16(9), p. e69745, doi: 10.7759/cureus.69745.