

Letters to the Editor

Using Video Technology to Improve Fairness in Objective Structured Clinical Examination Scoring

To the Editor: We read with great interest the article by Yeates and colleagues¹ addressing the impact of examiner variability on the Objective Structured Clinical Examination (OSCE) scores and using novel technology to combat this. As fourth-year medical students from Imperial College London, this article particularly resonated with us, and we encourage changes to improve the fairness of OSCE scoring.

We support the implementation of Video-based Examiner Score Comparison and Adjustment (VESCA) because it not only enhances fairness in OSCE scoring but it also has the potential to revolutionize the structure of the OSCE. As the Internet system is further developed, it could eventually negate the need for a live examiner. During our recent OSCE, we felt that examiners with favorable facial expressions and open body language may have reduced stress and helped students perform better than they would have otherwise. Although management of nonverbal cues may be covered in training, we believe that remote Internet scoring would be a more effective solution.

Conversely, VESCA does not take into consideration Simulated Patients' (SPs) behavior and portrayal of the condition. SPs' physical and verbal portrayal contributes significantly to how candidates interact with them. Tamblin and colleagues² found that SPs who received 2 training sessions depicted the condition more accurately than those who received 1. We have experienced how SPs can affect candidate scores, as they also grade students during the OSCE. Therefore, SPs must be trained regularly and be assessed during the OSCE. SP regulations, in combination with VESCA, could improve the overall fairness of the OSCE.

To further the study and encourage future implementation of VESCA, we suggest adopting an alternative approach to participant recruitment. Voluntary participation for VESCA raises the question of whether students who were more confident in their abilities opted to

participate. We suggest that an opt-out approach could be adopted to increase participant numbers and result in a more accurate depiction of the original cohort. In fact, it may be of interest to study if borderline students perform better with unobtrusive cameras rather than a live examiner in the room.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

Ailin Anto
Medical student, School of Medicine, Imperial College London, London, United Kingdom; ailin.anto18@imperial.ac.uk; ORCID: <https://orcid.org/0000-0002-4024-2237>.

Monica Kelada
Medical student, School of Medicine, Imperial College London, London, United Kingdom; ORCID: <https://orcid.org/0000-0002-0993-901X>.

Arunima Basu
Medical student, School of Medicine, Imperial College London, London, United Kingdom; ORCID: <https://orcid.org/0000-0002-8475-7275>.

The authors have informed the journal that they agree that A. Anto, M. Kelada, and A. Basu completed the intellectual and other work typical of the first author.

References

- 1 Yeates P, Moulton A, Cope N, et al. Measuring the effect of examiner variability in a multiple-circuit objective structured clinical examination (OSCE). *Acad Med.* 2021;96:1189–1196.
- 2 Tamblin RM, Klass DK, Schanbl GK, Kopelow ML. Factors associated with the accuracy of standardized patient presentation. *Acad Med.* 1990;65(suppl 9): S55–S56.

In Reply to Anto et al: We thank the authors for their comments on our article. We agree that the prospect of using video technology to remove the examiner from the room is an interesting proposition with the potential to reduce cueing of the candidate by the examiner.

Gormley and colleagues¹ have described the theatrical nature of the triadic interaction, which can occur between the examiner, candidate, and simulated patient; we agree that removing the examiner could therefore aid the clinical authenticity of the objective structured clinical examinations (OSCEs) and

recommend further research on the topic. Any such intervention would, however, require filming many more students than occurred in our study using Video-based Examiner Score Comparison and Adjustment (VESCA)—we videoed just 6 students—although Chen and colleagues² have successfully filmed OSCEs on a larger scale before. Moreover, if the VESCA intervention were to be used, we would recommend adherence to the principles of filming OSCEs, which we empirically developed to minimize intrusion while maximizing examiners' ability to make sufficiently informed judgments.³

We agree with the authors' latter 2 points: VESCA relies on uniformity by simulated patients, a limitation of VESCA that also warrants further investigation in due course, and randomization would enhance the selection of students for videoing. With respect to the second point, we used VESCA in a research setting, so voluntary participation was paramount. With routine use in assessment, it seems likely that randomization could (and should) be employed.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

Disclaimers: The views expressed are those of the author(s) and do not necessarily represent those of the National Health Service (NHS), the National Institute for Health Research (NIHR), or the U.K. Department of Health and Social Care.

Peter Yeates, MRCP, PhD
Senior lecturer in medical education research, School of Medicine, Keele University, Keele, Staffordshire, and a consultant in acute and respiratory medicine, Fairfield General Hospital, Pennine Acute Hospitals, NHS Trust, Bury, Lancashire, United Kingdom; p.yeates@keele.ac.uk; ORCID: <https://orcid.org/0000-0001-6316-4051>.

Richard Fuller, MA, FRCP
Director of Christie Education, Christie Hospitals NHS Trust/University of Manchester, Manchester, United Kingdom; ORCID: <https://orcid.org/0000-0001-7965-4864>.

Robert K. (Bob) McKinley, MD, FRCGP, FRCP
Emeritus professor of education in general practice, School of Medicine, Keele University, Keele, Staffordshire, United Kingdom; ORCID: <https://orcid.org/0000-0002-3684-3435>.

References

- 1 Gormley GJ, Hodges BD, McNaughton N, Johnston JL. The show must go on? Patients,

- props and pedagogy in the theatre of the OSCE. *Med Educ.* 2016;50:1237–1240.
- 2 Chen TC, Lin MC, Chiang YC, Monrouxe L, Chien SJ. Remote and onsite scoring of OSCEs using generalisability theory: A three-year cohort study. *Med Teach.* 2019;41:578–583.
- 3 Yeates P, Moulton A, Lefroy J, et al. Understanding and developing procedures for video-based assessment in medical education. *Med Teach.* 2020;42:1250–1260.

The Sky Should Be the Limit: USMLE Attempt Limits Will Not Reduce Disciplinary Actions Among Graduates of U.S. and Canadian Medical Schools

To the Editor: We are writing to offer a reasoned critique to Arnhart and colleagues' article.¹ We object to the article's conclusion, especially as stated in the abstract:

This study offers support for licensure and practice standards to account for physicians' USMLE [United States Medical Licensing Examination] attempts. The relatively small effect sizes, however, caution policy makers from placing sole emphasis on this relationship.

We believe a conclusion supported by the data they presented might read, "This study demonstrated that there is a small but significant relationship between USMLE attempts and disciplinary action. However, there is still insufficient evidence to support limiting USMLE attempts for the purpose of reducing disciplinary actions." The methods the authors describe do not account for the many confounding variables and factors that affect human behavior.

We are particularly concerned that Arnhart and colleagues did not discuss the reasons for multiple testing attempts, which might include health issues, personal dilemmas, and other unexpected or uncontrollable life events. Nor did the authors offer any discussion of the possible reasons for disciplinary action that may be unrelated to the design and purpose of the USMLE Step exams. Disciplinary issues of incompetence, inappropriate behavior, substance abuse, fraud, and malpractice were all lumped together, thus obfuscating any attempt to make sense of the identified relationship.

We are also troubled that the authors proposed policy while the results only indicated an association—the authors

themselves cautioned their readers about inferring a causal relationship. Even with the caveats listed in the abstract (and the long list of limitations in the paper itself), it is incorrect, misleading, and therefore irresponsible to even hint that this study supports limiting USMLE attempts. Suggesting that we substitute the number of examination attempts for scores to estimate the likelihood of disciplinary action is unsupported by this study. There may be other justifications for such a policy, but limiting disciplinary actions is not one of them.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

Marcel F. D'Eon, MEd, PhD
Director, Education Innovation Institute of the Medical College of Georgia, Augusta University, Augusta, Georgia; mdeon@augusta.edu.

A.J. Kleinheksel, MEd, PhD
Educational researcher, Education Innovation Institute of the Medical College of Georgia, Augusta University, Augusta, Georgia.

Reference

- 1 Arnhart K, Cuddy M, Johnson D, Barone M, Young A. Multiple United States Medical Licensing Examination attempts and the estimated risk of disciplinary actions among graduates of U.S. and Canadian medical schools. *Acad Med.* 2021;96:1319–1323.

In Reply to D'Eon and Kleinheksel:

We thank the authors for their critique and for highlighting the important role medical licensing examinations serve in the regulation of physician practice. Our article outlines existing United States Medical Licensing Examination (USMLE) and state medical board attempt limit policies and indicates that our "findings provide some support for the use of attempt limits as suggested by the USMLE program and implemented by state medical boards."¹ We want to emphasize that although our findings support a relationship between multiple USMLE attempts and increased likelihood of receiving disciplinary actions, the findings in isolation are not sufficient for proposing new policy on how many attempts should be allowed. We view our research as one of many factors for policy makers to consider as part of their broader decision-making process when reviewing minimum standards for medical licensure.

Although we did not discuss specific extenuating circumstances that may cause competent medical students and physicians to take the USMLE multiple times, we did acknowledge the legitimacy for multiple attempts by saying, "Attempt limit policies aim to balance providing individuals sufficient opportunities for demonstrating competence with protecting the integrity of the examination."¹ Though any of the extenuating circumstances noted in D'Eon and Kleinheksel's critique might be true (e.g., health issues, personal dilemmas), their cause for necessitating successive repeated attempts would be less likely over time.

D'Eon and Kleinheksel also indicate concern that disciplinary actions were measured dichotomously and failed to pick up nuances of why physicians received actions. In the Discussion of our article, we noted that exploring the number of USMLE attempts in relation to the risk of receiving a disciplinary action by type and severity would be a productive topic for future research. Furthermore, it is important to note we made efforts to focus on actions associated with problematic behavior by omitting administrative and other minor board actions in our analysis.

We understand the caution raised by the authors about a direct substitution of the number of examination attempts and examination scores to predict the likelihood of disciplinary action, as these 2 factors provide different types of information—even though there may be some overlap. Our intention is to consider that the number of USMLE attempts may provide one of several useful measures for understanding physician performance, and our results support the use of existing attempt limit policies as opposed to no limit at all.

Funding/Support: None reported.

Other disclosures: The authors are employed by the Federation of State Medical Boards (FSMB) and the NBME, which are cosponsors of the United States Medical Licensing Examination (USMLE).

Ethical approval: Reported as not applicable.

Katie L. Arnhart, PhD
Research project manager, Research and Data Integration, Federation of State Medical Boards, Euless, Texas; karnhart@fsmb.org; ORCID: <http://orcid.org/0000-0001-9975-6358>.

Monica M. Cuddy, PhD
Measurement scientist, Office of Research Strategy, NBME, Philadelphia, Pennsylvania; ORCID: <http://orcid.org/0000-0002-5756-9113>.