Letter

What if We Replace Pass/Fail Grading with Decision Making by Machine? Toward Competency-based Medical Education

Haniye Mastour [1], Toktam Dehghani [2, *] and Saeid Eslami [2, 3, 4]

[1]Department of Medical Education, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
[2]Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
[3]Pharmaceutical Sciences Research Center, Institute of Pharmaceutical Technology, Mashhad University of Medical Sciences, Mashhad, Iran
[4]Department of Medical Informatics, Amsterdam UMC (location AMC), University of Amsterdam, Amsterdam, the Netherlands

Corresponding author: Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. Email: dehghanit982@mums.ac.ir

Dear Editor,

Today a movement toward competency-based medical education (CBME), as an outcome-based method compared to the traditional time-based medical training method (1), is required. Worldwide adoption of competency-based assessment to guarantee the achievement and maintenance of competence will require evaluations, both during training and in unsupervised practice. The CBME needs many alterations and adaptations in academic health sciences systems. Although a series of medical high-stakes examinations are administered by the different national boards of medical examiners worldwide during medical education and even after completion for various reasons such as promotion to a higher year, graduation, obtaining a certificate, etc., it could be challenging to sustain a single and consistent approach for the decision-making process in such high volumes and repeatability. Therefore, achieving high consistency, low delays, resolving bottlenecks, and competency-based assessment in this process must be considered. Currently, the question is how much these pass/fail grading tests can evaluate a medical student's or physician's medical knowledge and clinical competencies. While using exam scores as a single measure has been widely critiqued (2).

Artificial intelligence (AI) is a field that allows medical education stakeholders to assess participants from different aspects of knowledge, skills, and even attitudes. With AI breakthroughs in various academic and education areas, such as tutoring and assessment, new point of view and novel methods have emerged. It integrates large quantities of unstructured or structured big data, calculates and demonstrates complex logical relations and models, and predicts outcomes in different fields of medical educa-

tion. Machine Learning (ML) algorithm quantifies performance and improves the granularity and precision of the classification of bimanual technical skills (3). Using real-time evaluations and interventions, ML techniques may analyze and integrate pragmatic data (4). There are many data structures and sources (like text data derived from consultation, index data derived from laboratory examinations, and medical image data derived from additional tests, etc.), which means heterogeneous and multimodal phenomena (5). The increasing complexity of medical decisions due to the plethora of data sources and the need to involve numerous specialists has shifted decision-making from a physician's process to a collective process made in multidisciplinary ways (6). As a result, accurate and objective performance assessment is essential for medical students, clinicians, and even certified physicians to be better prepared to advise patients (7, 8). However, current methods can be labor-intensive, time-consuming, and subject to bias. Once ML is designed and trained, it could provide immediate, automated, and reproducible feedback without needing expert reviewers (9). So various classical AI and ML methods, such as Artificial Neural Networks, Fuzzy and Genetic Algorithms, and novel techniques, such as Natural Language Processing, Text Mining, Agent-Based, and Deep Learning, can be used to achieve related purposes (10). According to Dias et al., the six Accreditation Council for Graduate Medical Education (ACGME) core competencies to which ML techniques for competence assessment could be applied are shown in Table 1 (4). In their systematic review, Lam et al. declared that ML could produce accurate and objective surgical skill assessments. They reviewed 66 from 1896 studies that were retrieved. The accuracy rates of ML methods were over 80%; however, participants and

Table 1. ACGME's Competencies Definition and Application of ML Techniques for Assessment (4)

ACGME's Core Competencies	Definition	ML Application for Competence Assessment
Patient care	This domain provides patient-oriented care that is proper, compassionate, and useful to treat health problems and promote health.	Extraction of data from electronic health records or motion tracking technologies (e.g., wearable devices); Text-mining techniques.
Medical knowledge	This domain is associated with knowledge of evolving and established clinical, biomedical, social-behavioral, and epidemiological sciences and the application of such knowledge in patient care.	Pattern extraction from non-structured data and assessing their relationship with an expert-based assessment (e.g., natural language processing to evaluate medical knowledge with data obtained from diagnostic reports, clinical notes, verbal communications, and written responses).
Interpersonal and communication skills	This domain indicates a physician's communication skills and interpersonal proficiency with a patient, his/her family, and health personnel.	Expression or facial recognition; Speech recognition (sentiment assessment); Gestures, gaze, or pose tracking; Objective measurement of various emotions and behaviors (engagement, nonverbal cues, stress, attention, empathy, and frustration).
Professionalism	This domain demonstrates the commitment to performing professional duties and compliance with ethical principles.	Extracting data from surveys, self-assessments, and patient-doctor conversations (audio and transcribed data).
Practice-based learning and improvement	This domain describes students'/physicians' ability to assess patient care and scientific evidence and constantly increase patient care according to continuous self-assessment and lifelong learning.	Extract relevant themes associated with physician performance; Pattern identification of clinical communication among interdisciplinary teams during handoffs.
System-based practice	This domain includes awareness, being responsive to the larger context and the healthcare system, and being able to efficiently be involved in other resources within the system to offer optimum healthcare.	Extract complex behavior patterns that humans could not observe alone; Speech recognition, and computer vision to assess physician knowledge, skills, and behavior.

tasks varied between studies (9). Nagaraj et al. developed and trained AI models to identify errors and provide automated evaluation and coaching for medical students. Two hundred sixteen suturing and knot-tying videos were used to train the models. The accuracy of the instrument holding and the knot-tying model was 89% with an F-1 score of 74% and 91% with an F-1 score of 54%, respectively (11). Generally, the accessibility and flexibility of ML methods have resulted in various successful applications in medical education and assessment (12-15).

To sum up, applying the intelligence processes by machines in medical education will help prepare tomorrow's physicians, and advances in the role of AI-driven methods indicate that they could inevitably transform previously accepted assessment models. During COVID-19 pandemic, we have moved towards a new technology-enhanced assessment, and this paradigm is supposed to be followed. Future ML-based competency assessment tools should move beyond evaluating basic tasks and towards objective structured examination and provide interpretable effective feedback with clinical value for medical students or physicians. Further research is required to integrate AI trends and the possible positive impacts of these innovative programs in evaluating medical students or physicians.

References

1. Kealey A, Naik VN. Competency-Based Medical Training in Anesthesiology: Has It Delivered on the Promise of Better Education? *Anesth Analg*. 2022;**135**(2):223–9. [PubMed ID: 35839492]. https://doi.org/10.1213/ANE.0000000000006091.

2. Burk-Rafel J, Reinstein I, Feng J, Kim MB, Miller LH, Cocks PM, et al. Development and Validation of a Machine Learning-Based Decision Support Tool for Residency Applicant Screening and Review. *Acad Med*. 2021;**96**(11S):S54–61. [PubMed ID: 34348383]. https://doi.org/10.1097/ACM.0000000000004317.

3. Fazlollahi AM, Bakhaidar M, Alsayegh A, Yilmaz R, Winkler-Schwartz A, Mirchi N, et al. Effect of Artificial Intelligence Tutoring vs Expert Instruction on Learning Simulated Surgical Skills Among Medical Students: A Randomized Clinical Trial. *JAMA Netw Open*. 2022;**5**(2). e2149008. [PubMed ID: 35191972]. [PubMed Central ID: PMC8864513]. https://doi.org/10.1001/jamanetworkopen.2021.49008.

4. Dias RD, Gupta A, Yule SJ. Using Machine Learning to Assess Physician Competence: A Systematic Review. *Acad Med*. 2019;**94**(3):427–39. [PubMed ID: 30113364]. https://doi.org/10.1097/ACM.0000000000002414.

5. Wei L, Yu Z, Qinge Z, Jan MA. Medical College Education Data Analysis Method Based on Improved Deep Learning Algorithm. *Mob Inf Syst*. 2022;**2022**:1–9. https://doi.org/10.1155/2022/3227316.

6. Drummond D. Between competence and warmth: the remaining place of the physician in the era of artificial intelligence. *NPJ Digit Med*. 2021;**4**(1):1–4. [PubMed ID: 33990682]. [PubMed Central ID: PMC8121897]. https://doi.org/10.1038/s41746-021-00457-w.

7. Blease C, Kharko A, Bernstein M, Bradley C, Houston M, Walsh I, et al. Machine learning in medical education: a survey of the experiences and opinions of medical students in Ireland. *BMJ Health Care Inform*. 2022;**29**(1). [PubMed ID: 35105606]. [PubMed Central ID: PMC8808371]. https://doi.org/10.1136/bmjhci-2021-100480.

8. Kolachalama VB. Machine learning and pre-medical education. *Artif Intell Med*. 2022;**129**:102313. [PubMed ID: 35659392]. https://doi.org/10.1016/j.artmed.2022.102313.

9. Lam K, Chen J, Wang Z, Iqbal FM, Darzi A, Lo B, et al. Machine learning for technical skill assessment in surgery: a systematic review. *NPJ Digit Med*. 2022;**5**(1):24. [PubMed ID: 35241760]. [PubMed Central ID: PMC8894462]. https://doi.org/10.1038/s41746-022-00566-0.

10. Chassignol M, Khoroshavin A, Klimova A, Bilyatdinova A. Artificial Intelligence trends in education: a narrative overview. *Procedia Comput Sci*. 2018;**136**:16–24. https://doi.org/10.1016/j.procs.2018.08.233.

11. Nagaraj MB, Namazi B, Sankaranarayanan G, Scott DJ. Developing artificial intelligence models for medical student suturing and knot-tying video-based assessment and coaching. *Surg Endosc*. 2022:1–10. [PubMed ID: 35982284]. [PubMed Central ID: PMC9388210]. https://doi.org/10.1007/s00464-022-09509-y.

12. Yanik E, Intes X, Kruger U, Yan P, Diller D, Van Voorst B, et al. Deep neural networks for the assessment of surgical skills: A systematic review. *J Def Model Simul*. 2021;**19**(2):159–71. https://doi.org/10.1177/15485129211034586.

13. Ward TM, Mascagni P, Madani A, Padoy N, Perretta S, Hashimoto DA. Surgical data science and artificial intelligence for surgical education. *J Surg Oncol*. 2021;**124**(2):221–30. [PubMed ID: 34245578]. https://doi.org/10.1002/jso.26496.

14. Rogers MP, DeSantis AJ, Janjua H, Barry TM, Kuo PC. The future surgical training paradigm: Virtual reality and machine learning in surgical education. *Surgery*. 2021;**169**(5):1250–2. [PubMed ID: 33280858]. https://doi.org/10.1016/j.surg.2020.09.040.

15. Luan H, Tsai C. A review of using machine learning approaches for precision education. *J Educ Techno Soc*. 2021;**24**(1):250–66.