Artificial Intelligence in Medical OSCEs: Reflections and Future Developments

Tse Kiat Soong1,2
Cheng-Maw Ho1

1Department of Surgery, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan; 2Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Abstract: With the advent of the age of Artificial Intelligence (AI), we seek to consider how AI could shape clinical examinations, specifically Objective Structured Clinical Examinations (OSCEs). OSCEs, whilst having its own limitations, could be further enhanced with new technologies like AI to help better assess and prepare our future clinicians. With the everchanging requirements on the modern clinician, we deliberate the strengths and weaknesses of AI, and the need for emphasis on different skills to complement rather than resist the tides of change. In conclusion, we feel that AI has the potential to be a strong driving force in remodelling OSCEs to support future doctors and could serve as a new frontier in medical education and beyond. That being said, we recognize the technology and its applications are still in its infancy, and further study will be needed to elucidate the role of AI in medical education and in the greater landscape of medical practice.

Keywords: medical education, assessment, perspective, future doctor

Introduction

As the age of Artificial Intelligence (AI) begins, an increasing number of industries are exploring the applications and possible ramifications of this new era on current practice.1 In that sense, medicine and healthcare is the same; there have been numerous commentaries regarding the increasing role of AI in healthcare, with both proponents and opponents for the further application of AI and deep machine learning in clinical practice.2,3 Various arguments have been cited, with advocates citing the efficiency and accuracy of AI in easing the clinical workload, given the fact that network-based interfaces allow for knowledge and information to be shared across servers to allow for shared mastery of the fields programmed and learned by the machines.2,4,5 However, critics ruminate the possibility of a bleak future of employment in the healthcare industry and ethical dilemmas where responsibility for medical errors are sought.4,6 However, much of these remain conjecture, for there has been minimal research and study into the application of this rapidly progressing technology in medical practice.

While the future remains to be seen, ranging from a completely replaced healthcare workforce to one where AI plays a supportive role of guiding clinical practice, or anywhere in between, AI would undeniably be playing a larger role in clinical practice in the future as compared to now.5 Therefore, the question arises—what skillsets and mindsets are required by clinicians in this new era, and are the current evaluation methods employed adequate in assessing the preparedness of our candidates?
Objective Structured Clinical Examinations (OSCEs) are a dominant assessment tool in healthcare education, allowing educators to assess practical performance reliably. The goal is to prepare medical students for practice-based learning and to train and test competence under standardized conditions. However, these standardized settings may artificially simplify the complexity of nonstandard, authentic patient encounters in real clinical environments. Clinical communication is reduced to tick-boxes on an examiner’s checklist, and learners simply strive to demonstrate behaviors under time pressure in pursuit of marks. OSCEs, in its current form, risk being just a barrier to gaining the title of “doctor,” rather than of being able to truly assess their ability to practice as a doctor. Given the changing climate, the objective of education should and will need to refocus back to patient care with greater emphasis than before.

AI applications in higher education can broadly be classified into four domains: profiling and prediction, assessment and evaluation, adaptive systems, and personalization, and intelligent tutoring systems. Zawacki-Richer et al has shown in their systematic review that AI applications can perform assessment and evaluation with great accuracy and efficiency. For example, Sanchez et al used an algorithm to match students to the professional competencies and capabilities required by companies, in order to ensure concordance between courses and industry needs. From perusal of the available literature, while many authors have discussed the diverse characteristics, skills, and knowledge with the advent of AI and the gaps in the current medical education framework, few, if any, have discussed how exams will be influenced by developments in these technologies. Thus, we seek to consider the various possibilities of how AI could shape clinical examinations, specifically OSCEs, in terms of changes in design, curricular relevance, and methods of appraisal or assessment of these new skillsets, with or without the use of AI.

The “Evolving” Modern Clinician

Demands on the modern clinician are ever-changing. While the clinicians of yesterday might have been expected to diagnose various complex conditions solely based off a simple consultation, examination, and rudimentary tests, clinicians of today have, at their disposal, advanced tools (including searches on a trusted medical information sites or phone applications) to aid their decision-making and diagnosis. Thus, the skillsets and problems faced by clinicians of different times vary greatly, and while some may be of the same flavor, these often take on very different contexts for the clinician to overcome. As the mercurial tides of medical reform are creeping up on us, the skillsets and demands on the clinician will continue to change.

The Switch from a Knowledge-Intensive Field

In most places of the world, information flows more freely than ever before, with the increasing connectivity of the world and the decreased barriers to information. Gone were the days where the only source of information was from pouring over tomes of medical texts for a single piece of information; a simple search on a preferred search engine or literature site would likely yield answers to the question that was sought after. With the evolution of medical technology and advances in our understanding of medical conditions, the knowledge and information for all fields grow vastly, hopefully yielding more and more answers to the questions that were asked of yesterday and tomorrow. However, we live in paradoxical times—one where the ease of access to information does not necessitate greater knowledge, owing to the burgeoning knowledge base and limited human capacity. This limited human capacity is addressed by that of advancements in AI; an AI system’s knowledge is limited not by capacity, for servers and storage space can be easily expanded, but by the progress of our own human knowledge, given that AI (in its current state for diagnostics) can only learn what is fed to it and draw logical deductions from the conclusions fed to it by human researchers and clinicians. This state of omniscience (of available knowledge) that AI possesses shows the advantage of AI over humans in terms of knowledge, and one that humans are unlikely to surpass.

This incongruity brings into focus one of the key changes in the skillset of the modern clinician: that of the increasing importance of correct knowledge capture over knowledge retention, as effectively discussed by Wartman and Combs. Given that a modern clinician is highly unlikely to surpass an AI system in terms of the amount of knowledge it drives, it is thus important that a clinician is broad-based, covering more breadth than depth. Of essence, however, is the ability to find reliable sources of information and knowing how to interpret and apply the information sought. This should thus prompt shifts in medical curricular planning to focus on the salient features of the above and to ensure that future clinicians have this skillset in their arsenal.
Use of AI Systems – Design, Interpretation, and Application

The implementation of AI into the medical field is an unstoppable force set to revolutionize the current landscape, for better or for worse. Thus, a new breed of clinician must be trained, one where doctors play a critical role in the design and planning of these systems and the direction that AI would evolve in, and in anticipating healthcare needs of the future. This necessitates knowledge in the domains of coding, big data, and user interface planning, which we believe would be core skills of the future. These clinical designers would combine clinical experience and knowledge and apply these principles in the design of AI systems which can then be extrapolated into more complex applications. The multidisciplinary team would also likely evolve to include computer or data scientists, to provide expertise on matters regarding these AI systems.

While design is one important aspect of these AI systems, another important facet would be the interpretation and application of what these AI systems provide; in other words, how can we make use of the predictions and recommendations of AI efficiently in clinical practice? This brings us to a core skill in the arsenal of the modern clinician–data interpretation and translation into clinical practice. AI, in its current form, draws off past examples and the previously reached conclusions to structure and guide its future decisions. Decisions, diagnosis, and management are recommended based off the input of signs and symptoms of the informant, where that is the patient or the clinician. This, while infinitely helpful for common, recurring conditions, can prove to be our drawback in the event of new diseases and infections. While AI is able to determine new phenotypes and act as decision aids as to when to start resuscitation or certain supportive and life-saving managements, deep learning has a significant disadvantage in the way it functions; it requires data (large amounts of it due to clinical diversities, in fact) to draw conclusions before proceeding. Thus, we postulate that clinical acumen, contrary to what some might argue, is more important than ever, for while AI can generate more differentials (with confidence intervals) more accurately than a human, humans are still important to discern which patient has the common cold, and which needs to be isolated for a potentially species-threatening new virus. This is likely why AI is still unlikely to replace clinicians in the foreseeable future.

A modern clinician should, thus, have excellent clinical acumen, be able to discern and categorize the various complaints of a patient, and be able to utilize the aid of AI in making clinical decisions, without the algorithms replacing the clinician reasoning process.

The Human Touch

“To cure sometimes, to relieve often, and to comfort always.” This central tenet (or dare I say, central dogma) of modern medicine is echoed by clinicians worldwide in medical education. This simple yet deeply profound saying epitomizes human touch; that of compassion and empathy and understanding of another human being in suffering. The human touch is arguably essential in the field of medicine, which is often said to be both an art and a science. While efficiency and accuracy might arguably be improved by the implementation of AI systems, a dearth yet to be addressed has surfaced. AI systems have yet to, nor are expected to, fully replicate the human touch that clinicians can provide. While AI interfaces can offer simple lines expressing empathy or compassion, the absence of true emotion before these neural interfaces bequeath the main issue—that the human touch cannot be replaced by AI. This aspect of AI, considering its unstoppable nature of integration into medical care, has made communication skills and these soft skills of compassion and empathy ever more important, and is something that should be strongly honed by clinicians in this era of supposed “replacement of roles” by technologies and a shifting focus from the patient to the screen.

OSCEs in the Era of AI: Exam Conduct, Simulated Patient, Clinician

The OSCE is a widely used clinical examination for the assessment of the clinical competency since its inception in 1975. Largely considered as the assessment of choice, the OSCE format has been modified in a multitude of ways to suit the syllabus and needs of each institution; for example, candidates at the National Taiwan University School of Medicine are assessed separately on each domain, whereas candidates at the National University of Singapore undergo what are known as Clinical Skills and Clinical Reasoning stations; in Clinical Skills stations, students are expected to perform a physical examination and generate possible differentials from physical signs, while in Clinical Reasoning stations, candidates are to take a history and perform a relevant examination before a discussion with the examiner about their differentials, most likely diagnosis, investigations,
and management. Regardless of the format, the domains of history taking, physical examination, clinical discussion, and procedural skills often feature strongly and form the core of most, if not all, OSCEs. In keeping with the discussion above regarding the skillsets of the modern clinician, how then, will medical education, and specifically, OSCEs, evolve to bridge and meet the changing requirements of tomorrow?

History Taking
An essential and indispensable skill of a clinician is that of history taking, as echoed in the aphorism by Hampton, “A careful history will lead to the diagnosis 80% of the time.”\(^\text{22,23}\) While the number presented might be arbitrary, there is certain truth in that saying—for a detailed account of a patient’s presenting complaint and events leading to illness often gives critical clues and hints to the medical detective work that we are often tasked with doing.

In the age of AI, this skill is once again highlighted as something essential, for the use of AI interfaces to input presenting symptoms and complaints are restricted by several flaws, with 2 major ones that will be discussed here further. First, the AI system provides differentials and possibly a question list to ask from what is inputted into the system by the clinician. Systems like this would undoubtedly streamline processes and help doctors consider possibilities along that track, but this assumes that the original presenting complaint is interpreted correctly. The use of the algorithms of AI would potentially lock off the true diagnosis in the face of an error by a less astute clinician, which means that history taking is ever more important. Moreover, the algorithms of AI, while extremely powerful and robust, lose something in the process, that of the subtlety of patient complaints. While efforts can be made to attempt to differentiate the various symptoms from each other, a patient may not have thought something to be significant and reported to a front-facing interface of AI that it has a symptom. From the above, the need for good and effective history taking has been illustrated, and it is ever important to be able to pick up on the subtleties that patients may not have provided.

History taking assessment should, thus, be focused on evaluating these skills as set forth above. Perhaps, since AI systems (even at the lowest level) should be able to generate algorithms and lists of questions to lead questioning by the clinician, the focus might shift slightly from just clinching the various diagnosis and generation of differential diagnosis but place more emphasis on the differentiation of the various presenting symptoms and complaints, and to identify the subtleties that separate the entities from one another. These subtleties should also include when to decide when a patient might be malingering, or providing false symptoms, as well as when to consider a patient’s history as unclear testimony requiring further revision and clarification.

Moreover, assessment should also consider the situations and scenarios that future doctors might operate in, with AI systems on hand to assist in exams. For academic rigor, however, perhaps only the systems with the minimum capabilities will be provided, to ensure that enough competencies are achieved before allowing a student to progress on their journey into being a junior doctor. This also allows for assessors to gauge how well students know how to prioritize their lines of questioning in the limited time available; something that is already being tested now but will be even more important in the future given the immense amount of knowledge that AI is likely to have amassed. Thus, such a setup in assessment would allow for students to be proficient, independent of how AI might evolve – where it could take on roles to complement clinicians as described in the “minimum” above, or in a larger role set to fully support clinicians with more robust systems to overcome the various operational bottlenecks we now face.

Another aspect worth considering would be the impact of AI on the conduct of such assessments. With advancement of technologies allowing for front-facing and interactive interfaces, these deep-learning systems present an opportunity for increased objectivity, cost efficiency, and standardization. Since these systems have a vast data network, clinicians and educators can set real-life cases to test students, while also ensuring standardization, for the AI system is one central network and would help reduce interpersonal variation. The use of such systems, coupled with other newer technologies such as Virtual Reality (VR) simulations,\(^\text{24}\) allow for many more students to take the exam at the same time, and would help save time and resources, as a one-time investment would save costs in the long run, compared to compensating patients for their time per exam.

Physical Examination
In the domain of physical examination, while AI can aid in the integration of the history with the physical signs and the various possible differential diagnosis, we are of
the view that a clinician is still required for a physical examination and assessment. While some might argue that a pan-scan of every patient would give all the information required, current capabilities are not able to provide an affordable, quick imaging method with acceptable levels of side effects, with the pricey MRI lacking the speed, and CT-scans exposing patients to large amounts of radiation. Thus, in this aspect, clinicians are still required to have good physical examination skills and to pick up the relevant signs before there can be much aid from AI systems.

OSCEs should, thus, still emphasize the need to pick up important and relevant physical signs; as with history taking, perhaps OSCEs can now be taken together with the aid of a minimal-assistance AI system, one which can suggest a physical examination that should be done in the presence of the previous history, and with some integration abilities after the reporting of the physical signs found. This would also provide an opportunity to test a candidate’s ability to interpret the integrated data and the suggested differentials before they can choose what they believe is the most likely diagnosis. In terms of OSCE conduct, front-facing interactive interfaces can be implemented into models with certain physical signs to aid testing and save resources.

Clinical Discussions and Clinical Decision-Making

Clinical discussions and viva questioning have been an essential part of most OSCEs, as they allow the examiner a glimpse of the train of thought of the candidates, providing an opportunity to gauge the abilities of each student. Traditionally, these questions have focused on the interpretation of various pieces of history and physical signs, as well as differential diagnosis and relevant investigations. Able students would also progress on to discussions about management of these patients.

However, in the age of AI, perhaps there would be a shift in emphasis in line with the shift of the knowledge-intensive era. This shift is likely to be that of a much stronger emphasis on approaches rather than conditions, where a candidate’s algorithm and ability to discern a condition from another might be more valuable than knowledge about a few conditions, for that is something that AI can provide with much less effort. While this is true, it does not give students an excuse to skive off without knowing about conditions, for basic information regarding each condition is still required to generate a sound clinical algorithm. This is especially crucial since one can walk down the wrong road with one wrong clinical judgment with AI, and it is of utmost importance that a clinician avoids this scenario, and where they already have made a mistake, be able to identify it early and quickly as well as set the clinical path on the correct track before any damage is done. This perhaps could be assessed as well; the signs, symptoms, and parameters that might suggest that something is suspected, as well as how to investigate in these scenarios.

Procedural Skills

In the realm of procedural skills, the use of AI could greatly enhance the efficiency and success rate of these repetitive, simple procedures. A medical student graduating to become a junior doctor is expected to be well versed in performing basic procedures, as they will be called to assist when there is difficulty in said procedures in patients in the ward. While there are convincing arguments that AI should be able to reliably complete these more menial tasks with higher success rates than humans, one must always consider the possibility that not all institutions and clinics might be equipped with such systems and capabilities. Thus, it is still an essential skill to know these basic procedures, such that in the event of AI interface failure or lack of such facilities at various institutions or clinics, the clinician is still able to reliably complete these procedures to continue the various steps in patient care.

When applying this concept to OSCEs then, perhaps one of the more drastic changes to this would be the marking scheme. In an exam, candidates are often graded on various administrative measures, such as certain steps for identification of patients, and other safeguards to prevent performing an incorrect procedure on the wrong patient, such that it is possible to pass the station even if they are unable to successfully complete the procedure. While essential to ensure patient safety, patient preparation and aftercare can easily be handled by AI and should not be the emphasis of such exams; perhaps, these procedural skills stations should place a heavier weightage on the success and completion of the procedure itself, such that our clinicians would be proficient at completing said procedures upon graduation. Perhaps, further assistance as with the minimal level required can be provided, such as image guidance for blood draws or the like. This would, however, be dependent on the resources available in each
country and the “minimum standard” should be drawn from daily operational requirements of various hospitals.

OSCE conduct can also be greatly enhanced with AI, for it allows for a “standardized patient” to communicate with a student while completing said procedure required on a manikin or model. This might also help increase efficiency by reducing the number of examiners required for conducting exams in these stations. Moreover, AI can create simulated clinical environments for expansive learning, lessening environmental tension, and facilitate learners toward being fit to practice in the future. Pros and cons of AI applications in OSCE are summarized in Table 1. Musings from a future doctor (TKS) about the implementation of AI into healthcare proper are discussed in Supplement.

## Conclusion

Herein, we discussed the possibilities that AI might provide to OSCE examinations, which will reflect the changes in the skillsets that is required of the modern clinician. While AI still is in its infancy, we believe that it will play a pivotal role in the future, whether it be in healthcare or medical education and examinations, and while we do not claim to be prophets, we believe that this is the general direction toward which AI would lead medical education and hence its assessments. To stay relevant, we must continue to adapt and evolve as we navigate and overcome these uncharted territories that is our ever-evolving healthcare landscape; although we believe that there will always be a role for clinicians (albeit in different capacities), we must keep ourselves updated and be ready to accept and even influence change to stay as gatekeepers of these technologies. Given the theoretical nature of AI in OSCE (since the technology is still in its infancy), further study is required to further elucidate the role of AI in OSCE, and to the greater landscape of medical practice.

## Acknowledgment

The authors thank Prof. Jann-Yuan Wang for providing critical comments.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Alpaydin E. *Introduction to Machine Learning*. 4th ed. MIT press; 2020.
2. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380:1347–1358. doi:10.1056/NEJMr1814259
3. Sidey-Gibbons J, Sidey-Gibbons C. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol*. 2019;19:64. doi:10.1186/s12874-019-0681-4
4. Shinners L, Aggar C, Grace S, et al. Exploring healthcare professionals’ understanding and experiences of artificial intelligence technology use in the delivery of healthcare: an integrative review. *Health Informatics J*. 2020;26:1225–1236. doi:10.1177/1460458219874641
5. Buch VH, Ahmed I, Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities. *Br J Gen Pract*. 2018;68:143–144. doi:10.3399/bjgp18X695213
6. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. *J Med Internet Res*. 2020;22:e15154. doi:10.2196/15154
7. Harden RM, Stevenson M, Downie WW, et al. Assessment of clinical competence using objective structured examination. *Br Med J*. 1975;1:447–451. doi:10.1136/bmj.1.5955.447
8. Reid H, Gormley GI, Dornan T, et al. Harnessing insights from an activity system - OSCEs past and present expanding future assessments. *Med Teach*. 2020;1–6. doi:10.1080/0142159X.2020.1795100.
9. Zawacki-Richter O, Marín VI, Bond M, et al. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *Int J Educ Technol High Educ*. 2019;16:39.
10. Sánchez LE, Santos-Olmo A, Álvarez E, et al. Development of an expert system for the evaluation of students' curricula on the basis of competencies. *Future Internet*. 2016;8:22. doi:10.3390/fi8020022
11. Wartman S, Combs CD. Reimagining medical education in the age of AI. *AMA J Ethics*. 2019;21:E146–152. doi:10.1001/jamainethics.2019.146
12. Master K. Artificial intelligence in medical education. *Med Teach*. 2019;41:976–980. doi:10.1080/0142159X.2019.1595557
13. Paranjape K, Schinkel M, Nannan Panday R, et al. Introducing artificial intelligence training in medical education. *JMI Med Educ*. 2019;5:e16048. doi:10.2196/16048
14. Rampton V, Mittelman M, Goldhahn J. Implications of artificial intelligence for medical education. *Lancet Digital Health*. 2020;2:E111–112. doi:10.1016/S2589-7500(20)30023-6
15. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. *Acad Med*. 2018;93:1107–1109. doi:10.1097/ACM.0000000000002044
16. Loftus TJ, Upchurch GR Jr, Bihorac A. Use of artificial intelligence to represent emergent systems and augment surgical decision-making. *JAMA Surg*. 2019;154:791–792. doi:10.1001/jamasurg.2019.1510
17. Taylor RB. *Medical Wisdom and Doctoring: The Art of 21st Century Practice*. Springer Science+Business Media, LLC, 2010.
18. Brown S. Preserving the human touch in medicine in a digital age. *CMAJ*. 2019;191:E622–623. doi:10.1503/cmaj.109-5757
19. Alkureishi MA, Lee WW, Lyons M, et al. Impact of electronic medical record use on the patient-doctor relationship and communication: a systematic review. *J Gen Intern Med*. 2016;31:548–560. doi:10.1007/s11606-015-3582-1
20. Harden RM, Gleson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ*. 1979;13:39–54. doi:10.1111/j.1365-2923.1979.tb00918.x
21. Harden RM. Revisiting ‘Assessment of clinical competence using an objective structured clinical examination (OSCE)’. *Med Educ*. 2016;50:376–379. doi:10.1111/medu.12801
22. Hampton JR, Harrison MJ, Mitchell JR, et al. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. *BMJ*. 1975;2:486–489. doi:10.1136/bmj.2.5969.486
23. Cooke G. A is for aphorism - is it true that ‘a careful history will lead to the diagnosis 80% of the time’? *Aust Fam Physician*. 2012;41:534.
24. Pottle J. Virtual reality and the transformation of medical education. *Future Health J*. 2019;6:181–185. doi:10.17861/fhj.2019-0036
25. Hungerford C, Walter G, Cleary M. Clinical case reports and the viva voce: a valuable assessment tool, but not without anxiety. *Clin Case Rep*. 2015;3:1–2. doi:10.1002/ccr3.225
26. Hashimoto DA, Rosman G, Rus D, et al. Artificial intelligence in surgery: promises and perils. *Ann Surg*. 2018;268:70–76. doi:10.1097/SLA.0000000000002693