

## Embracing the next frontier in assessment

### Adopter la prochaine frontière de l'évaluation

Christina St-Onge<sup>1</sup>

<sup>1</sup>Université de Sherbrooke, Quebec, Canada

Correspondence to: Christina St-Onge, PhD, email: stoc3101@usherbrooke.ca

Published: May 1, 2025. CMEJ 2025 16(2) Available at <https://doi.org/10.36834/cmej.81477>

© 2025 St Onge; licensee Synergies Partners. This is an Open Journal Systems article distributed under the terms of the Creative Commons Attribution License. (<https://creativecommons.org/licenses/by-nc-nd/4.0>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited.

A frequent saying in our field is that “assessment is the engine driving medical education,”<sup>1</sup> and perhaps no moment in recent history has offered greater potential for reimagining that engine than now. As the Canadian Medical Education Journal (CMEJ) marks its 15th anniversary, we find ourselves at a crossroads: technological advances—particularly in artificial intelligence (AI)—are poised to transform how we assess and support learners. Will we harness these innovations to foster wellness and inclusivity, or will we remain anchored to our traditions?

## Why assessment matters more than ever

For decades, assessment has shaped not only what students learn but also how they learn.<sup>1–5</sup> Traditional tools such as multiple-choice questions and the Objective Structured Clinical Examination (OSCE) have served as benchmarks for clinical knowledge and skills.<sup>6,7</sup> However, the landscape of medical education has expanded dramatically. Issues such as learner wellness, equity, diversity, inclusion (EDI), and professional identity formation now demand more holistic approaches to understanding and guiding learner development and assessment.<sup>8,9</sup>

In recent years, the community has engaged in robust discussions on competency-based medical education (CBME), formative feedback, and collaborative assessments. These conversations echoed a growing consensus: assessment must evolve from a high-stakes, summative gatekeeper to a process that encourages

reflection, growth, and psychological safety.<sup>8,10</sup> Assessment can be a powerful tool for growth, belonging, and fairness when it is used to foster honest reflection on both successes and shortcomings—without fear of punitive consequences.

## A changing landscape: AI and assessment

The interface between AI and assessment is both exciting and complex. Recent literature highlights AI's potential to support educators by streamlining assessment development, such as through automated item generation and quality assurance.<sup>11,12</sup> This could alleviate the significant workload associated with assessment development,<sup>13,14</sup> allowing educators to focus on other tasks such as ensuring the quality of AI-generated assessment.

AI also holds promise for advancing EDI. For example, AI-driven tools could help identify and remove biased language or content in exam items, promoting a more equitable testing environment. Similarly, AI could flag potential biases in clinical scenarios, prompting educators to develop more inclusive materials. However, these tools must be deployed judiciously, with human oversight, to ensure contextual appropriateness and fairness. Moreover, language remains an issue as most LLMs are trained in English.<sup>15,16</sup> Having limited data to train in other languages might make those productions of lesser quality than the LLM trained on English data.

AI-driven formative assessments can provide low-stakes, personalized feedback for learners, supporting self-

regulated learning and lifelong learning mindsets.<sup>17–19</sup> Using AI to self-test one's knowledge, for example, can normalize trial, error, and reflection cycles. Repeated testing can contribute to reducing anxiety often associated with traditional assessments<sup>20</sup> while reinforcing the notion that learning is a shared journey. A required safeguard, of course, would be putting in place strategies to ensure learners have the oversight to appraise AI-generated content critically.

## Navigating risks and ethical considerations

With rapid technological advances come legitimate concerns: inadvertent bias, over-reliance on algorithms, and cheating.<sup>21</sup> These challenges demand robust, evidence-based approaches and ongoing ethical scrutiny. They might require that we revisit our definition of validity and our approaches to validation to ensure that we can attest that the assessment results appropriately reflect the knowledge, skills, attitude, or competence we aimed to document. Faculty members are also faced with ethical dilemmas when contemplating the use of AI. Is it appropriate to use it to scrub exam questions and stations to make them less prone to bias? Is it also acceptable to use AI to generate feedback? The different uses require careful consideration.<sup>22</sup> As stated by Lambert Schurwirth in his address at the 2024 Royal College Summit on AI and put forth by other researchers in the field,<sup>22</sup> it is not whether we adopt AI-driven assessment but how to do so ethically, equitably and in alignment with the holistic values of medical education.

## Charting a path forward

This moment calls for nuanced dialogue, rigorous research, and multi-stakeholder collaboration—an ethos CMEJ has championed since its founding. CMEJ stands ready to publish, discuss, and refine the ideas that will shape the next era of medical education. We invite our community to engage in this critical conversation and to help build an assessment system that is not only innovative but also just, compassionate, and evidence-informed.

In this issue of the CMEJ, we have many articles addressing multiple problems that fit these criteria. We trust you will find them informative and perhaps even transformative as we look forward to many others in the future, some touching on crucial issues of assessment.

## Original Research

[Exploring the perspectives of new-in-practice specialists about the Health Advocate role: “I didn’t even know where to start”](#) by Cochrane and team explored early-career physicians' perspectives on health advocacy training and its relevance in practice. They found that while the training prepared them for patient-level advocacy, the new physicians felt underprepared for system-level advocacy.

[Considering the potential unintended consequences of RateMDs: an exploratory study in one specialty](#) by Pulkki and co-authors explored how negative ratings on a physician-rating website might inform lifelong learning for physicians. They found that the criticisms often reflected deficiencies in areas corresponding to key CanMEDS roles such as professionalism, communication, and leadership. They noted that while this feedback provided opportunities for professional improvement, it could also harm physician confidence.

In Bartman et al.'s [Multi-source feedback in undergraduate medical education: a pilot study](#), the authors adapted the MCC 360 multisource feedback tool for undergraduate medical education and piloted it with clerkship students. The team found that while some students found the process burdensome, they valued the patient feedback and facilitator support.

## Brief Reports

In their report, [Are we optimizing medical students' preparation for clerkship? A content analysis of narrative comments on clinical skills during preclinical training](#), Bergeron and team looked at whether narrative comments on clinical skills help with the transition from preclinical medical training to clerkship. They found that narrative comments on clinical skills before clerkship were mostly positive and lacked suggestions, thus limiting their usefulness.

In [Diagnostic hypothesis generation through engaged peer observation: a quantitative descriptive study in a clinical simulation context](#) by Stéphanie Benoît et al., the authors found that student observers could generate diagnostic ideas similar to clinical students during clinical simulations, thus demonstrating peer observation as a helpful tool for learning clinical reasoning. This is a French article.

[Can relaxation exercises improve students' OSCE grades: a prospective study](#) by Massalou et al. studied whether human performance optimization techniques would improve medical students' performance during OSCEs. While the techniques, such as relaxation methods, were

successful at reducing anxiety, they did not improve their scores.

### Reviews, Theoretical Papers, and Meta-Analyses

In [Optimizing feedback reception: a scoping review of skills and strategies for medical learners](#) by Rowe et al., the authors reviewed studies on skills and strategies to help medical learners receive, evaluate, and use feedback. All identified studies used educational workshops, reporting learners' perceived improvement in feedback reception; however, none evaluated the strategies in practice.

### Black Ice

[To be \(virtual\) or not to be: six ways to get a grip on choosing a delivery method for your educational program](#) by Lahouaoula and team provided tips – such as considering the desired interactivity –to help medical educators choose between the various online and in-person teaching methods. Their tips can help teachers select the right format for their educational program.

### Canadiana

In the article, [Beyond the classroom: lessons in empathy and accessibility as a student clinician serving Calgary's vulnerable populations](#) by Brandon Azer, Azer shared his personal experience volunteering at a student-run wellness centre for individuals facing homelessness and poverty. He maintained that this type of experience is important for future doctors as it increases empathy and improves understanding of patient-centered care.

[Bridging gaps in orthopedic residency admissions: embracing diversity beyond research metrics](#) by Kumar and team advocated for a more inclusive approach in the Canadian residency selection process. They contended that by prioritizing things such community engagement and health advocacy in residency selection process, it will create a more effective medical field.

### You Should Try This!

[Rethinking global health training: making the links between theory and practice](#) by Fisher and co-authors describes their revised global health program, which aims to address ethical concerns in international medical electives and train equity-oriented healthcare workers.

[Teaching compassion through a community-led, experiential learning activity for undergraduate medical students: the Empathy Project](#) by Shoppoff and team described an immersive experience that simulated the challenges of homelessness for first-year medical students. The feedback they received suggested that the initiative

increased both empathy and compassion, which will help medical students provide better care for their patients.

### Commentary and Opinions

Sinha and team's commentary, [Breaking bad news, building better learners: using the SPIKES framework for medical education feedback](#),<sup>24</sup> proposed adapting the SPIKES framework, used for delivering bad news to patients, to provide quality feedback in medical education. The authors contended that by following the six steps of SPIKES—setting, perception, invitation, knowledge, emotions, and strategy/summary—educators can provide both positive and constructive feedback.

[A critical analysis of The Royal College of Physicians and Surgeons of Canada examination experience](#) by Malcolm M MacFarlane studied the role of high-stakes examinations in the Royal College of Physicians and Surgeons of Canada certification process. The paper reviews validity, reliability, and fairness best practices in exam construction, and calls for greater scrutiny of the Royal College of Physicians and Surgeons of Canada examination processes.

[The hidden curriculum](#) by Mohammed Kasim Ali commented on the “hidden curriculum” in medical education, which normalizes exhaustion and self-sacrifice as signs of dedication. Ali argued for cultural change in medicine that includes advocating for well-being over silent endurance.

### Letters to the Editor

The letter [Disclosure of bad news: a challenging practice?](#) by Modesto dos Santos and Modesto Sugai responded to Preti and Sanatani's [Five ways to get a grip on the personal emotional cost of breaking bad news](#). The authors commended Preti and Sanatani's focus on enhancing the training of healthcare workers in breaking bad news, and offered additional studies to advance the discussion.

The letter, [Royal College exams, examined](#) by Bandiera et al. responded to MacFarlane's commentary, [A critical analysis of the Royal College of Physicians and Surgeons of Canada examination experience](#). While agreeing on the importance of transparency and fairness in the certification exams, the authors caution against releasing detailed exam data, as it could be misleading or risk candidate privacy.

### Images

Azer and team's article and image, [A glimpse beyond](#), is digital artwork that highlights the need for empathy in medicine—reminding learners that behind every chart is a human being with a complex life story.

[Constellations and the sound of hooves](#) by James Huntley is an ekphrastic 55-word commentary and acrylic art piece that discusses how pattern recognition drives medical advancements but also contributes to biases when patterns are misinterpreted. This artwork is the cover image for this issue.

## Scientific Reports

[Objective Structured Clinical Examinations practices across Canadian medical schools: a national overview](#) by Gilchrist and Bismilla examined the frequency, type, and timing of Objective Structured Clinical Examinations (OSCEs) in Canadian medical schools. They reported that pre-clerkship assessments had a balanced mix of formative and summative OSCEs, while clerkship OSCEs were predominantly summative.

Enjoy!

**Edited by:** Marcel D'Eon (editor-in-chief)

## References

1. Van Der Vleuten CP. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ*. 1996;1(1):41-67. <https://doi.org/10.1007/BF00596229>
2. Brown GTL. The past, present and future of educational assessment: a transdisciplinary perspective. *Front Educ*. 2022;7. <https://doi.org/10.3389/educ.2022.1060633>
3. Hattie J, Timperley H. The power of feedback. *Rev Educ Res*. 2007;77(1):81-112. <https://doi.org/10.3102/003465430298487>
4. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9):S63-7.
5. Ben-David MF. The role of assessment in expanding professional horizons. *Med Teach*. 2000;22(5):472-7. <https://doi.org/10.1080/01421590050110731>
6. Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356(4):387-96. <https://doi.org/10.1056/NEJMr054784>
7. Norcini J, Anderson MB, Bollela V, et al. 2018 Consensus framework for good assessment. *Med Teach*. 2018;40(11):1102-9. <https://doi.org/10.1080/0142159X.2018.1500016>
8. Boud D, Molloy E. Rethinking models of feedback for learning: The challenge of design. *Assess Eval High Educ*. 2013;38(6):698-712. <https://doi.org/10.1080/02602938.2012.691462>
9. Cruess RL, Cruess SR, Steinert Y. Amending Miller's pyramid to include professional identity formation. *Acad Med*. 2016;91(2):180-5. <https://doi.org/10.1097/ACM.0000000000000913>
10. Edmondson AC. The fearless organization: Creating psychological safety in the workplace for learning, innovation, and growth. Hoboken (NJ): John Wiley & Sons; 2018.
11. Gierl MJ, Haladyna TM. Automatic item generation: Theory and practice. New York: Routledge; 2013.
12. Lee J, Wu AS, Li D, Kulasegaram KM. Artificial intelligence in undergraduate medical education: a scoping review. *Acad Med*. 2021;96(11S):S62-70. <https://doi.org/10.1097/ACM.00000000000004291>
13. García-Ramírez Y, Bijelić V. Assessing task prioritization for professors through affinity and satisfaction scores. *Cogent Educ*. 2024;11(1):2321355. <https://doi.org/10.1080/2331186X.2024.2321355>
14. Link AN, Swann CA, Bozeman B. A time allocation study of university faculty. *Econ Educ Rev*. 2008;27(4):363-74. <https://doi.org/10.1016/j.econedurev.2007.04.002>
15. Naveed H, Khan AU, Qiu S, et al. A comprehensive overview of large language models. *ArXiv Prepr ArXiv230706435*. 2023;
16. Liu Y, Cao J, Liu C, Ding K, Jin L. Datasets for large language models: a comprehensive survey. *ArXiv Prepr ArXiv240218041*. 2024. <https://doi.org/10.48550/arXiv.2402.18041>
17. Sardi J, Candra O, Yuliana DF, Yanto DTP, Eliza F. How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review. *Int J Eng Pedagogy*. 2025;15(1):94-108. <https://doi.org/10.3991/ijep.v15i1.53379>
18. Panadero E. A review of self-regulated learning: six models and four directions for research. *Front Psychol*. 2017;8:422. <https://doi.org/10.3389/fpsyg.2017.00422>
19. Roll I, Winne PH. Understanding, evaluating, and supporting self-regulated learning using learning analytics. *J Learn Anal*. 2015;2(1):7-12. <https://doi.org/10.18608/jla.2015.21.2>
20. Yang C, Li J, Zhao W, Luo L, Shanks DR. Do practice tests (quizzes) reduce or provoke test anxiety? A meta-analytic review. *Educ Psychol Rev*. 2023;35(3):87. <https://doi.org/10.1007/s10648-023-09801-w>
21. Dawson P, Bearman M, Dollinger M, Boud D. Validity matters more than cheating. *Assess Eval High Educ*. 2024;49(7):1005-16. <https://doi.org/10.1080/02602938.2024.2386662>
22. Corbin T, Dawson P, Nicola-Richmond K, Partridge H. 'Where's the line? It's an absurd line': Towards a framework for acceptable uses of AI in assessment. *Assess Eval High Educ*. 2025;1-13. <https://doi.org/10.1080/02602938.2025.2456207>
23. Cochrane JD, Dudek N, Crawford K, Cowley L, LaDonna KA. Exploring the perspectives of new-in-practice specialists about the Health Advocate role: "I didn't even know where to start" *Can Med Educ J*. 2025;16(2) <https://doi.org/10.36834/cmej.78570>
24. Pulkki KH, Pira S, Young M, et al. Considering the potential unintended consequences of RateMDs: an exploratory study in one specialty. *Can Med Educ J*. 2025;16(2) <https://doi.org/10.36834/cmej.77821>
25. Bartman I, St-Onge C, Roy M, et al. Multi-source feedback in undergraduate medical education: a pilot study. *Can Med Educ J*. 2025;16(2) <https://doi.org/10.36834/cmej.79283>
26. Bergeron L, Blanchette P, Chakroun M, et al. Are we optimizing medical students' preparation for clerkship? A content analysis of narrative comments on clinical skills during preclinical training. *Can Med Educ J*. 2025;16(2) <https://doi.org/10.36834/cmej.78569>
27. Benoit S, Bouchard-Lamothe D, Denis-LeBlanc M, Burnier I. Diagnostic hypothesis generation through engaged peer observation: a quantitative descriptive study in a clinical

- simulation context. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.77080>
28. Massalou D, Doyen J, Almairac F, et al. Can relaxation exercises improve students' OSCE grades: a prospective study. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.77739>
  29. Rowe J, Bouchard Lamothe D, Haggerty T, et al. Optimizing feedback reception: a scoping review of skills and strategies for medical learners. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.79722>
  30. Lahouaoula I, Buckley H, Nathoo N. To be (virtual) or not to be: six ways to get a grip on choosing a delivery method for your educational program. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.78281>
  31. Azer B. Beyond the classroom: lessons in empathy and accessibility as a student clinician serving Calgary's vulnerable populations. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.80338>
  32. Kumar R, Gosain A, Saintyl JJ, Zheng A, Chima K, Cassagnol R. Bridging gaps in orthopedic residency admissions: embracing diversity beyond research metrics. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.80775>
  33. Fisher LA, Aristizabal Londono C, Cropper K, Beresh G, Alhassan JA. Rethinking global health training: making the links between theory and practice. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.79921>
  34. Shoppoff L, Mathew C, Burkholder Harris K, et al. Teaching compassion through a community-led, experiential learning activity for undergraduate medical students: the Empathy Project. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.79452>
  35. Sinha S, Gilchrist C, Leung F-H. Breaking bad news, building better learners: using the SPIKES framework for medical education feedback. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.80847>
  36. MacFarlane MM. A critical analysis of The Royal College of Physicians and Surgeons of Canada examination experience. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.79678>
  37. Ali MK. The hidden curriculum. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.81103>
  38. Modesto dos Santos V, Modesto Sugai K. Disclosure of bad news: a challenging practice? *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.80918>
  39. Preti BT, Sanatani MS. Five ways to get a grip on the personal emotional cost of breaking bad news. *Can Med Educ J.* 2024;15(3):97-99. <https://doi.org/10.36834/cmej.78228>.
  40. Bandiera G, Watling C, Horsley T, Dagnone D, Taber S, Bhanji F. Royal College exams, examined. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.81283>
  41. Azer B, Dyson K, Ivaturi S. A glimpse beyond. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.80997>
  42. Huntley JS. Constellations and the sound of hooves. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.81116>
  43. Gilchrist C, Bismilla Z. Objective Structured Clinical Examinations practices across Canadian medical schools: a national overview. *Can Med Educ J.* 2025;16(2) <https://doi.org/10.36834/cmej.80144>