



University of
Nottingham

UK | CHINA | MALAYSIA

***If AI can pass it, should we
still assess it?***

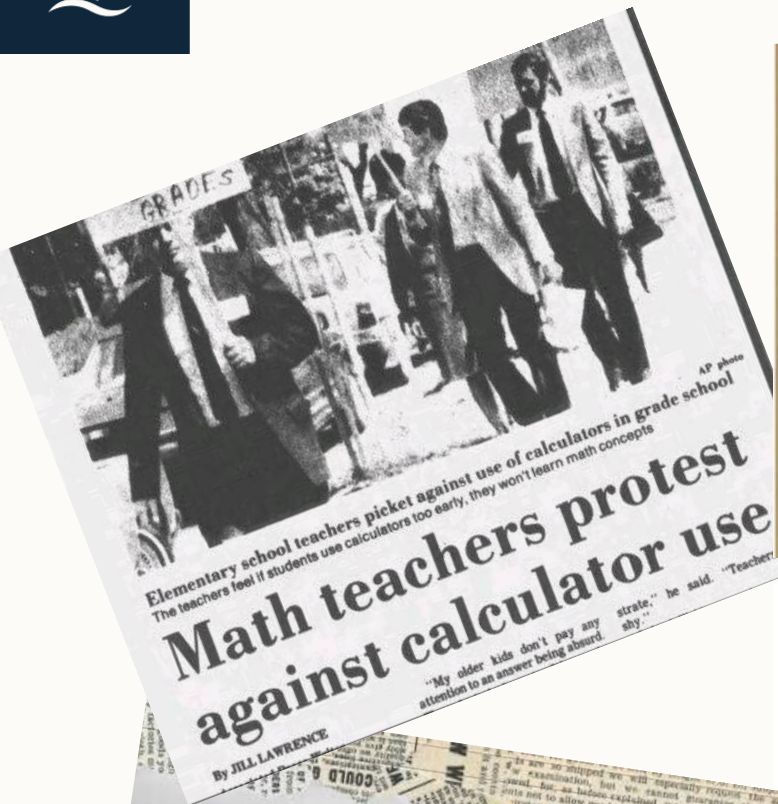
***Rethinking assessment in
Medical Education***

Professor Claire Stewart





Times are changing





Day-to-day use of AI?



How many of you use AI



How do you use it?



If you are refining the outputs - how can you do this if you haven't been taught, learned and been assessed on



What future are we training for?

If ai is the future – why are we resisting?





Assessment



Currently, assessment constructs are developed for ease, feasibility and reliability.



Digital adaptations are simply digitalizing the exam



Prepare for the future clinicians





Assessment pathway

AI can do all of these

Blueprinting

Content creation

Marking

Quality Assurance (psychometrics)

Student and Faculty Feedback



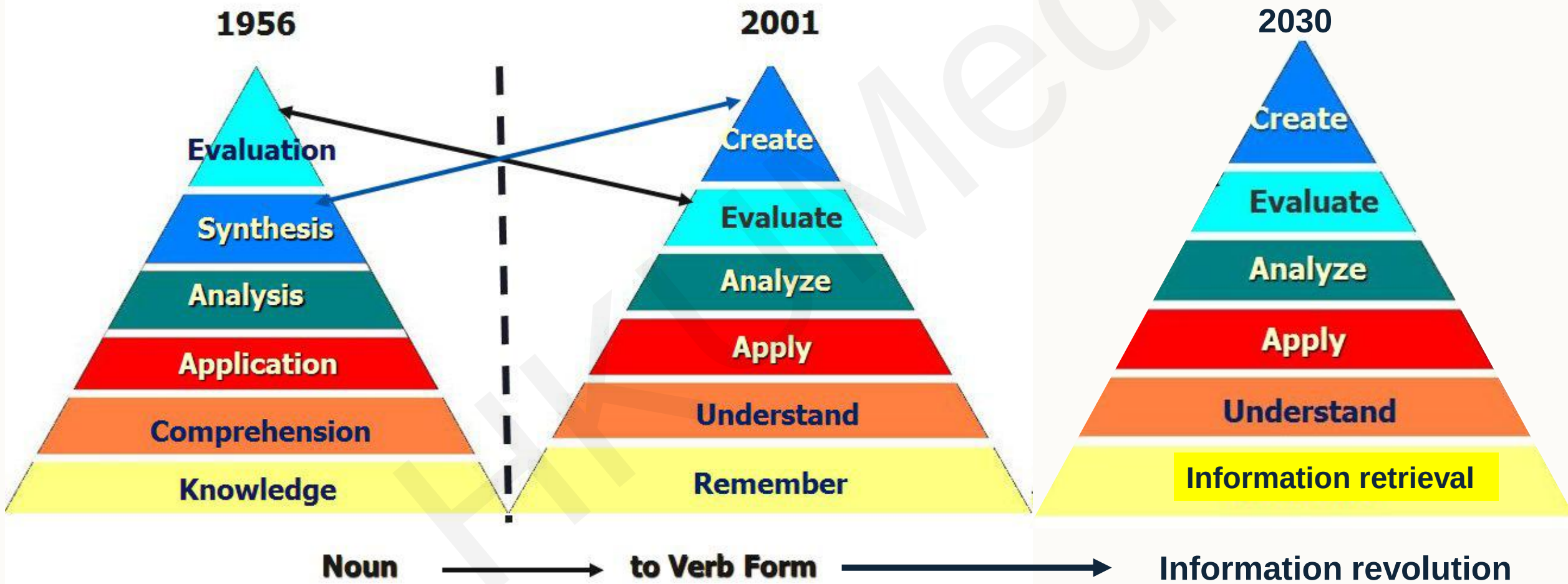
Why, then, are we resisting using AI to reduce our workload and make it more manageable, freeing up time to facilitate a true flipped classroom and discuss feedback meaningfully with our students?





What should we assess in an AI-enabled world?

Updated Blooms?





AI to create content

write me a MLA style SBA with 5 options for a first year medical student

Here is a single best answer (SBA) question formatted according to the General Medical Council's Medical Licensing Assessment (MLA) style, tailored for a first-year medical student studying core anatomy and clinical assessment. [DOCX](#)

Question

A 24-year-old man presents to the emergency department after falling onto his outstretched right hand while playing football. He reports severe pain on the radial side of his right wrist. On physical examination, there is marked tenderness to palpation within the anatomical snuffbox. Plain radiographs of the wrist taken on arrival do not show a clear fracture line.

Which of the following bones is most likely injured in this patient?

- ☐ A. Capitate
- ☐ B. Hamate
- ☐ C. Lunate
- ☐ D. Scaphoid
- ☐ E. Triquetrum

+ Ask Gemini



Flash



Correct Answer

D. Scaphoid

Explanation

- ☐ **D. Scaphoid:** A fall onto an outstretched hand (FOOSH) is the classic mechanism of injury for a scaphoid fracture. The anatomical snuffbox is bounded by the tendons of the extensor pollicis longus, extensor pollicis brevis, and abductor pollicis longus; the floor of this space is formed by the scaphoid bone. Tenderness in the anatomical snuffbox is highly sensitive for a scaphoid fracture. Early radiographs can frequently appear normal in non-displaced fractures, requiring clinical immobilization and repeat imaging in 10–14 days or advanced imaging (e.g., MRI) to prevent complications such as avascular necrosis due to its retrograde blood supply.
- ☐ **A. Capitate:** The capitate is the largest carpal bone and occupies the center of the wrist. It is less frequently fractured during a FOOSH injury and does not form the floor of the anatomical snuffbox.
- ☐ **B. Hamate:** Fractures of the hamate (specifically the hook of the hamate) typically occur from direct impact during sports involving swinging bats, clubs, or racquets, rather than a FOOSH mechanism. Palpation of the anatomical snuffbox does not elicit tenderness in isolated hamate injuries.
- ☐ **C. Lunate:** A FOOSH injury can cause a lunate dislocation or fracture, but lunate pathology typically presents with central wrist pain, reduced range of motion, or median nerve compression (carpal tunnel symptoms) rather than focal tenderness in the anatomical snuffbox.
- ☐ **E. Triquetrum:** The triquetrum is located on the ulnar side of the wrist (medial side in the anatomical position) and is the second most common carpal bone fractured. However, it presents with tenderness on the ulnar aspect of the wrist, not within the anatomical snuffbox.



AI-proof assessments

Core Principles of AI-Resistant Design

- **Process over Product:** Evaluating *how* a student arrived at an answer (drafts, reasoning, iteration) rather than just the final submitted text.
- **Personalisation and Context:** Tying assignments directly to hyper-local contexts, classroom discussions, or personal experiences.
- **Multi-Modal Communication:** Shifting away from standard take-home text essays toward oral, visual, or interactive formats

Common Examples of AI-Proof Assessments

- **Oral Defences & Viva Voces:** Students present their work and answer real-time, unscripted follow-up questions from peers or instructors to demonstrate genuine mastery.
- **Process Tracking & Version Histories:** Requiring work to be completed in cloud documents (e.g., Google Docs or Word Online) to review version history and verify natural writing development instead of large copy-pasted blocks.
- **Video & Audio Reflections:** Asking students to record short, informal video reflections explaining their thought process, key takeaways, or real-world applications.
- **Social Annotation & Collaborative Analysis:** Using tools like Hypothesis where students annotate, discuss, and debate a shared reading text collaboratively online.
- **In-Class Teach-Backs:** Assigning students to teach a core concept to their peers or role-play an interview with an industry professional.
- **Incomplete Data & Critical Evaluation Tasks:** Presenting students with intentionally flawed or incomplete information and grading them on their ability to identify missing variables and apply intellectual restraint.



AI-proof assessments – in practice

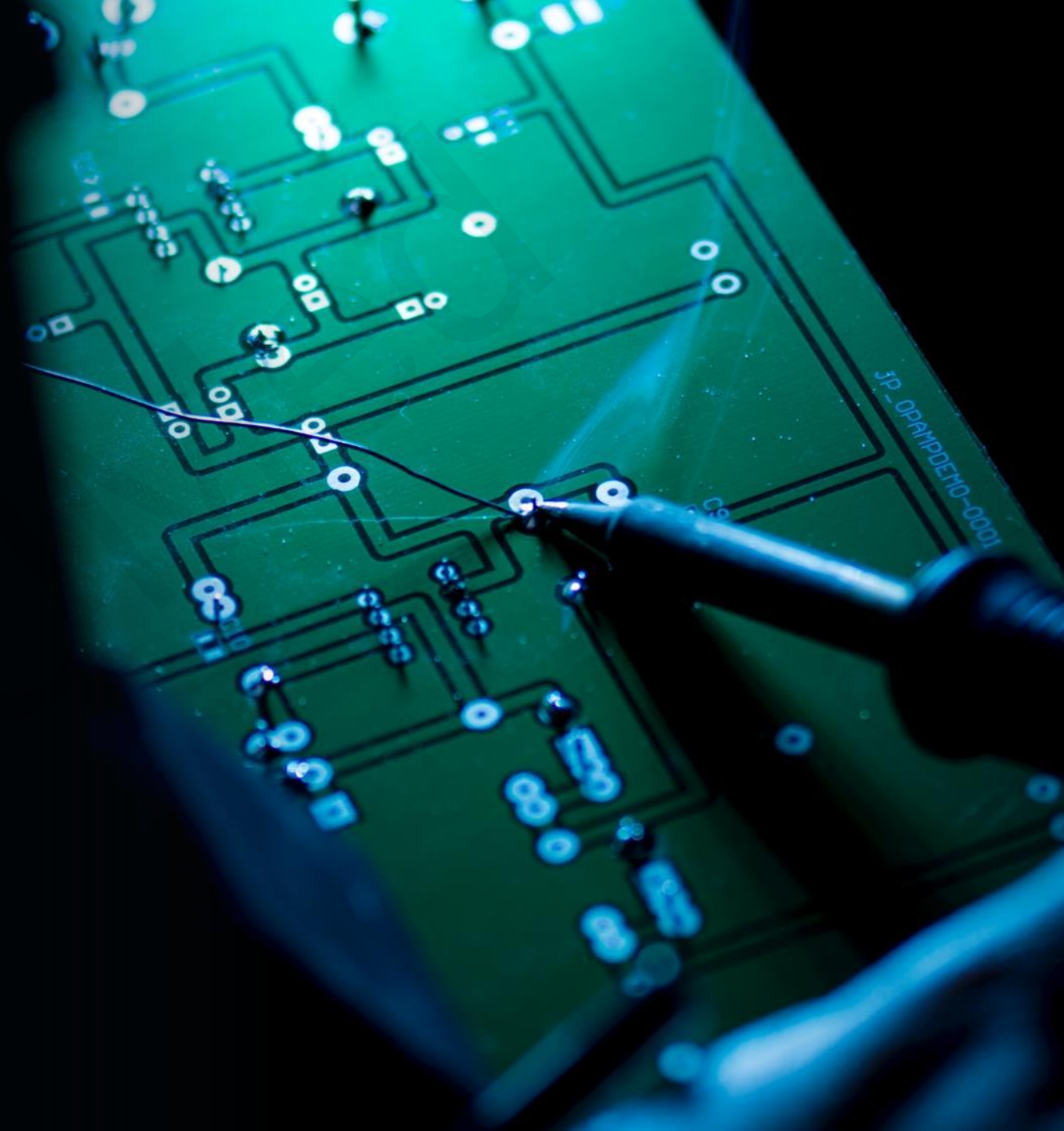
- **Viva – No**
 - Examiner Bias & Low Inter-Rater Reliability
 - High Anxiety Disproportionately Affects Performance
 - Resource Intensity & Scalability Issues.
 - Inconsistent Questioning & Scope Drift.
 - Lack of Transparency & Audit Trails
 - Susceptibility to "Hawthorne/Charisma" Effects
- **Examinations - yes!**
- **In-person activities** (team-based activities, peer review, presentations, timed in person work etc.) - **yes!**
- **Workplace-based assessments- yes!**



University of
Nottingham

UK | CHINA | MALAYSIA

**But is this still just
adapting to the AI
world in the same
way as we simply
digitalised
assessments?**





AI enhanced assessments

Rather than banning it, should we be assessing how they can use AI effectively

- Use AI to personalise assessments
 - Examinations with adaptive testing, thoroughly sampling to realise the limits of students' knowledge and competencies?
- Non-exam-based assessments
 - Permit the use of referenced AI
 - In the same way we permit internet searching or use of research databases
 - Ensure that the topics have elements of higher-order thinking requiring comprehension, analysis and synthesis



AI enhanced feedback

Continue to use digital & AI
algorithms to spot trends:

Individual

Cohort

Institutional

Facilitate the analysis of
the feedback

Synthesize development
plans



So we are in the midst of change – What should we do now?

Remember, the students will always be ahead of faculty

Embrace it - not fear it

Adapt

Remain agile

Research it.... Quickly, contextually and collaboratively



Remember

Everything that is in the world when you are born is normal

Anything that is developed between then and when you turn 30 is exciting and creative and with any luck you can make a career out of it.

Anything invented after you turn 30 is against the natural order of things and the beginning of the end of civilisation as we know it until its been around for 10 years when it gradually becomes alright really.

Douglas Adams



University of
Nottingham

UK | CHINA | MALAYSIA

Thank you

Any questions?



References

- **Alizadeh, M., & Jafar Sameri, M.** (2025). Intelligent Assessment Systems in Medical Education: A Systematic Review. *Journal of Advances in Medical Education & Professionalism*, 13(3), 133–144.
- **Astbury, R., et al.** (2026). The Impact of Artificial Intelligence on Medical Education: A Systematic Review. *Applied Science and Biotechnology Journal for Advanced Research*, 5(3), 37–41.
- **Elhassadi, A., et al.** (2026). Artificial Intelligence in Healthcare Education: A Systematic Review of Trends, Pedagogical Applications, and Outcomes. *PRISMA Systematic Review*.
- **General Medical Council.** (2024). *Achieving Good Medical Practice: Guidance for Medical Students*. General Medical Council
- **Masters, K.** (2024). Generative AI in medical education assessments: Threat, tool, or catalyst for systemic reform? *Medical Teacher*, 46(4), 452–461.
- **Nouri, H.** (2025). Performance of large language models in medical licensing examinations: A systematic review and meta-analysis. *Journal of Educational Evaluation for Health Professions*, 22, 36.
- **Preiksaitis, C., et al.** (2025). Computerized adaptive testing and AI feedback loops in health professions education: A systematic review. *Medical Education*, 59(2), 184–196.
- **Quality Assurance Agency for Higher Education.** (2023). *Reconsidering Assessment for the ChatGPT Era: QAA Advice on Developing Sustainable Assessment Strategies*. QAA
- **Roveta, A., et al.** (2025). Artificial Intelligence in Medical Education: A Narrative Review on Implementation, Evaluation, and Methodological Challenges. *Medical Sciences*, 6(9), 227.
- **Russell Group.** (2023). *Russell Group Principles on the Use of Generative AI Tools in Education*. Russell Group Universities
- **Xie, Y., et al.** (2026). Generative artificial intelligence in medical education: From knowledge assessment to clinical reasoning and professional competence. *Frontiers in Medicine*, 13, 1855759
- **Zhao, J., Chapman, E., & Sabet, P. G. P.** (2026). Generative AI and Educational Assessments: A Systematic Review. *Educational Psychology Review*, 38(2), 112–134.