# Day 1: Friday December 6th, 2024 – Workshop Meeting

***12:45 – 18:00\* (\*All times are local)***

***Chairs****:* ***Maeve Durkan, Ambrogio Fassina, Danny Mathysen, Albert Mifsud, Gian Battista Parigi***

**12.45 – 13.00**

Name badges and registration

**13.00**

Welcome and introduction: Maeve Durkan

Dr. Maeve Durkan opened the meeting with a warm welcome and invited all participants to introduce themselves. Each attendee briefly presented their name, role, and affiliation to begin the session.

* 1. **- 13.20**

Appraisals update – collaborations

The UEMS-CESMA appraisal system was initiated by Prof. Danny Mathysen in 2013. Since then, 13 appraisals have been conducted. The purpose of the appraisal process is to ensure that European medical specialty exams are valid, safe, and uphold high standards of medical practice. A key element of the appraisal process is collaboration, often involving volunteers. While collaboration brings value, it also brings challenges. Discussions emphasized the importance of having 2–3 appraisers per appraisal and the need for all to sign confidentiality agreements, as sensitive information is often shared. In 2024, a 22-page appraisal report template was finalized and adapted for future use in 2025, and terminology in the appraisal process is under review—for example, possibly replacing "safety procedure" with "security procedure." During the Spring CESMA meeting in Antwerp, high expectations were set, particularly regarding the role of volunteers. There is limited availability of qualified appraisers, as appraisers must work in pairs. The fee for appraisals has increased significantly—from €500 to €3,000—to reflect the growing workload and ensure sustainability:

* + The fee covers multiple aspects: the exam, report, and diploma.
  + The price is the same regardless of whether the exam is in one part or several parts.
  + CESMA ensures at least two CESMA appraisers and one academic partner are involved in each appraisal.

Dr. Celine Carr raised concerns about the fee increase and asked if any customer satisfaction surveys had been conducted but Dr. Maeve Durkan explained that the fee increase had been discussed over several meetings and is necessary to maintain the quality and sustainability of CESMA appraisals. Dr. Durkan emphasized the importance of maintaining formative assessments—evaluations meant to support learning rather than simply pass/fail outcomes. Appraisal reports are crucial to protect the integrity of CESMA processes and ensure exams do not lose credibility.

**UEMS-CESMA Appraisals – Updates and Standards Discussion**

Appraisal Status Update (Prof. Danny Mathysen)

* Provided an update on the status of UEMS-CESMA appraisals, including:
  + Appraisals that have been completed
  + Those still outstanding
  + Appraisals that are expected or upcoming

Standards Framework (Albert Mifsud)

* Explained the definition of a “standard” as a document that establishes an agreed-upon method of operation.
* Recalled that in 2021, two key documents were adopted as CESMA’s standard references (Document 1 and Document 2).
* The main goals of these standards include:
  + Increasing flexibility
  + Making documents more accessible
  + Reconsidering criteria for:
    - Admission to UEMS examinations
    - Awarding of UEMS Diplomas and Fellowships

Albert Mifsud posed a set of strategic and philosophical questions for future consideration:

* Should CESMA only approve exams that are exclusively available to medical doctors?
* Must European Training Requirements (ETRs) be UEMS-approved in order to assess exam quality?
* If ETRs are not UEMS-approved, how should quality be assessed?
* Should UEMS exams be open to candidates from non-UEMS member states?

**13.20-13.30**

* Treasurer’s Report
* Certifications and Appraisals
* Final certificates issued for:
* Legal & Liability Aspects (Session with Me Adrien Hanoteau)
* Legal insights on UEMS exam liability and contract structures.
* Artificial Intelligence in Medical Exams
* AMME – GDPR & AI Ethics

**13.30 -13.40**

Rules and regulations - responsibility of appraiser is to CESMA alone Confidentiality

Appraisal update - done/ outstanding/challenges

Responsibility of Appraiser:  
The appraiser is accountable solely to CESMA. All appraisal activities must align with CESMA's guidelines, and the appraiser must maintain objectivity, transparency, and fairness throughout the evaluation process.  
**All appraisal data, documentation, and discussions are strictly confidential.** Information must not be disclosed to any unauthorized party under any circumstances. A breach of confidentiality may lead to disciplinary action. Efforts are ongoing to address these challenges through regular follow-ups and improved coordination with relevant departments.

**13.40 - 13.55**

Scope of examinations for appraisal - UEMS/ UEMS aligned (e.g. cardiac echo/ embryology) What does alignment mean

* The UEMS Council supports the development of European specialty exams as well as other exams aligned with UEMS rules and standards,
* An exam is considered "aligned" if it meets any of the following
  + Follows UEMS appraisal standards
  + Is linked to a UEMS Section or MJC (Multidisciplinary Joint Committee)
  + Is designed for medical doctors who have completed a medical degree
  + Typically originates from a European country

Alignment enables CESMA to appraise exams that are under the umbrella of a UEMS Section, even if they are not officially labelled "UEMS exams."

Concerns Raised:

* Dr. Chris Plummer: In Cardiology, the specialty includes formal training and certification before independent practice. Clear distinction is necessary between general appraisals and specialist certification.
* Dr. Feliciano Ramos: Raised concerns about the validity and recognition of diplomas issued from appraised exams not clearly within a UEMS-recognized specialty. He questioned whether an exam that falls outside a recognized specialty should be appraised at all.
* Dr. Vesna Kusec: Urged adherence to the Rules of Procedure and stressed that CESMA should focus on exclusively medical specialties and avoid diluting the definition.
* Dr. Krysta Krysztof: Highlighted variations across Europe in how specialties are structured (e.g., Maxillofacial Surgery training for dentists vs. doctors). Warned that appraising exams like those for dentists might send mixed messages about CESMA's scope and intentions.
* Dr. Arthur Felice: Clarified that exams appraised under CESMA must be officially connected to a UEMS Section or MJC, as recognized by the UEMS Council.
* Dr. Feliciano Ramos (follow-up): Raised the issue of liability, especially when CESMA appraises exams that may not be universally recognized or understood as medical specialties.

Next Steps / Recommendations:

* Consider a clear definition and official list of eligible specialties for appraisal.
* Possibly update or reinforce the Rules of Procedure to clarify:
  + What constitutes alignment
  + Which exams fall within CESMA’s remit
* Future discussions needed on recognition, liability, and value of appraised diplomas.

**13.55 -14.40**

Selecting a Provider for Your Exam:

Define the Purpose of the Exam

* Is the exam assessing:
  + Knowledge only?
  + Competency?
  + Or both?
* Clarify whether the exam is meant for:
  + A European-level certification
  + A national summative exit examination
  + A formative assessment

Curriculum and Alignment

* Is the exam aligned with the European Training Requirements (ETR)?
* Is there a clear blueprint that maps exam content to:
  + Curriculum goals?
  + Clinical competencies?

Format of the Exam

* What is the appropriate structure?
  + MCQ (Multiple Choice Questions) – single-stage or multi-stage?
  + OSCE, oral exams, or clinical scenarios?
* Will there be one stage (e.g., only a written test) or three stages (written + clinical + oral)?
* Should the exam be training-aligned or independent?

Eligibility Criteria

* Who can take the exam?
  + Medical doctors only?
  + Trainees at a specific stage?
  + Specialists from non-EU countries?
* Determine whether the exam should be open or restricted based on:
  + Level of training
  + Geographic or institutional affiliations

Standard Setting Methods

* How will the pass mark be determined?
  + Based on the exam’s purpose and the expected competence level of candidates.
* Consider different standard-setting methods:
  + Angoff Method – panel estimates minimum competence per question
  + Modified Angoff – incorporates performance data to adjust difficulty
  + Hofstee Method – blends expert judgment with acceptable failure rates and score ranges

Which is the best exam?

There is no *one-size-fits-all* answer, but the best exam is one that is taken at the right time. Measures what it claims to measure by knowledge, clinical reasoning, technical skill, or professional behaviour—aligned with the curriculum and ETR. Uses a reliable and fair scoring system with pass marks must reflect the minimum competence required, and not just a fixed percentage.

Setting the Pass Mark

* Should be based on:
  + Difficulty of the exam
  + Level of expected competence
  + Profile of the candidate (e.g., final-year trainee vs. recent graduate)
* Common methods:
  + Angoff or Modified Angoff: Expert judgment-based
  + Hofstee Method: Combines expert opinion and acceptable failure rates

The pass mark must be defensible, reproducible, and consistent with exam purpose.

Updates from 4 groups - 5 minutes each.

## 14.40 - 15.00 Coffee Break

**15.00 - 15.20**

What happens when your platform collapses

Online exam- appraiser enters as candidate and can cheat\* On site vs off site

When CESMA appraisers review an exam, they may:

* Attend in person or virtually to observe:
  + The delivery process
  + The professionalism of staff and invigilators
  + Candidate behavior (e.g., cheating, stress management, rule compliance)
* Evaluate:
  + Exam integrity
  + Clarity of instructions
  + Security measures
  + Candidate identification procedures

Common issues include:

Online Exam Delivery Failure

* Server crashes, internet issues, login problems
* Can cause panic and unfairness

Back-Up Plans Required

* Extra time allocated
* Technical support on standby
* Contingency protocol for partial completions

Security and Cheating

* Identity fraud, collusion, external communication
* Proctors must be trained to monitor behaviour
* Use of AI proctoring, browser lockdown, or live invigilation

Cheating & AI – Major Concern for Online MCQs

* MCQs online are highly vulnerable to AI-based cheating.
* Tools like ChatGPT can now pass Single Best Answer (SBA) exams by exploiting repeated question formats.
* Future AI may detect AI-generated answers, but risk is growing.

Mitigation Strategies:

* Randomize answer order and question pools
* Use scenario-based MCQs, less prone to AI
* Include live-proctored oral components
* Implement exam analytics to detect outliers and guessing patterns
* Contracts should include clauses on candidate misconduct and result invalidation.

On-Site vs. Online Exams – Strategic Choices

* COVID accelerated online exam adoption, but it wasn’t driven by cost alone.
* Many Boards still prefer on-site formats for better supervision and integrity.
* Decision depends on:
  + Type of competencies being assessed
  + Risk tolerance
  + Candidate pool (international vs. local)

**15.20 - 15.35**

OSCE exam - are they sustainable in the long run

Challenges & Constraints

High Cost:

* + OSCEs are resource-intensive: venue, materials, standardized patients, multiple examiners.
  + Travel and accommodation (for both candidates and assessors) add to costs.
  + Sustainability is questioned, especially when scaling to large groups (e.g., 40–50 candidates).

Human Resource Demands

* + Require manpower across all stations: often 2 examiners per station (one asks, one observes).
  + Significant planning to maintain examiner consistency and standardized performance.

Logistics and Flow

* + Smooth movement between stations is essential.
  + Timing and coordination of 6–10 rotating stations.
  + Pre-determined candidate numbers critical for efficiency.

|  |
| --- |
|  |
|  |
|  |

Will OSCEs Become Obsolete?

OSCEs assess competencies that MCQs cannot:

* Communication
* Professionalism
* Clinical reasoning
* Procedural skill

But it depends on:

* Smart design
* Financial planning
* Examiner engagement
* Evolving tech solutions

**15.55 - 17.40**

AI in examinations - friend or foe

The Positive Role of AI in Examinations

* Efficiency and Accuracy: AI can automate grading, especially for objective tests, reducing human error and saving time for educators.
* Personalized Assessments: Adaptive testing powered by AI can adjust question difficulty in real time based on student responses, offering a more accurate measure of ability.
* Proctoring and Integrity: AI-based remote proctoring tools help maintain examination integrity by monitoring eye movement, background noise, and screen activity, deterring cheating.
* Accessibility: AI tools can assist students with disabilities, offering voice-to-text, translation, or personalized interfaces to make exams more inclusive.

The Risks and Concerns

* Privacy and Surveillance: AI proctoring tools raise concerns about excessive monitoring, data privacy, and the psychological stress of being constantly watched.
* Bias and Fairness: Algorithms can reflect and amplify existing biases, potentially leading to unfair assessments or unequal treatment of certain groups.
* Cheating with AI Tools: On the flip side, students can exploit AI-powered tools (like ChatGPT or translation apps) to cheat during exams if proper safeguards are not in place.

Presentation 1. & intro - Danny Matthysen: 15 min

AI in medical

Give some thought.

Exercise.

AI in Medical Exams: Challenges and Opportunities

Presentation: Prof. Chris Plummer (15 mins)

* AI for Exam Creation:
  + Generates high-quality multiple-choice questions (MCQ’s), clinical vignettes, and essay prompts.
  + Saves time: What takes an expert 1 hour/question can be shortened to minutes.
* AI for Grading:
  + Natural Language Processing (NLP) for automated essay marking.
  + Objective, fast, and scalable.
* Multilingual Support:
  + Enables translation of exams into multiple languages, broadening accessibility.
  + Useful where English is not the primary language.

Challenges

* Validity and reliability must be preserved.
* Still requires expert validation and human oversight.
* Difficulty in assessing professionalism, communication, and clinical judgment.

Presentation 3 (Zoom): Morne Wolmarans – *AI and Cheating* (15 mins)

Focus: *How AI is undermining exam integrity and how to fight back.*

1. Essay Writing – AI generates convincing, plagiarism-free answers.
2. Real-time MCQ Solving – Bots or AI models quickly respond to live questions.
3. Wearable Tech – Smart glasses, Bluetooth earbuds, covert mics/cameras.
4. Identity Fraud – Use of fake IDs or stand-ins for online exams.
5. Plagiarism and Rewriting Tools – AI tools that paraphrase to evade detection.

Real Cases

* AI outperforms students in essay questions and diagnosis scenarios.
* Fake ID students have passed exams using AI help – major ethical risk.

Combating AI Cheating: What Can Be Done?

Tech Safeguards

* Proctoring tools:
  + Facial recognition
  + Locked browsers
  + Keystroke analysis
* Secure platforms:
  + Multi-layered encryption
  + Real-time AI detection of suspicious behavior
* Audits & analytics:
  + Post-exam data patterns can reveal anomalies.

AI as a Tool for Equity

Multilingual Capability

* AI can translate exam content reliably across languages.
* Promotes inclusion, especially in EU contexts with diverse linguistic backgrounds.

Beyond Cheating: AI as a Teaching Assistant

* Real-time feedback during practice exams.
* Personalized preparation strategies.
* Formative assessment tool, not just summative.

*“AI is not the enemy — misuse is.”*

* Use AI to enhance quality, not replace judgment.
* Invest in technology, training, and trust.
* The goal: valid, fair, and secure assessments for future medical professionals.

**17.40 – 18.00**

Ǫ G A

Dr. Danny Mathysen: "It can be difficult to cheat if the exam has multiple parts." He emphasizes that multi-part exams (e.g., written MCQs, oral interviews, OSCEs) are inherently more secure and reduce the risk of cheating, particularly compared to single-format online assessments.

Why multi-part exams are safer:

* Diverse formats test knowledge, reasoning, and communication skills.
* Face-to-face or live interactions are harder to manipulate.
* Cheating in one part (e.g., MCQ) doesn’t guarantee success in another (e.g., oral exam).
* Requires greater preparation, discouraging shortcuts like AI-assisted answers.

19.30 CESMA DINNER :

Restaurant ***Aux Armes de Bruxelles****,* Rue des Bouchers 13, 1000 Bruxelles

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Day 2: Saturday December 7th, 2024

***09:30 – 16:30\* (\*All times are local)***

***Chairs****:* ***Maeve Durkan, Ambrogio Fassina, Danny Mathysen, Albert Mifsud, Gian Battista Parigi***

## Commercial companies exclude 9.30 to 11 closed meeting.

**09.30-09.40**

Introduction

The meeting began with a warm welcome from CESMA President Maeve Durkan. She invited all participants to introduce and re-introduce themselves in turn, allowing everyone to share their name, affiliation, and role.

**09.40 - 09.50**

Update from President

Talked about the ‘trainer trainer’ around Europe

A discussion was held regarding the current landscape of trainer ‘training programs across Europe. Participants shared insights into how different countries implement and structure training for medical educators and examiners.

It was noted that while some regions have well-established, mandatory training programs for trainers involved in medical education and assessments, others still rely on more informal or voluntary systems. Concerns were raised about the lack of consistency and standardization in trainer preparation across member states.

The group highlighted the importance of harmonizing trainer training to ensure fairness, quality, and reliability in medical examinations. It was agreed that CESMA could play a valuable role in promoting best practices and possibly contributing to a common framework or recommendations for trainer development at the European level.

**09.50 – 10.00**

Update rules of appraisal – Danny Mathysen

* New applications must now include two proposed candidates for appraisal (reciprocity of exam appraisal) contribution.
* New appraisal templates are introduced with more detailed information required.
* Appraisers are encouraged to offer greater support and availability to CESMA, given the growing demand.
* For online exams, it is essential to always have a backup plan. Insurance may help but legal advice is often required in case of serious issues.

Dr. Maeve Durkan underlined the issue of security breaches or significant flaws in the exam delivery that can result in annulment of the exam; such cases must be explained, and centers must fix the issues before re-applying. Dr. Danny Mathysen, then strongly supports creating a comprehensive guideline to help appraisers and exam organizers respond effectively to issues and criticism.

Appeals & Fair Process

Dr. Maeve Durkan:

* Candidates can appeal the appraisal process, but not the exam result itself.
* The focus of appeals should be on the exam construction process (e.g. quality of questions, consistency, fairness).

Dr. Ashraf Butt:

* Candidates may challenge the exam if there is evidence of bias or administrative misconduct.
* This includes how questions were asked and how the exam was conducted.

Dr. Felician Ramos:

* Raised concern about language issues, particularly for non-native English speakers.
* Need for clarity and perhaps language accommodations or support, depending on the context.

**9.50 - 10.10**

Update from the UEMS Autumn Council by Bertrand Daval (CEO UEMS)

* The last Council meeting was a joint meeting with UEMS-EMO.
* Key outcomes:
  + Adoption of the European Training Requirements (ETR).
  + Endorsement of the UEMS Planetary Health initiative.
* Regarding constitutional matters:
  + New updates on Belgian law are ongoing.
  + An explanatory note is currently being prepared.
* Decision to create a new UEMS Section in Sports Medicine.
  + This new section is expected to be an important asset to UEMS.
* Presentation from Montenegro, for admission to UEMS
* The first UEMS Congress will be held in Leuven in 2026, consolidating previously separate conferences.

Additional Points: EPAs & Training

* Dr. Maeve Durkan spoke about integrating EPAs into evaluations and ETR.
  + The biggest challenge is when and how to develop EPAs.
* Dr. Chris Plummer emphasized the importance of assessing skills and expanding tools like case discussions and success metrics tied to EPAs.
* Dr. Danny Mathysen explained the “Train the Trainer” program in Belgium:
  + Trainers must complete a specific training course to remain certified.
  + The program includes workshops with role-playing to simulate real-life scenarios.
  + Workshops typically last 2 to 3 hours, although the law regarding the exact length is somewhat vague.
  + There is interest in possibly inviting workgroup members to share the Belgian model for ideas.

**10.10 - 10.25**

Treasurer report

* For the first time, the numbers are positive showing a favourable financial trend.
* Refer to Battista Parigi’s slides for detailed insights on the budget trend from 2017 to 2024.
* The current treasurer will be retiring in May 2025.
* If anyone is interested in the position:
  + You need to submit a CV.
  + Also, provide a letter of endorsement from your National Medical Association (NMA).

**10.25 - 10.40**

Presentation and awards appraisal certificates

The following specialties have received their final appraisal certificates:

* Regional Anaesthesia and Pain Management (CESMA-certified: ESRA-EDPM)
* Medical Genetics and Genomics
* Neuroradiology
* Paediatric Neuroradiology
* Interventional Neuroradiology
* Hermes-ERS-2 (Adult and Paediatric Respiratory Exam)
* Neurology
* Hand Surgery

## 10.40-11.00 Coffee break

**11.00 – 12.15**

Session of Q&As with UEMS Lawyer: Me Adrien Hanoteau

The legal perspective

Session of Q&A with UEMS Lawyer – Me Adrien Hanoteau

Introduction:  
Bertrand gave a quick introduction.

Legal Concepts & Advice on UEMS Exams

* Liability of UEMS:  
  UEMS faces some liability concerns in organizing exams (e.g., candidate issues, technical problems). Although no major legal disputes have occurred, zero risk doesn’t exist. It is important to be cautious.
* Risk Mitigation:  
  Always seek amicable resolutions to disputes as legal procedures are long and costly.
* Legal Concepts at Stake:  
  Civil liability can be contractual or extra-contractual. Victims must prove the harm and ideally be restored to the previous situation.
* Contractual Protections:
  + Only Belgian law applies, and one jurisdiction is chosen.
  + Contract clauses limit UEMS's commitments and liabilities strictly to those in the contract.

Questions & Discussions (around 28 minutes in)

* Candidate Guidelines & Appeals:
  + TT suggested candidates sign a guideline document.
  + Dr. Felice discussed using the correct terminology for cheating under Belgian law; Maître advised caution with the word “cheating.”
  + Dr. Mathysen questioned if candidates can appeal after the appeal period; Maître confirmed late appeals can be limited but possible.
* Recognition of UEMS Diploma:
  + Dr. Battista Parigi recalled a lawsuit by the German government against UEMS over diploma recognition for a non-European candidate.
  + The diploma is a “quality mark,” not a legal qualification.
  + Discussion about clearly stating that commission decisions are final and not subject to further dispute.
* Jurisdiction Clause:
  + Dr. Durkan emphasized the need to specify jurisdiction in contracts, ideally Belgium.
  + Maître confirmed Belgium law is applicable, but choice of jurisdiction can be flexible; Swiss jurisdiction example given for EBNR.
* Liability and Appeals:
  + If a system failure occurs during an exam, responsibility lies with the third-party provider.
  + Dr. Battista Parigi requested UEMS draft liability templates for all CESMA members.
  + Dr. Felice stressed the need for clear legal liability allocation—only legal entities can be sued, never individuals personally (except for criminal liability).
  + Dr. Durkan recalled shared liability in MoU agreements.
* Eligibility Criteria & Documentation:
  + Dr. Serdar Ceylaner raised issues about eligibility for Medical Genetics exams and submission of diplomas.
  + Clarified differences in eligibility by specialty.
* Volunteers and Intellectual Property:
  + Dr. Durkan noted volunteers who write questions must have agreements covering confidentiality, intellectual property, data protection.
  + Maître confirmed these should be included in common agreements, ideally at the Board level.
  + Mister Daval believed liability lies mainly with UEMS President, but volunteers are not exempt from individual liability, especially for criminal acts.
  + Shared accountability applies for quality assurance of question writers.

**12.15 -12.45**

The candidate who challenges the result

When a candidate challenges the exam outcome, a robust and fair appeal process must be in place:

* Clear appeals policy, communicated before the exam
* A defined time window to submit a challenge
* Process for review by independent examiners
* Transparency about what aspects can be appealed (e.g., marking error vs. disagreement with result)

## 12.45-13.45 CESMA Lunch

**DME Ground Floor**

**13.45 -14.15**

AI

Danny Mathysen (10-15 min) presentation with general introduction of AI in medicine

* Quote to Reflect On:  
  “I am not young enough to know everything.” – Oscar Wilde  
  Used to highlight the humility and evolving nature of expertise, even in AI.
* Definition and Context:  
  Artificial Intelligence (AI) in medicine refers to the simulation of human intelligence processes by machines, particularly in areas like diagnostics, treatment planning, and predictive analytics.

History and Development

* Outlined the historical evolution of AI in healthcare.
* Emphasis on how initial skepticism has turned into cautious acceptance as tools become more accurate and supportive.

Case Example: AI-Driven Safety Improvement

* Video Reference:  
  A video from the Faculty of Medicine and Health Science featuring Michael Hugh Mosley was shown.
* Case Summary:
  + A child was misdiagnosed with chickenpox in 1999.
  + The child’s actual condition was more serious, and the misdiagnosis contributed to their death.
  + This tragic error prompted the development of clinical checklists, which laid the groundwork for AI-assisted diagnostic tools.
* Impact:
  + Today’s AI-based checklists have reached high levels of accuracy, significantly improved diagnostic safety and reduced preventable errors.

**AI in Medical Education**

**AMME - GDPR confidentiality - using AI to dictate your notes**

Use of AI for Dictating Notes

* Highlighted the growing practice of using AI tools (e.g., speech-to-text) to dictate clinical and educational notes.
* Emphasis on ensuring compliance with GDPR and confidentiality protocols, especially when dealing with patient data or sensitive educational records.
* AI transcription tools must be:
  + Secure
  + Encrypted
  + Compliant with EU data protection regulations

Application of AI in Medical Education

* A series of recently published articles on AI in education were reviewed.
* Some publications (from 1.5 years ago to present) were later found to be inaccurate or scientifically flawed.
  + Demonstrates the need for critical evaluation of AI-related research.
  + Encourages educators and clinicians not to take AI claims at face value.

Insights from AMEE (Association for Medical Education in Europe)

* AMEE published a practical guide on implementing AI in medical education and research.
  + Encourages the thoughtful and ethical integration of AI technologies.
  + Emphasizes collaboration with IT professionals and researchers.
  + Guide includes:
    - AI fundamentals
    - Use cases in teaching and assessment
    - Ethical and legal frameworks
* The AMEE AI Guide will be:
  + Linked in Danny Mathysen’s slides
  + Published by Amelia Donighian for broader access.

Ethical Considerations

* Urgent need to raise ethical standards for students and faculty using AI.
* Recommendations include:
  + Establishing Research Ethics Committees that are AI-literate.
  + Appointing a Chief AI Officer or equivalent AI governance role.
  + Incorporating AI ethics education into medical curricula.
  + Promoting responsible AI use and awareness of potential biases, limitations, and misuse.

**14.15 – 15.00 Debate: Danny Mathysen vs Chris Plummer**

AI is the best to happen in medicine VS AI is the worst in medicine

* Multidisciplinary Collaboration
  + *Dr. Celine Carrera*: AI encourages a cross-disciplinary approach and highlights the need for ethical training in AI usage.
* Efficiency & Speed
  + *Dr. Thomas Giesen*: AI accelerates processes, making diagnosis and decision-making faster and more accurate.
* Enhanced Translation & Communication
  + *Dr. Dragos Bumbacea*: AI tools improve translation, helping in multilingual exams, cross-border collaboration, and patient communication.
* Reach & Accessibility
  + *Dr. Tali Turgut*: AI can extend access to care, particularly in radiology and diagnostics, reducing waiting times and addressing shortages.
* Existing Use in Medical Devices
  + *Dr. Ashraf Butt*: AI is already embedded in tools like MRI and CT scanners, and its presence in medicine isn’t new—just more visible now.
* Promoting Digital Literacy
  + *Dr. Patricia Peralta*: AI tools like prompting, summarization, and simulation can empower educators and learners to adapt more quickly.
* Pattern Recognition
  + *Dr. Chris Plummer*: AI can detect subtle patterns in large datasets that are invisible to human cognition, enabling new discoveries.

BAD Aspects of AI in Medicine

* Bias & Limited Scope
  + AI is only as good as the data it is trained on—this can lead to biased or exclusionary results, especially in underserved populations.
* Overreliance Risk
  + *Dr. Dragos Bumbacea*: There’s a risk of depending on AI too much, losing the critical ability to distinguish between what is learned vs. what is found.
* Lack of Human Perception
  + *Dr. Patricia Peralta*: AI lacks empathy, nuance, and intuition, which are crucial in complex medical decisions.
* Ethical Deficiencies
  + AI systems operate based on algorithms, not human values or ethics. This presents challenges in patient-centered care.
* Erosion of Trust
  + *Prof. Plummer*: Relying on machines may undermine patient trust, especially if AI decisions are opaque or appear impersonal.

Reflections & Conclusions

* AI is Already Integrated: From diagnostics to workflow automation, AI has been silently embedded in medicine for years—its growing visibility brings new scrutiny.
* Balanced Integration is Key: The goal isn’t to choose between humans and AI, but to blend both strengths. Human judgment + AI efficiency = better care.
* Ongoing Research Needed:
  + *Dr. Mathysen*: We must study how AI impacts medical exams, assessments, and training systems.
  + Develop evidence-based strategies for incorporating AI ethically and safely.
* AI is Inevitable—but Manageable: As *Tali Turgut* noted, it’s unavoidable, but if used wisely, more than 50% of its applications can be beneficial.

**15.00 -15.15**

Chris Plummer- experience of cardiology with AI and European exams.

* Objective: Assess ChatGPT’s performance on the European Exam in Core Cardiology (EECC).
* Initial Reaction: High performance by ChatGPT triggered concerns—*Was there cheating? A databank breach?*
* Dataset:
  + 500 exam questions were selected.
  + Image-based and video-based questions were excluded, as ChatGPT-3.5 cannot process visuals.
  + Final dataset: 362 text-only questions.
* Process:
  + Questions were input into ChatGPT.
  + ChatGPT responses were compared to the correct answers.
  + The aim was to simulate performance under standard exam conditions.

Conclusions

* AI (ChatGPT) would not currently pass the EECC.
* The performance was impressive but not threatening, as the exam still tests higher-order thinking and image interpretation.
* No evidence of cheating or databank breach was found.
* The study’s methodology lacked rigor, and conclusions should be taken cautiously.

Policy Implications

* AI use by candidates is not allowed during the EECC or similar exams.
* Need to monitor AI developments and update exam protocols accordingly.
* Raises an important discussion about future-proofing assessments against AI-assisted tools.

Slides of Interest:

* See Slides 13–15 in Dr. Plummer’s presentation for data and detailed results.

**15.15 -15.35**

Danny Mathysen - mock exam in AI (experience in Antwerp)

Exam Setup

* Structure: 2-part exam
  + Part 1: Oral
  + Part 2: MCQ with 100 questions
* Goal: Test AI performance against a standard medical exam framework.

AI Bots Used

* Tools evaluated included:
  + ChatGPT-3.5 & GPT-4
  + Microsoft Bing
  + Bard Palm 2
  + Claude+ / Claude Instant
* Accessed mostly via free versions, including through Poe (multi-bot access platform).
* Limitations noted: No image processing capabilities; hallucinations were frequent.

Translation & Local Context

* Questions were translated from Dutch to English using AI tools.
* Manual review followed, correcting for local context or translation errors.
* Some questions were tailored for the Belgian educational system, potentially influencing results.

Observations & Pitfalls

* Errors & hallucinations:
  + Example given: Incorrect and mathematically flawed responses to a basic risk assessment question (e.g., Down syndrome risk in a 40-year-old woman).
* Takeaway: AI can mimic accuracy but still makes critical reasoning errors.

Guidelines & Ethical Concerns

* AMEE guidelines on AI in medical education were emphasized.
* Danny urged CESMA members to read them thoroughly and identify missing aspects relevant to exam integrity.

Discussion Points

* Maeve: Question writing must still be anchored in human expertise with compliance checklists.
* Danny: Questions should be reviewed by both experts and non-experts to ensure balance and clarity.
* Dragos Bumbacea: Proposed evaluating future outcomes of AI-trained vs. non-AI-trained students.
* Danny: Stressed the need for long-term studies to assess AI’s impact on medical education.

Final Note

* Call to action: CESMA must remain at the forefront of defining quality and ethical standards in exams amid the evolving role of AI.
* Bottom line: AI is here—not to replace clinicians or examiners, but to challenge and improve how we teach, assess, and maintain trust in medical qualification.

**15.35 - 15.50**

Using AI in exam -development and delivery - where are we going?

Survey of Examination Portfolio (Presented by Dr. Durkan)

Responses: 27 (from 24 UEMS Sections/MJCs)

Exam Format

* 1-part exams (MCQ-only): 6 Sections
* 2-part exams (MCQ + oral or OSCE): 14 Sections
* 3-part exams: 4 Sections

Exam Delivery (Pre- & Post-COVID)

* Pre-COVID: Primarily on-site
* Current status:
  + 8 Sections: Fully on-site
  + 3: Off-site via Pearson VUE/training centers
  + 2: Mixed (on-site + off-site)
  + Remaining: Exclusively online

Exam Costs (Examples)

* National exam: Free
* European candidates: €250–€550
* Non-European: Up to €920
* Costs often depend on:
  + Training location (EU vs. non-EU)
  + Number of exam parts (2-part: €900, 3-part: €1200)

Training Origin of Candidates

* 6: Only EU-trained
* 6: >80–90% EU-trained
* 5: ~2/3 EU-trained
* 2: <1/3 EU-trained

Partnerships & Key Concerns

* Collaborations with scientific societies:
  + Sharing: IP rights, costs, control, liability, profits
  + Managing Pan-European standards
* Legal and standard issues arise with non-European candidates

Discussion Points

Dr. Tali Turgut: Requested survey results be shared.

Dr. Arthur Felice: Support for benchmarking strong non-EU candidates for employability.

Dr. Joseph Briffa: Concerned about verifying non-EU training standards — risk to exam integrity.

Dr. Ashraf Butt: Diplomas serve as validation of current standards; peer acceptance is key.

Dr. Octavian Toma: UEMS diplomas (e.g., in anaesthesiology) are accepted internationally (e.g., Abu Dhabi).

Dr. Battista Parigi: Clarified CESMA’s founding principle – *exams are open to all*:

* EU = "Diploma of..."
* Non-EU = "Certificate of..."
* Suggested using AI to update this communication clearly (per *Glasgow Declaration*).

Standard Setting & Validity

Dr. Maeve Durkan: Cited *Royal College of Physicians of Ireland*'s standard setting approach:

* Plan and standard set over 2 days
* Identify "minimally competent candidate"
* Rate difficulty of questions
* Discuss borderline cases

Dr. Danny Mathysen:

* Warned against flawed pass/fail logic if 1–2 questions decide the outcome.
* Suggested poor passing rates (e.g. 40%) signal problems in standard setting or training.
* Professors often fail the exams meant for trainees → reflects on question quality.

Database & Score Benchmarking

* Dr. Ambrogio Fassina: Proposed creation of a shared question database, including historical data.
* Dr. Tali Turgut: Asked for pass score suggestions to be developed and shared.
  + Action point recorded.

Call for Engagement

Dr. Maeve Durkan: Invited members to email ideas, proposals, or contributions

Reflections on AI in Exams

* Integration of AI must address:
  + Quality assurance
  + Fairness across geographies
  + Ethical use in question setting and evaluation
* AI may assist in clarity, translation, standard setting, but human oversight remains crucial.

**16.00 – 16.15**

Provisional Meeting Dates and Locations

* 2025 Meeting:
  + Dates: 2–3 May 2025
  + Location: Cork, Ireland
  + Travel Connections: Accessible via KLM, Ryanair, and Aer Lingus.
* 2026 Meeting:
  + Host: Feliciano Ramos
  + Location: Zaragoza, Spain

AOB

We look forward to having you with us at the CESMA Autumn Meeting 2024!