A Community of Practice Approach to an Effective AI Working Group

Prof Peter Schmidt, Chair Academic Board, Holmes Institute
Prof Hamish Coates, Assessment Consultant, Australian National University

Coping with the ‘challenge’ / ‘threat’ of AI

The magnitude of challenge has been compared to the metaphor of painting the Sydney Harbour Bridge (Prince’s Bridge for Melbourne audiences)

Adjustment is required to the assessment regime of every single unit but also to the curriculum of every single unit

It is clear that considerable time and effort is required, and this is an ‘in-flight’ change since teaching and assessment relentlessly continue

Coping with AI: Assessment Review and Curriculum Review

- June 2024: All 202 HEPs, large and small, submitted plans to TEQSA. The relief was palpable
- November 2024: TEQSA's AI-Tool Box mandated, for every HEP, the establishment of a **Working Group** given the ongoing challenge posed by AI
- Since then?

A progress review is overdue. **Tech has moved, what about education?**

Developing an Effective AI Working Group

- Clear to several HEPs that we should ‘join forces’ and that CABF, (a classic Community of Practice), was set up precisely to address challenges of this size and complexity
- Our Working Group (mandated by TEQSA) would be a benchmarking group
- Sharing our hard-won Intellectual Property (IP)?
  - Could be vital for a small HEP
  - But what’s in it for us if we’re a larger HEP?
    Community Service Obligations; Social Licence
- That we were able to ‘pull this off’ at all demanded an extra-ordinary level of sharing which may have been too much for some:
  We were five; we are now three But: no henny penny moments yet!

Developing an Effective AI Working Group

- With MOUs and NDAs in place, we met several times the first key question was how would be go about the work?
- One of our members volunteered to develop and propose a suitable Methodology

First Step: Deciding on a Methodology for the Working Group

- **Project Title:**
  - "Benchmarking AI Education in College Curricula: A Comparative Study"
- **Project Objectives:**
  - Assess the current AI curriculum in colleges and benchmark it against industry and academic standards.
  - Compare AI courses across institutions based on content, pedagogy, and real-world applicability.
  - Identify gaps and best practices in AI education.
  - Provide recommendations for improving AI education in colleges.
- **Scope of the Study:**
  - Focus on undergraduate AI courses in computer science, engineering, and interdisciplinary programs.
  - Analise core subjects, electives, hands-on learning, and research opportunities.
  - Include benchmarking against industry certifications and top global AI programs.

HOLMES INSTITUTE

First Step: Deciding on a Methodology for the Working Group

- **Data Collection:**
  - Review the syllabi from multiple colleges.
  - Conduct surveys/interviews with students, faculty, and industry professionals.
  - Analise online AI course offerings (MOOCs, certifications).
- **Benchmarking Criteria:**
  - Core AI topics (Machine Learning, Deep Learning, NLP, etc.).
  - Practical applications (projects, industry partnerships, internships).
  - Assessment methods and pedagogy (theory vs. practical balance).
  - Emerging trends (ethics, AI safety, generative AI).
- **Analysis & Comparison:**
  - Compare curriculums across institutions.
  - Map AI courses to industry needs (job postings, skills in demand).
  - Evaluate hands-on experience (labs, competitions, hackathons).

HOLMES INSTITUTE

First Step: Deciding on a Methodology for the Working Group

- **Key Deliverables:**
  - Benchmarking Report summarizing curriculum strengths and weaknesses.
  - Comparative Matrix ranking AI programs based on set criteria.
  - Recommendations for curriculum enhancement.
  - Future-Proofing Guide with AI education trends.
- **Timeline & Milestones:**

  Week 1-2: Literature review & defining benchmarking criteria.
  Week 3-4: Data collection (syllabi, surveys, industry analysis).
  Week 5-6: Data analysis and curriculum comparison.
  Week 7: Drafting the benchmarking report.
  Week 8: Presentation of findings & recommendations.
- **Expected Impact:**
  - Enhanced AI education aligning with industry needs.
  - Improved hands-on learning opportunities in AI programs.
  - A roadmap for colleges to stay competitive in AI education.

Chat-GPT Prompt for a Methodology for the Working Group

"Please give a bullet point design of a project to study AI in a college curriculum suitable for a benchmarking exercise”

This underlines the seriousness of the AI challenge

HOLMES INSTITUTE

The Scope of Works

It was clear that the scope of works would be to address:

- Assessment Review
- Curriculum Review

But in doing so, issues of **scalability and cost** would arise and we would need to be alert to potential efficiencies that AI could provide:
In the words of Vice Chancellor Theo Farrell
"AI is in everything we do at La Trobe"

The scope, initially thought to be two-fold, evolved to three-fold

HOLMES INSTITUTE

Starting with Assessment Review: The Actual Methodology

- To limit the **academic/corporate risk**, each HEP would run a **pilot** of a small group of their units. Demonstrable success in the pilot should encourage Course Development Committees to adopt the approaches
- We decided on a **case approach** choosing units where the existing assessment regime was clearly in desperate need of change
  - Any changes we would make would be 'in-flight'
  - We quickly recognised an **adaptive approach** would be necessary The process of finding improvements might take a few iterations
    - § "Assessment in the age of GenAI is an ongoing negotiation, requiring continual adaptation rather than prescriptive one-size-fits-all solutions."
    - § "Requires a shift from seeking definitive answers to engaging in ongoing, adaptive work shaped by competing priorities and evolving conditions."

§ Thomas Corbin, Margaret Bearman, David Boud & Phillip Dawson (03 Sep 2025): The wicked problem of AI and assessment, Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2025.2553340
Centre for Research in Assessment and Digital Learning (CRADLE)

Assessment Review

AI is a threat to significant portions of current assessment regimes

- The essay
- The individual assignment

In response, in an effort to reinforce and prolong the use of the current assessment regime, a **punitive approach** has been the first reaction

The AI Challenge to Assessment Integrity

The Punitive Tech V Tech Approach

- Coming on top of the struggle with Contract Cheating, it is not surprising that there was a strong focus on **'the punitive approach'**
- **Key question:**
  Is *Turn-it-in* or *Safe-Assign* better at detecting the cheats?

- Some focus on authorship-detection software: CIA-style

The AI Challenge to Assessment Integrity

However, the second part of the AI challenge seeks to embrace AI in the curriculum i.e. Curriculum Reform so as to produce AI-literate, work-ready graduates.

Pushing back against AI, as in the punitive approach, did not seem appropriate to our group.

Towards a new Assessment Paradigm

Eschewing the Punitive Approach

§ "Opens up the possibility of more meaningful, inclusive, and future-relevant assessment practices. It makes space for **assessments that prioritise assuring learning over policing**".

§ Thomas Corbin, Margaret Bearman, David Boud & Phillip Dawson (03 Sep 2025): The wicked problem of AI and assessment, Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2025.2553340
Centre for Research in Assessment and Digital Learning (CRADLE)

‡ "Rather than investing primarily in detection mechanisms, institutions need to emphasise the redesign of assessment to capture authentic demonstrations of student capability and comprehension"

‡ Lodge, J. M., Bearman, M., Dawson, P., Gniel, H., Harper, R. Liu, D., McLean, J., Ucnik, L. & Associates (2025). Enacting assessment reform in a time of artificial intelligence. Tertiary Education Quality and Standards Agency, Australian Government, p3.

Towards a new Assessment Paradigm

Instead of trying to bolster the existing paradigm, we sought **a new assessment paradigm**

Assurance of learning being the measure of assessment effectiveness.

A new Assessment Paradigm?

Yes, it has happened before….

- The Exam Paradigm (1960s)

U of Q: 2nd Yr Pure Science	Morning	Afternoon
Monday	3 hour Exam	3 hour Exam
Tuesday	3 hour Exam	Free!
Wednesday	3 hour Exam	3 hour Exam
Thursday	Free!	3 hour Exam
Friday	3 hour Exam	3 hour Exam
Saturday	3 hour Exam	No exams

- Gave way to the more-enlightened Continuous Assessment Paradigm

A new assessment paradigm?

“Embodied Assessment”

The Working Group recognised that the actual persona of the academic needs reinstatement squarely into the emerging assessment paradigm.
[The innovative elements of this will be outlined later in this talk].

A new assessment paradigm?

“Embodied Assessment”

When all is said and done:

One-on-one interaction between the experienced academic professional and the student is the best guarantee of assurance of learning.

→ Beyond “situated” and “authentic” assessment

Seeking a new assessment paradigm

"Embodied Assessment"

But, there are obvious risks and costs

- **Risk**: First-mover disadvantage

Is there a case for a blanket directive by the Regulator so as to remove the disadvantage to the provider who moves first to tighten assessment regimes?

Seeking a new assessment paradigm

"Embodied Assessment"

Reinstating the actual persona of the academic squarely in the assessment regime incurs obvious costs

- **Cost:** as issues of **scalability and cost** arise and we need to be alert to potential efficiencies that AI could provide

Perhaps AI itself can enable the reduction of some costs
e.g. the cost of marking

Getting the robot-human balance right…**Marking Case Study**

Case Studies

Three Cases

- A Human-Robot Marking Case
- A post-graduate Unit in Project Management
- An under-graduate Unit in the Internet of Things

Tech platforms are NOT “HE”—same limitations for last 25 years!

NOT A LEARNING PLATFORM

- The fundamental purpose of an LMS is administration and management, not pedagogy.
- It serves as a tool for teachers to deliver content, but falls short of actively facilitating student learning.

RIGID

- The learning approach focuses on students receiving information rather than actively engaging with it.
- The educational method is not centered around the student's own exploration and discovery.
- The curriculum follows a rigid, one-size-fits-all structure instead of adapting to individual student needs.

LIMITED ASSESSMENT AND FEEDBACK FEATURES

- The assessment capabilities are limited.
- There is minimal support for different item types.
- Only simple question types can be graded.
- No feedback is provided.
- AI support is limited or nonexistent.
- There are significant operational overheads.

TECHNOLOGICAL STAGNATION

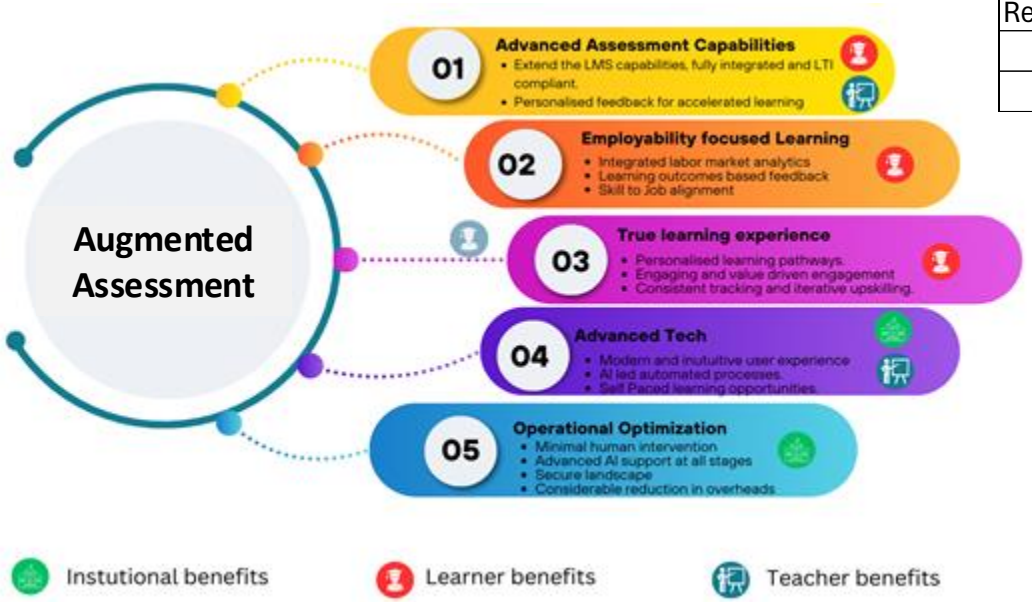
- Reliance on plugins and external services
- Using non-integrated tools, such as Microsoft Excel, can disrupt workflow.
- Vulnerable security environment
- Difficult and separate integrations
- Outdated and cumbersome user interface

Augmented assessment “solutions” have arrived

- AI-driven assessments categorise learning content, generate targeted questions, and support creation, delivery and grading of assessments
- Automated assessment task generation along with rubrics and marking guides
- Automated assessment of student responses against rubric delivering quantitative
- Detailed grading feedback to students of their proficiency and highlights areas for improvement and directs students to relevant content to facilitate improvement
- Evaluation of authenticity/integrity of student response
- Human review to validate and oversee task generation and marking
- Generation of customised models tailored to specific organisational needs
- Seamless integration with widescale education platforms

Assessment process decomposition comparison

Phase	Activity	Current		Augmented		Comparison	
		Human	Platform	Human	Platform	Quality	Cost
Development	Specifying outcomes	X		X	X	↑	↓
	Selecting task formats	X		X	X	↑	↓
	Drafting materials	X			X	↑	↓
	Specifying rubric	X			X	↑	↓
	Task quality review	X		X	X	↑	→
	Material production	X	X		X	↑	↓
	Administration planning	X		X	X	↑	↓
Implementation	Organising facilities	X		X	X	↑	→
	Student management	X	X	X	X	↑	→
	Administration	X	X	X	X	↑	→
	Resolving problems	X		X	X	↑	↓
	Collating results	X	X	X	X	↑	→
Grading	Marking performance	X	X	X	X	↑	↓
	Checking integrity	X	X	X	X	↑	↓
	Producing data	X	X	X	X	↑	↓
	Cross-validating results	X	X	X	X	↑	↓
	Producing grades	X	X	X	X	↑	↓
Reporting	Analysis and comments	X		X	X	↑	→
	Reporting	X	X	X	X	↑	→
	Reviewing and improving	X		X	X	↑	↓



Cost analysis

Parameter	Current context				Augmented
	Single	Semester	HEI/Semester	National	National
Student count	1	1	50,000	1,500,000	1,500,000
Unit count	1	3	3	3	3
Assessment count	2	2	2	2	2
Marking events	2	6	300,000	9,000,000	9,000,000
Marking dollars/event	$12.5	$12.5	$12.5	$12.5	$4
Marking hours/event	0.25	0.25	0.25	0.25	0.1
Marking dollars	$25	$75	$3,750,000	$112,500,000	$39,000,000
Marking hours	1	2	75,000	2,250,000	750,000

- Assumptions of similar sunk costs and risks
- Not adjusted for obvious quality gains
- Costs for first-run implementation not economies of scale
- Minimum upskilling and change costs
- JUST marking: Not development, validation, administration, management, governance, reporting, etc.
- Better, more reliable, more manageable

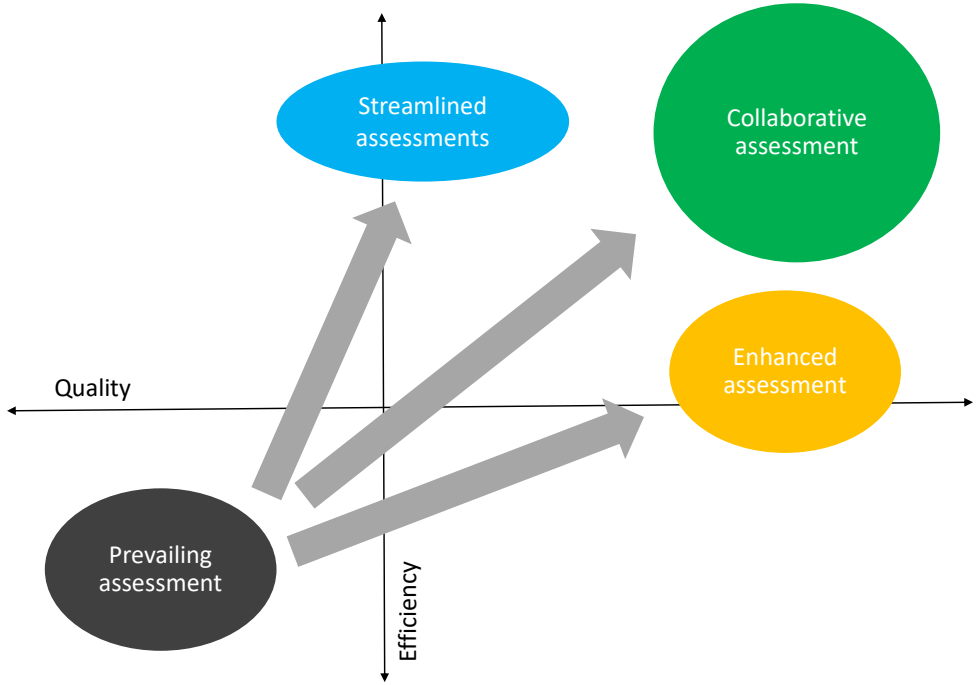
Main insights

- Strong alignment between human and robot scoring.
- The robot comments are far better than the human comments, though they need human quality control.
- Correlations between the robot and human scoring are influenced by idiosyncratic variations in marking due to localised conditions and procedures.
- Higher-level human oversight is needed to monitor quality and apply situational adjustments.

Decision space

Platform

- Regulatory/moral risk
- Disengagement



Current

- Error-prone
- Expensive
- Outdated

Augmented

- Risk-managed
- AI-supported
- Enhanced results

HOLMES INSTITUTE

CASE: A post-graduate unit in Project Management

New paradigm elements are highlighted in red font

Current	New
Two MCQ tests10% each: Total 20%	Three On-line tests: A hurdle, Formative assessment – diagnostics for tutor meeting
A Group Case Study: 40%	A Group Case Study: 50% all group work in tutorial and a Group Video Dialogue, A hurdle
An Individual Assignment: 40% [An essay in comparing project management methodologies]	An Individual Assignment: 50% Peer Review of designated Group Assignment Tests critical appraisal ability which is a major Learning Outcome on a local or novel source to which AI has no access

The AI Challenge to Curriculum Design

The second part of the AI challenge seeks to embrace AI in the curriculum so as to produce AI-literate, work-ready graduates

We see AI very much in the curriculum in this next case – very natural and easy for a unit from the IT curriculum

HOLMES INSTITUTE

CASE: An under-graduate unit: Internet of Things

New paradigm elements are highlighted in red font

Current	New
IoT System Fundamentals: 20%	IoT System Fundamentals 20%: Students use AI tools for research but must manually validate outputs.Peer critique sessions focus on clarity, technical accuracy, and ethical AI use
IoT Data Pipeline Design: 30%	End-to-End IoT Data Pipeline: 30% Allows AI support. Includes peer review other group's pipeline: robustness and ethical AI integration.
IoT Dashboard & Analytics: 50%	IoT Dashboard & Predictive Analytics: 50% AI used for predictive analysis and visualization recommendations. Peer evaluations focus on interpretability and responsible AI application.

HOLMES INSTITUTE

The AI Challenge to Curriculum Design

To produce AI-literate, work-ready graduates for business means embracing AI in the B-School curriculum,
To do that we need answers to questions like:

- What are banks doing with AI? What are legal firms doing with AI?
- What are the state and federal governments doing with AI?
- What are financial planners doing with AI?

Is AI a better financial planner than a human?

Is AI better at interpreting an X-Ray or MRI than a radiographer?

To answer these questions, we need to engage with recent successful graduates – the alumni - and mount a serious research effort

Insights from the Cases

Main Insights in the "Embodied Assessment" Paradigm
reinstating the actual persona of the academic in assessment

1. All group work to be undertaken on campus **in tutorials**
2. All key assignments subject to **live review of the tutor** [stratified semi-random sampling to face an audit]
3. MCQs generated from large pool and marked by the technology
4. MCQs may be used as a diagnostic tool **in formative assessment**
5. Peer Review by fellow students to be introduced

Another Option for the Assessment Review

The Assessment Designer might take advantage of the known weaknesses of AI to design assignments which would not be susceptible to AI-generated solutions

……. But at the current pace of change……?

HOLMES INSTITUTE

Exploiting AI-weaknesses

1. Cross Disciplinary Synthesis
   - Assignments designed across more than one domain

2. Personal Contextualisation
   - Students required to connect theory to their own employment

3. Process Transparency
   - Staged submissions

4. Introduce Ambiguity and Uncertainty
   - Assessments that require data from grey areas, emerging concepts or unresolved debates

Exploiting AI-weaknesses

5. Original Data and Artefacts
   - Require students to collect small data sets workplace interviews or document observations

6. Ethics and Critical Appraisal of AI
   - Require students to critique AI outputs and identify flaws and propose improvements

7. Novel or Local Sources
   - Assessment design around latest reports not in the AII domain e.g. Peer Review of other students work

HOLMES INSTITUTE

Summary and Conclusions

1. The AI Challenge cannot be ignored
2. Hoping to cope with AI impacts using technological solutions degenerates into an AI-arms-race which cannot succeed as AI continues to develop apace
3. HEPs would be wise to explore more than one line of change initiatives:
   e.g authorship software? Assignments designed in domains of AI weakness
4. Fundamental change will need to be **adaptive** – can't guarantee immediate success
5. **Assurance of Learning** is the key criterion of what constitutes effective assessment
6. Since first-mover change can be a disadvantaged, there may be a need for the Regulator to mandate across-sector directives to ensure first-movers are not disadvantaged
7. The 'new paradigm' will involve greater human innovation and presence

HOLMES INSTITUTE

Summary and Conclusions

Looking back, in years to come, the current situation will be understood as a period of turbulence associated with fundamental change

HOLMES INSTITUTE

Facilitated Dialogue

SURELY, Australia's Academic Governors must take national lead on this matter, and contribute reforms which make difference.

What agile, courageous and perspicacious steps can we take TOGETHER (OZCABS+CABF+TEQSA) to make productive change?