Changing Assessment in an Age of Artificial Intelligence: a principled approach

David Boud

Centre for Research in Assessment and Digital Learning, Deakin University

Emeritus Professor, University of Technology Sydney

Work and Learning Research Centre, Middlesex University

DEAKIN UNIVERSITY AUSTRALIA

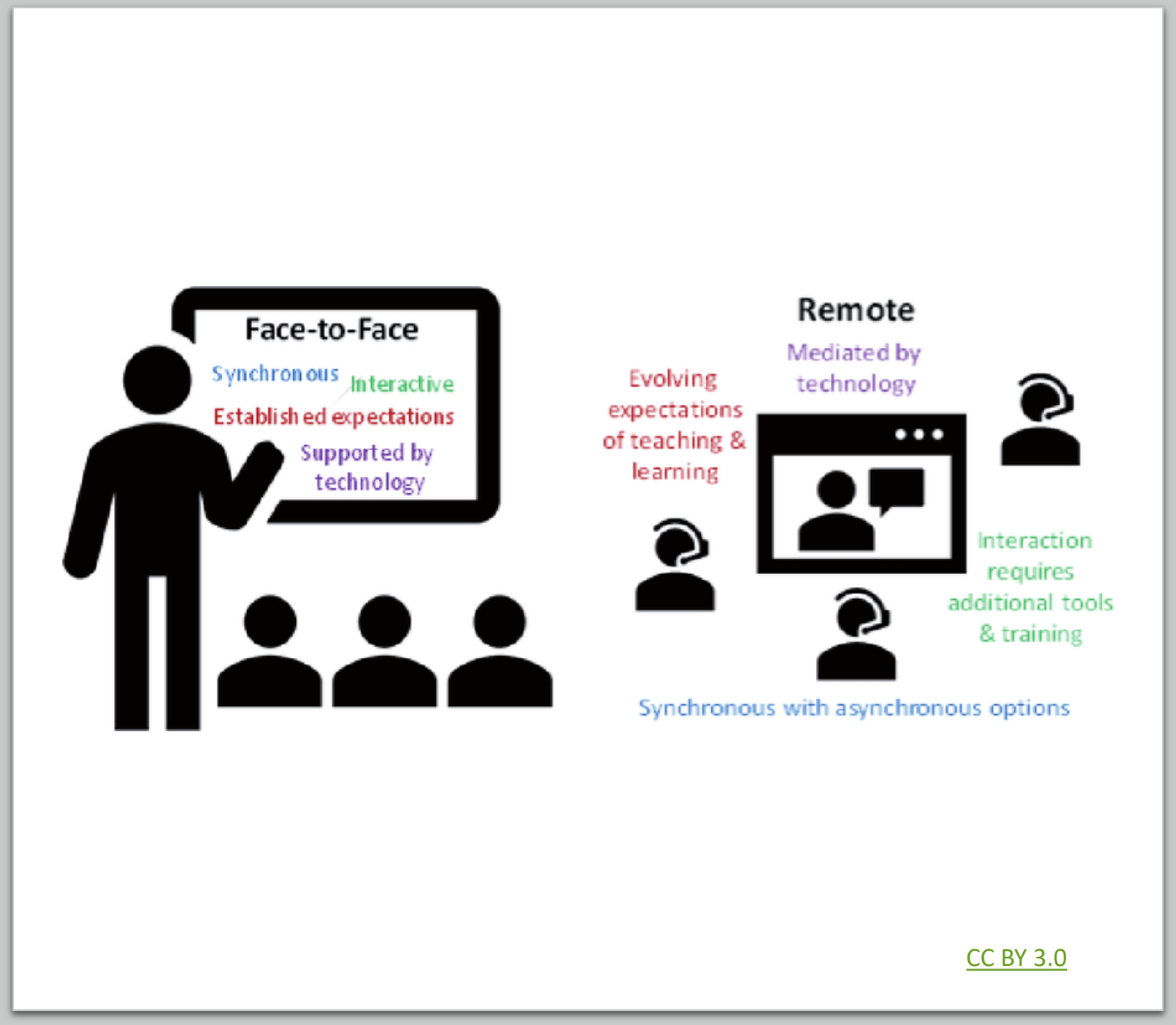
cradle Centre for Research in Assessment and Digital Learning

Middlesex University London

UTS UNIVERSITY OF TECHNOLOGY, SYDNEY

Overview

- The problem
- What stays the same in assessment?
- What needs to change?
- Guiding principles for progress
- What should assessment particularly emphasise now?



The problem

- Artificial Intelligence (AI) is permeating the whole of society
- Higher education is facing a dual challenge: to pedagogy and to assessment
- GenAI is a useful aid to learning even in its present early forms
- But it also provides great temptations to those who just want to avoid learning

The immediate AI challenge: generative AI

- Based on a Large Language Model, genAI produces seductive text that *appears* to answer any question posed to it
- It has current limitations and biases. Some will be fixed, some are intrinsic
- Versions of it are freely available to all (but equitably?)
- Students can use it for getting comments on their work at any time and generate ideas and text for assignments
- Students can use it to ‘augment’ their own work
- If a student uses it in assessment we can’t know if they can meet the appropriate learning outcome—it is a challenge to assessment validity (as each output is unique, detection is problematic)

What does assessment now need to do? Three key purposes

Assure	Enable	Build
Assure that learning outcomes have been met	Enable students to use information to aid their learning now	Build students' capacity to judge their own learning
• Summative assessment	• Formative assessment	• Sustainable assessment

When is each needed?

Assure	Enable	Build
Assure that learning outcomes have been met *Late in a course, once a period of learning is complete*	Enable students to use information to aid their learning now *Early in a course to help students reach the level of performance needed to meet the learning outcomes*	Build students' capacity to judge their own learning *Throughout a course to continually develop students' evaluative judgement*

2. Assessment for assurance

Assure

Assure that learning outcomes have been met

- Summative assessment

What are our fundamental assessment obligations?

"***Assessment ensures that the qualifications are awarded only to those students who meet specified learning outcomes.*** Learning outcomes are specified for each course, which are consistent with the relevant national qualification frameworks' descriptors, and assessment determines whether each student has achieved them"

"Providers operate processes that ensure learning outcomes are consistent with the requirements of the relevant national qualifications frameworks. ***They ensure assessments measure the extent to which students achieve the learning outcomes both at, and beyond, the threshold level***."

UK Quality Code for Higher Education. Advice and Guidance: Assessment. QAA November 2018
www.qaa.ac.uk/quality-code

A standards-based approach to assessment

- Judges whether students can demonstrate attainment of learning outcomes to a given standard.
  - Poor achievement on one outcome cannot be compensated for by good achievement on another
  - Judging students against each other (norm-referencing) is prohibited
- Requires transparent standards for each task
  - Setting a pass mark is not setting a standard. A threshold level is needed for each outcome
  - Universal Grade Point Averages are educationally meaningless

Implications for the context of genAI

Given that assessment must ensure that students have met specified learning outcomes at least at a threshold level, what might we do?

No generic answer, however, examples can illustrate the kinds of approaches that might be adopted

Example 1: Sequenced or nested tasks

Complex assessment tasks arranged in a sequence of sub-tasks allow for assessment for learning at early stages and assessment for assurance at the end of the sequence

- Discontinuity between products at different stages can indicate help derived illegitimately
- Selective random orals may be used for confirmation
- Knowing our students. Rethink anonymity for all tasks—when is anonymity really needed?

Example 2: Avoid tasks answerable by genAI

Check tasks by seeing what ChatGPT generates. Retain those it finds hard to cope with

- Avoid all standard problems/essay topics
- Make students accountable for their responses by always being prepared to justify and articulate their answers
- Set highly contextualised, personalised and recent tasks which require close attention to contemporary information

As genAI improves the latter may become less and less appropriate!

Example 3: Programmatic assessment

Stop *assuring* unit or module learning outcomes and move to assurance of program outcomes

- Include fewer acts of assessment for assurance that need to be guarded
- Include more data points, most of which are low-stakes
- Tools of curriculum mapping used to involve all those who contribute to a program—they make suggestions for, and commit to, alignment
- Identify in which course unit each program learning outcome is introduced, developed or assured?

Key takeaways for assessment for assurance

- Invest energy into integrity for assurance only, not for every act of assessment
- Don’t penalise students for early failures to meet standards, eg. through GPAs. Ensure that what they end up being able to do is formally recorded, not just my grades
- For assurance, additional checks may be needed. For example, oral confirmation (but not orals generally). Triggered without a *prima facie* case of misbehaviour being established.

Pause for discussion and questions

3. Assessment to enable learning

Enable

Enable students to use information to aid their learning now

- Formative assessment

3. Assessment to enable learning

- What needs to sit alongside acts of assessment to aid learning—feedback processes?

Key points about feedback

- Feedback is one of very few ways courses can be tailored to the individual needs of students
- Feedback processes need to be carefully designed
  - Providing comments to students is *only a part* of any feedback process
  - Without active engagement from students (eliciting/processing/acting), feedback hasn't occurred because it can't influence learning
- Feedback must *always* be judged in terms of its effect on learning
- Students get inputs from many sources (staff, peers, family, LMS, genAI)

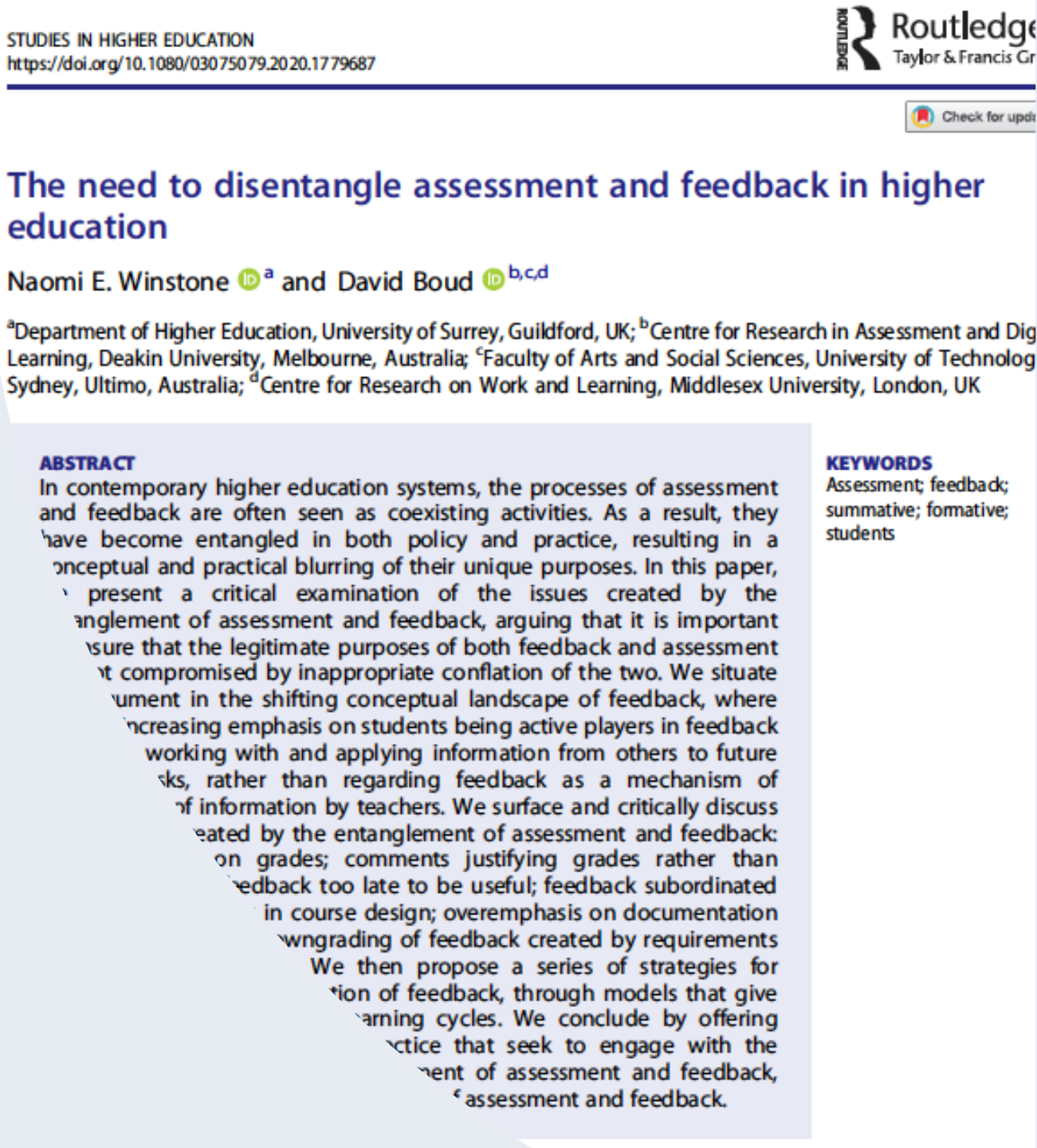
An important distinction

Mark justification

- Judgements and comments about what students have completed
- Essentially backward-looking

Feedback information

- Comments about what students can do to improve their work
- Essentially forward-looking



STUDIES IN HIGHER EDUCATION
https://doi.org/10.1080/03075079.2020.1779687

Routledge
Taylor & Francis Gr

Check for upd

The need to disentangle assessment and feedback in higher education

Naomi E. Winstone [a] and David Boud [b,c,d]

[a]Department of Higher Education, University of Surrey, Guildford, UK; [b]Centre for Research in Assessment and Dig
Learning, Deakin University, Melbourne, Australia; [c]Faculty of Arts and Social Sciences, University of Technolog
Sydney, Ultimo, Australia; [d]Centre for Research on Work and Learning, Middlesex University, London, UK

ABSTRACT

In contemporary higher education systems, the processes of assessment and feedback are often seen as coexisting activities. As a result, they ave become entangled in both policy and practice, resulting in a nceptual and practical blurring of their unique purposes. In this paper, present a critical examination of the issues created by the anglement of assessment and feedback, arguing that it is important sure that the legitimate purposes of both feedback and assessment t compromised by inappropriate conflation of the two. We situate ument in the shifting conceptual landscape of feedback, where ncreasing emphasis on students being active players in feedback working with and applying information from others to future ks, rather than regarding feedback as a mechanism of f information by teachers. We surface and critically discuss eated by the entanglement of assessment and feedback: on grades; comments justifying grades rather than edback too late to be useful; feedback subordinated in course design; overemphasis on documentation wngrading of feedback created by requirements We then propose a series of strategies for tion of feedback, through models that give arning cycles. We conclude by offering ctice that seek to engage with the ent of assessment and feedback, assessment and feedback.

KEYWORDS

Assessment; feedback; summative; formative; students

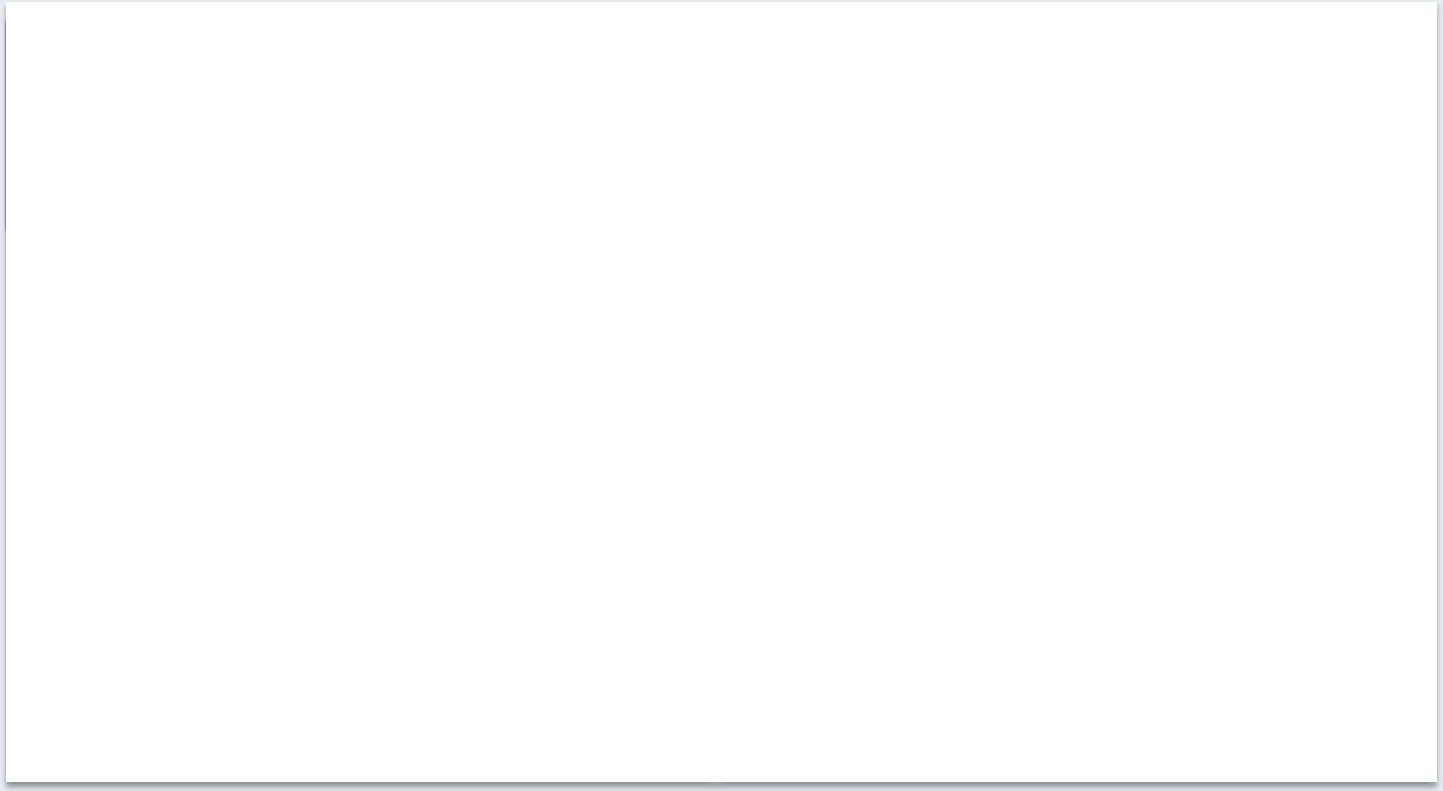
Disentangle grading from feedback

- Not all student tasks can or should be marked
  - at least in a way that leads to a permanent record on students' files or generates a GPA
- Feedback can occur with assessed or non-assessed work
  - They are not synonymous nor need to occur together
- Not all assessed work needs to be linked to a feedback process
  - eg. end of course products
- Feedback is needed when students can do something about the information they receive, not when they can't

Key takeaways for assessment for learning

- Help students know that engaging in assessment for learning is essential to enable them to meet assurance standards
  - persuade using evidence from previous cohorts of their own course
- If they cheat on assessment *for learning* they *only* cheat themselves and will disadvantage themselves later
- Talk always in terms of meeting standards, not grades
- Students must learn to actively use feedback processes to benefit themselves.

4. Assessment to build capacity for judgement



Build

Build students' capacity to judge their own learning

- Sustainable assessment

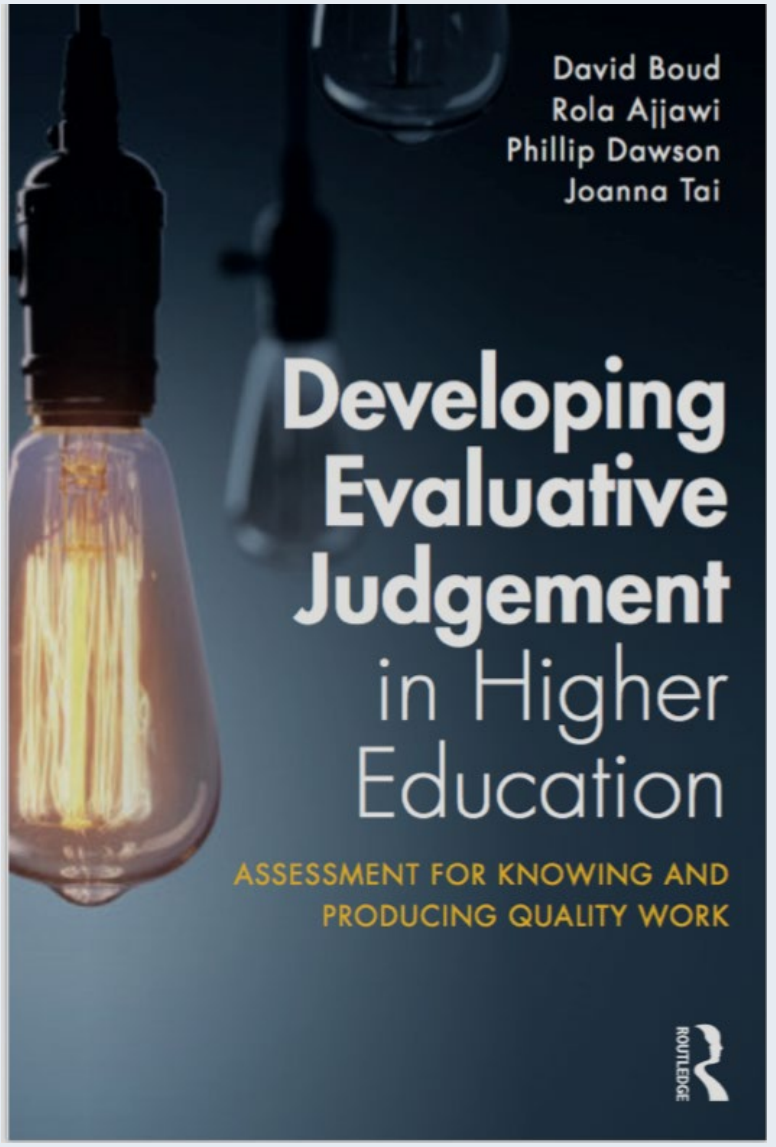
Developing evaluative judgement

- If students can’t judge the quality of their own work, how can they learn effectively?
- If graduates can’t judge the quality of their own work, how can they practice effectively?
- If students or graduates can’t help each other judge the quality of their work, how can they work effectively with each other?

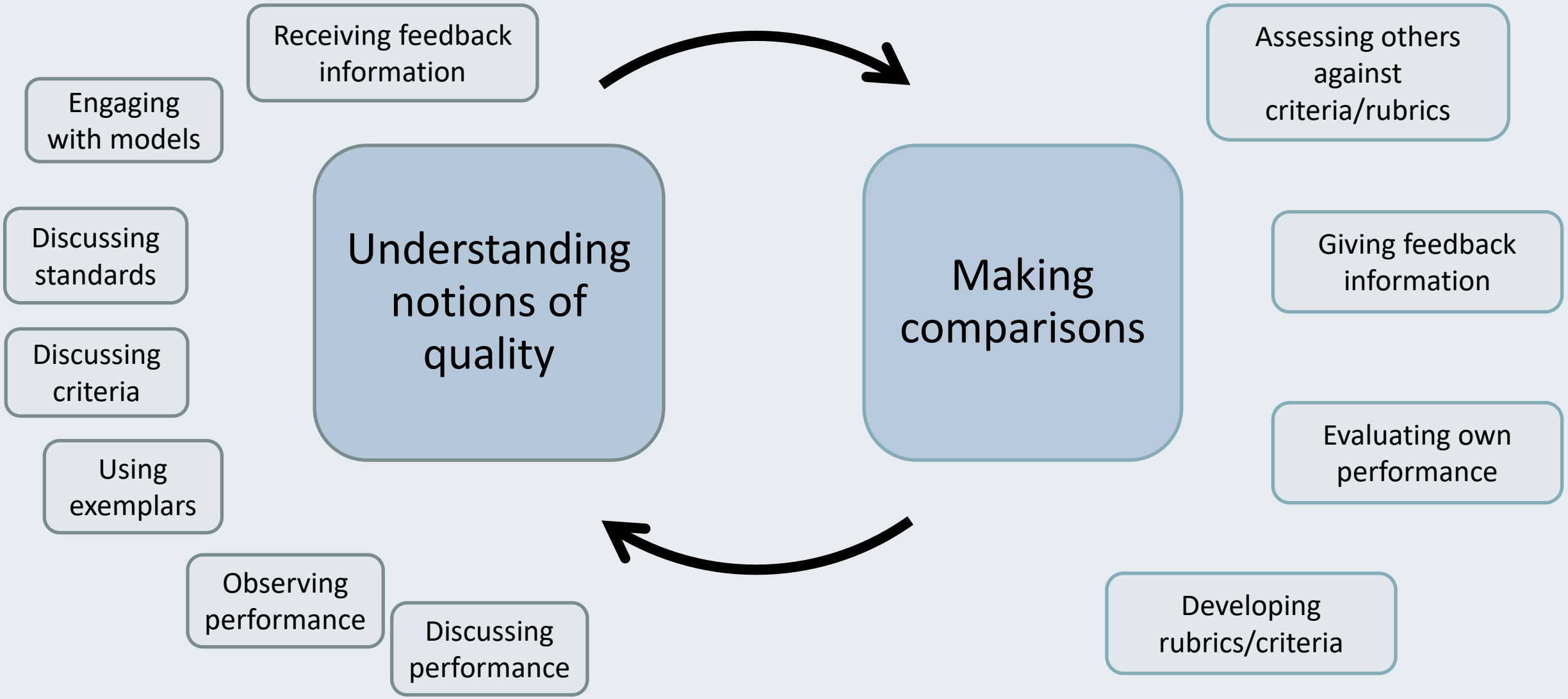
“If you can't describe what you are doing as a process, you don't know what you're doing.”
— *W. Edwards Deming 1900-1993*

Evaluative judgement is:

the capability to make
informed decisions
about the quality of work
of self and others



Integral components of evaluative judgement



Key takeaways for assessment to build judgement

- Focus students on the question of ‘what does good look like’ at every opportunity
- Provide multiple opportunities for students to judge their work and that of others as part of all courses
- Build students evaluative judgement to use AI tools effectively (ie. develop prompting/judging skills, and the importance of checking and iteration)

Balancing different purposes of assessment

- *Assessment for assurance*
  needed after a period of learning is complete, assessment before then does not reflect what students can do
- *Assessment to enable learning*
  needed during a period of learning—at the end it is too late
- *Assessment to build students capacity to make effective judgements*
  needed throughout

Where does this leave us now—assessment?

- Regression to traditional exams handicaps students and undermines standards
  - important learning outcomes will not be addressed
  - Assessment will not equip students for professional practice
- Major redesign of assessment needed to
  - avoid unrealistic tasks or ones that can be outsourced to available technology
  - have multiple occasions to assess each program learning outcome
  - have fewer, more complex tasks which relate to threshold standards
  - have more specific, unique, current and contextualised tasks
- Securely protect a few key tasks
  - more assessment for learning, less but higher quality risk-managed assessments for assurance
  - Assure the program not each module!

Pause for discussion and questions

What is happening more widely?

- Proliferation of Institutional guidelines of how to cope with AI and assessment
  - Too often they are simplistic, out of date and apply to only some versions of genAI
- genAI platforms and versions are coming to market every month
  - Better versions often involve payment, so how do we deal with issues of equity?
- Starting to see more principle-based approaches
  - They don’t tell you how to do it, but provide principles that need to be met
- Research is starting to be published on effects of different strategies and the effect of the same strategy on different students
  - Disturbingly, the same strategy may have a very positive effect on stronger students and a very negative effect on weaker students

Resources

Useful webpage of Australian accrediting/quality agency (TEQSA) about assessment and AI:
https://www.teqsa.gov.au/guides-resources/higher-education-good-practice-hub/artificial-intelligence?utm_source=sendgrid.com&utm_medium=email&utm_campaign=website

Bearman, M., Ajjawi, R., Boud, D., Tai, J. & Dawson, P. (2023). *CRADLE Suggests… assessment and genAI*. Centre for Research in Assessment and Digital Learning, Deakin University, Melbourne, Australia. doi:10.6084/ m9.figshare.22494178

Check out my colleague Phill Dawson on cheating and threats to integrity in using AI on YouTube

https://blogs.deakin.edu.au/cradle/wp-content/uploads/sites/188/2023/06/CRADLE-Suggests-Assessment-and-genAI.pdf

In a world of genAI:

Enact principles of good assessment design

- Deploy resources to assure assessment when it matters most
- Design feedback sequences to support learning
- Develop student capability to identify 'what good looks like' through assessment
- Devise multiple submission formats to make assessment more equitable
- Focus on evidencing that outcomes are met

Adapt current assessment to account for genAI

- Have open conversations about genAI with students
- Review rubrics and other forms of assessment criteria
- Specify assessment situations where it is appropriate or inappropriate to use genAI
- Design tasks to promote students' portrayal of their unique achievements
- Develop and assess critical digital literacies

A national response

- In Australia, TEQSA is the regulator that accredits entire universities and guards the statutory Higher Education Quality Standards
- TEQSA decided on a light touch (initially) and commissioned a series of online seminars conducted by CRADLE at Deakin University. Enrolment for each 2,000+
- It then sponsored a group of experts to develop in consultation with other experts a document *Assessment Reform for the Age of Artificial Intelligence*. A set of principles, not a how to do it guide.
- By mid-2024 each institution accredited by TEQSA will have to say how they are dealing with the issue.

We need guiding principles to consider impact on assessment, not ad hoc responses

Assessment Reform for an Age of Artificial Intelligence

Guiding principles

I. Assessment and learning experiences equip students to participate ethically and actively in a society where AI is ubiquitous

II. Forming trustworthy judgements about student learning in a time of AI requires multiple, inclusive and contextualised approaches to assessment

https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf

Assessment Reform for an Age of Artificial Intelligence

Propositions

Assessment should emphasise...

1. ... appropriate, authentic engagement with AI
2. ... a systemic approach to program assessment aligned with disciplines/qualifications
3. ... the process of learning
4. ...opportunities for students to work appropriately with each other and AI
5. ... security at meaningful points across a program to inform decisions about progression and completion

https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf

Assessment Reform for an Age of Artificial Intelligence

Propositions

Assessment should emphasise...

1. ... appropriate, authentic engagement with AI

AI is ubiquitous in society, denying it in university courses is not only unrealistic, but dysfunctional for students' lives

Assessment Reform for an Age of Artificial Intelligence

Propositions

Assessment should emphasise...

2. ... a systemic approach to program assessment aligned with disciplines/qualifications

A move to programmatic or coursewide assessment is needed as the level of protection of integrity for assessment in all units is not achievable. This implies less summative assessments and relatively more formative assessments

Assessment Reform for an Age of Artificial Intelligence

Propositions

Assessment should emphasise...

3. ... the process of learning

A focus on how students approach tasks rather than simply on a final result is needed, as it is this that transfers to new situations

Assessment Reform for an Age of Artificial Intelligence

Propositions

Assessment should emphasise...

4. ...opportunities for students to work appropriately with each other and AI

Students need to be enabled to develop skills of working with AI, eg prompt engineering, as well as with each other to future life.

Assessment Reform for an Age of Artificial Intelligence

Propositions

Assessment should emphasise...

5. ... security at meaningful points across a program to inform decisions about progression and completion

Maintaining assessment security is an expensive process and done naively threatens to diminish the quality of student achivement. What are the appropriate points?

Dilemmas to be addressed

- Many existing problems with assessment have still to be addressed
  - Assessment is still enacted in ways that don't address learning outcomes or meet standards
- Don't leave it to each staff member or adopt lock-down approaches to address the problem
  - Even with good guidelines, it is too costly, generates resistance and leads to poor outcomes
- Different students will respond differently to the same conditions
  - Strong students will utilise AI for learning, weak students to use it to stagger over the line

Further discussion