# Assessment design for a time of AI

Phillip Dawson

Centre for Research in Assessment and Digital Learning (CRADLE)

Deakin University



How to draw an owl

1.

2.

1. Draw some circles     2. Draw the rest of the f***ing owl

# Disclaimer

- I'm a 'standards-based assessment' person

- I support AfL

- I think there are many interesting things to talk about with AI

- I receive research funding from ed tech companies
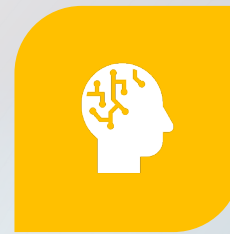
- My mum helped me contract cheat in year four
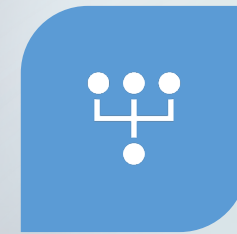
# Four things to take from this presentation

The need for assessment change for a time of AI

How assessment needs to change for a time of AI

Key considerations to keep in mind as you change assessment

The necessity for structural assessment changes

If AI can do your assessment to the pass level, and your task matters for assurance of learning, there might be a problem

- Now, and in the future, with a good prompt and iterative refinement
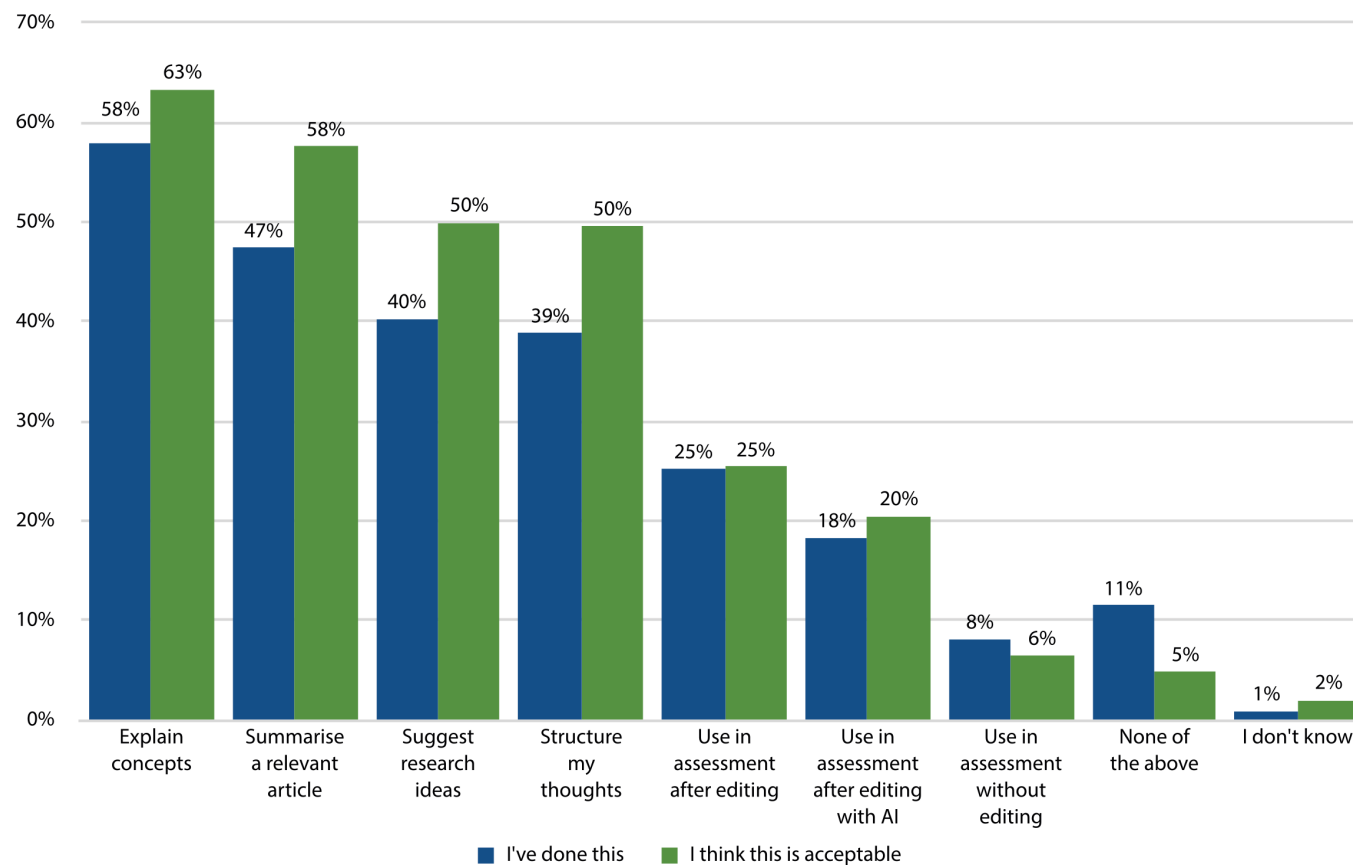- Most of the time when someone thinks it can't, and they ask me, it can

"AI can't do xyz…"

Here's where we were in 2021

# A Brief History of Artificial Intelligence

Michael Wooldridge

What It Is, Where We Are, and Where We Are Going

arithmetic (1945)

sorting lists of numbers (1959)
} easy

playing simple board games (1959)

playing chess (1997)

recognizing faces in pictures (2008)

usable automated translation (2010)

playing Go (2016)

usable real-time translation of spoken words (2016)
} solved, after a lot of effort

driverless cars

automatically providing captions for pictures
} real progress

understanding a story & answering questions about it

human-level automated translation

interpreting what is going on in a photograph

writing interesting stories

interpreting a work of art

human-level general intelligence
} nowhere near solved

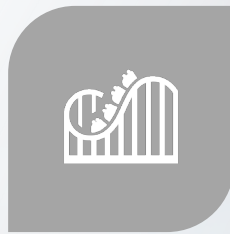*Figure 5 What students have done themselves versus what they consider acceptable*



Legend: ■ I've done this   ■ I think this is acceptable

| Category | I've done this | I think this is acceptable |
|---|---|---|
| Explain concepts | 58% | 63% |
| Summarise a relevant article | 47% | 58% |
| Suggest research ideas | 40% | 50% |
| Structure my thoughts | 39% | 50% |
| Use in assessment after editing | 25% | 25% |
| Use in assessment after editing with AI | 18% | 20% |
| Use in assessment without editing | 8% | 6% |
| None of the above | 11% | 5% |
| I don't know | 1% | 2% |

## Quiz: Collective Learning

⊞ Google Classroom

**What is collective learning?**

☑

Choose 1 answer:

(A) The ability of living organisms to change from generation to generation

(B) A form of communication that uses symbols

(C) The act of relying on nature for sustenance; hunting and gathering

(D) The ability to share, preserve, and build upon ideas over time

Start over **Do 12 problems** ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Check



QuizSolverAI: https://quizsolverai.com/

Operator Video credit: Cursive on YouTube
https://www.youtube.com/watch?v=X887F7NeCR4

# Four things to take from this presentation



The need for assessment change for a time of AI

How assessment needs to change for a time of AI
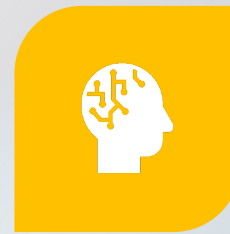
Key considerations to keep in mind as you change assessment
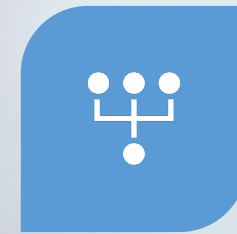
The necessity for structural assessment changes

**Lodge, Howard, Bearman, Dawson & Associates (2023)**

2 principles
5 propositions

Developed with 18 experts in assessment, higher education and artificial intelligence

Lodge, J. M., Howard, S., Bearman, M., Dawson, P, & Associates (2023). *Assessment reform for the age of Artificial Intelligence*. Tertiary Education Quality and Standards Agency.

Australian Government
Tertiary Education Quality and Standards Agency

## Assessment reform for the age of artificial intelligence

November 2023

TEQSA

## bit.ly/teqsaAI

(or Google TEQSA and the document name)

cradle

DEAKIN UNIVERSITY

# Two guiding principles

## Guiding Principle 1

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

## Guiding Principle 2

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

# Five 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*

# Addendum: three thoughts since talking with people about the document

We don't mention cheating, but it's not a useful framing

We also don't mention validity, but it underpins the whole thing

We can't substitute all our assessments of disciplinary outcomes with assessments of AI use/critique outcomes

# Four things to take from this presentation

The need for assessment change for a time of AI

How assessment needs to change for a time of AI
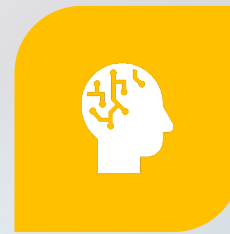
Key considerations to keep in mind as you change assessment
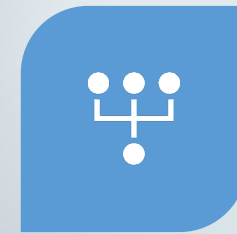
The necessity for structural assessment changes

*"many of the dilemmas we face are not about assessment per se but are at heart 'debates about what should be assessed'."*
*(McDowell, 2010, citing Wiliam 2007)*

McDowell, L. 2010. "Challenging Assessment?" Assessment & Evaluation in Higher Education 35 (3): 263–264.
Wiliam, D. 2007. "Balancing Dilemmas: Traditional Theories and New Applications." In Balancing Dilemmas in
Assessment and Learning in Contemporary Education (pp. 267–281). Routledge.
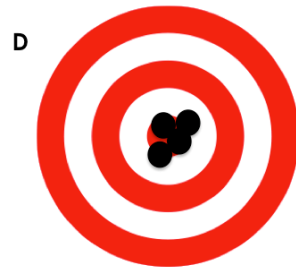
# Validity matters more than cheating
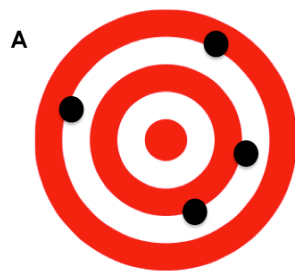


A — Unreliable & invalid
B — Unreliable but valid?
C — Reliable but invalid
D — Reliable & valid
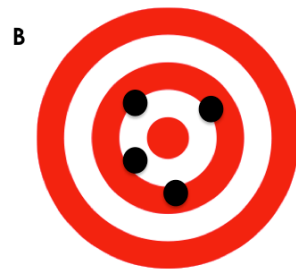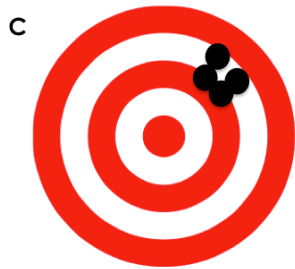
From a validity perspective, does it matter if a student uses AI in this task?

## Validity matters more than cheating

Phillip Dawson[a], Margaret Bearman[a], Mollie Dollinger[a,b] and David Boud[a,c,d]

[a]Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, Australia; [b]Faculty of Health Sciences, Curtin University, Perth, Australia; [c]Faculty of Arts and Social Sciences, University of Technology Sydney, Sydney, Australia; [d]Work and Learning Research Centre, Middlesex University, London, UK

Image: https://learningspy.co.uk/assessment/when-assessment-fails/#post/0

# Future-authentic assessment

"assessment that faithfully represents not just the current realities of the discipline in practice, but the likely future realities of that discipline"
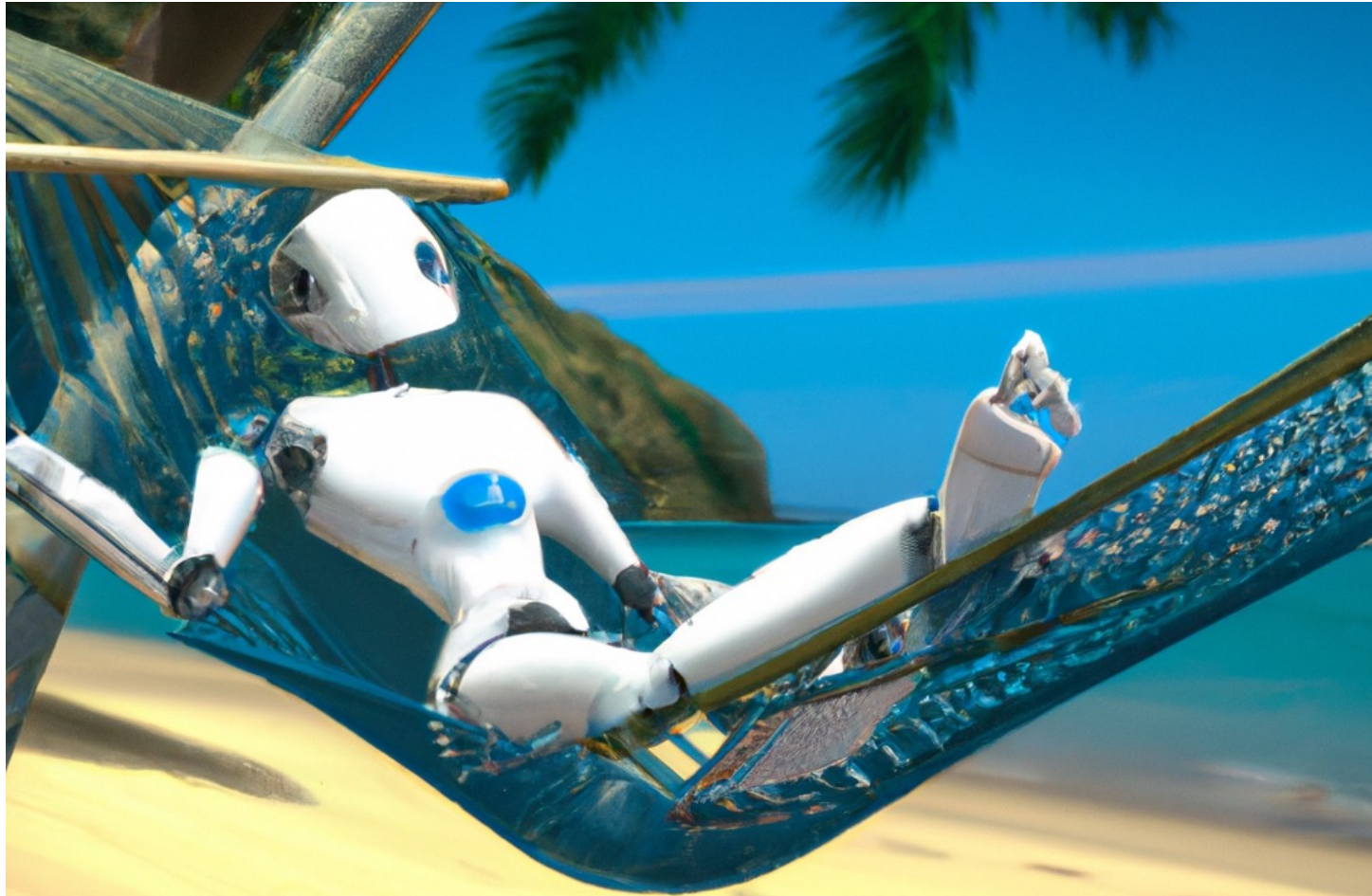
# (Reverse) Scaffolding

**Scaffolding**: allow use of genAI to support students for outcomes they are developing

**Reverse scaffolding**: allow use of genAI for outcomes students have sufficiently mastered

AI tools for production don't respect the Zone of Proximal Development

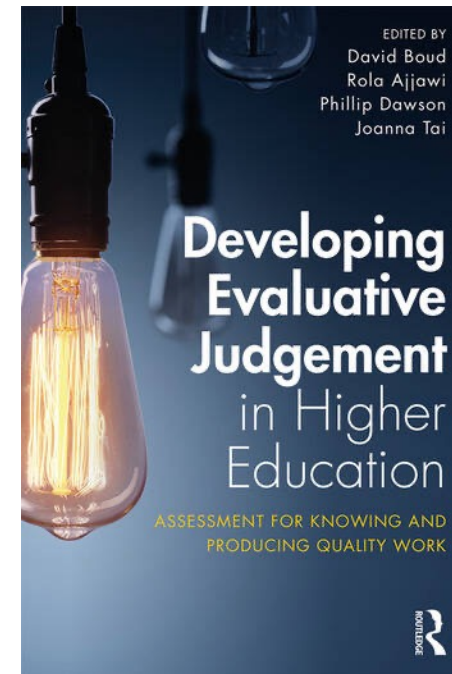Cognitive offloading

# Evaluative judgement



Judgements about AI products

Judgements of AI processes

AI judging our work

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 1-13. https://doi.org/10.1080/02602938.2024.2335321

# Those 7 concepts again

**Assessment matters, but so does what is assessed** (does it need to change in a time of AI?)

**Validity matters more than cheating** (is AI panic more of a threat to validity than AI itself?)

**Future-authentic assessment** (prepare for their future, not our past)

**Reverse scaffolding** (use AI once you can do it yourself)

**Zone of Proximal Development** (tools for production vs tools for learning)
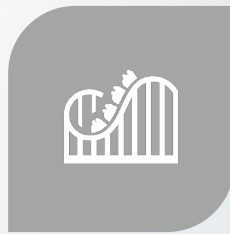
**Cognitive offloading** (extraneous vs intrinsic)

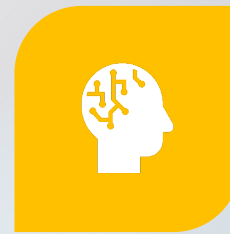**Evaluative judgement** (but it can't be the only thing)

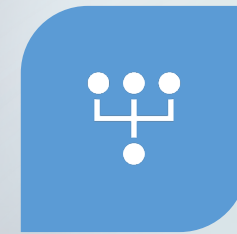# Four things to take from this presentation

The need for assessment change for a time of AI

How assessment needs to change for a time of AI

Key considerations to keep in mind as you change assessment

The necessity for structural assessment changes

# Don't just tell students what to do

**Routledge**
Taylor & Francis Group

OPEN ACCESS · Check for updates

## 'Where's the line? It's an absurd line': towards a framework for acceptable uses of AI in assessment

Thomas Corbin[a] (iD), Phillip Dawson[b] (iD), Kelli Nicola-Richmond[b] (iD) and Helen Partridge[b] (iD)

[a]Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Australia; [b]Deakin University, Australia;

**Routledge**
Taylor & Francis Group

OPEN ACCESS · Check for updates

## Talk is cheap: why structural assessment changes are needed for a time of GenAI

Thomas Corbin[a] (iD), Phillip Dawson[a] (iD) and Danny Liu[b] (iD)

[a]Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Melbourne, Australia; [b]DVC Education Portfolio, The University of Sydney, Sydney, Australia

Discursive: addressing AI in assessment through the instructions we give to students, e.g. "you can use AI for editing but not writing"

Structural: addressing AI in assessment through changes in the task that are unavoidable, e.g. "we will have a conversation with you about your work"

If you aren't supervising, you can't be sure how AI was/wasn't used

So don't set restrictions that can't be enforced
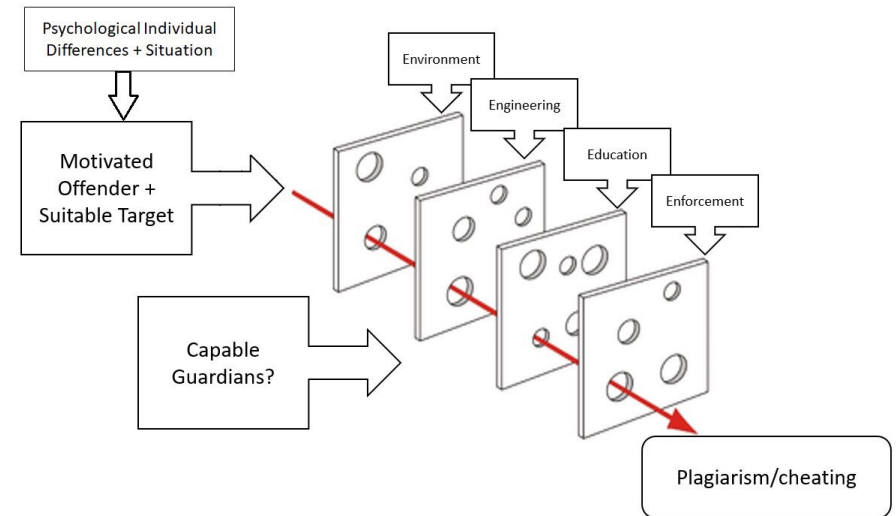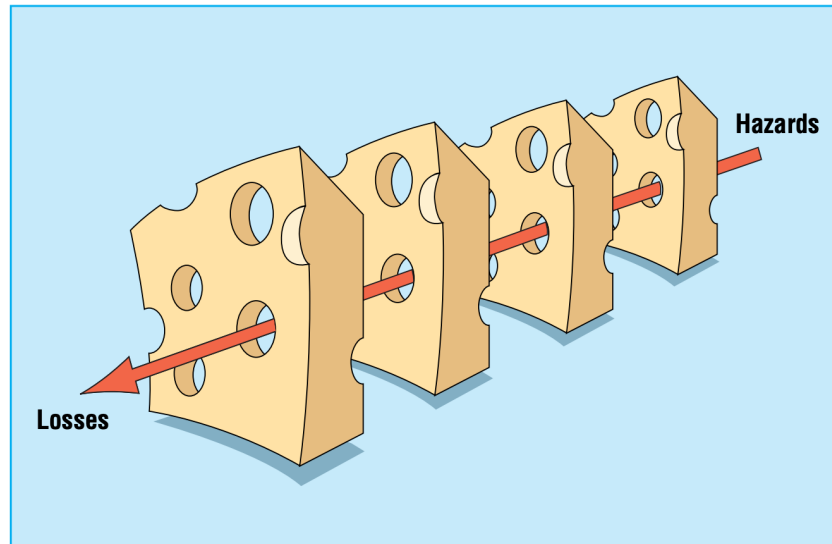
It just hurts validity

Secure some tasks

Accept the others will see significant AI use

There is no such thing as an AI-proof assessment.

Where it matters that students don't use AI, we need multiple approaches.

Reason, J. (2000). Human error: models and management. BMJ, 320(7237), 768. https://doi.org/10.1136/bmj.320.7237.768

Rundle, K., Curtis, G., & Clare, J. (2020). Why students choose not to cheat. In T. Bretag (Ed.), *A Research Agenda for Academic Integrity. Edward Elgar Publishing.*
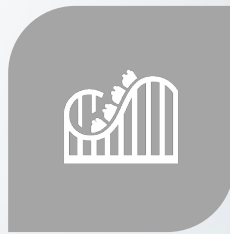
# 10 key concepts

1.  **Assessment matters, but so does what is assessed** (does it need to change in a time of AI?)

2.  **Validity matters more than cheating** (is AI panic more of a threat to validity than AI itself?)

3.  **Future-authentic assessment** (prepare for their future, not our past)

4.  **Reverse scaffolding** (use AI once you can do it yourself)

5.  **Zone of Proximal Development** (tools for production vs tools for learning)

6.  **Cognitive offloading** (extraneous vs intrinsic)

7.  **Evaluative judgement** (but it can't be the only thing)

8.  **Make structural not discursive changes** (no bogus rules)

9.  **No such thing as AI-proof assessment** (beware of anybody who says they have one)

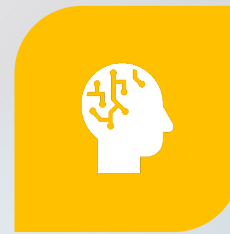10. **Swiss Cheese, programmatic** (layers of imperfect assessments tell us more than one good one)

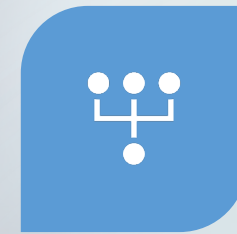# Four things to take from this presentation

**The need for assessment change for a time of AI**

**How assessment needs to change for a time of AI**

**Key considerations to keep in mind as you change assessment**

**The necessity for structural assessment changes**