

Authentic assessment

Singapore Seminar on Authentic assessment (Part 2)
Prof. Gavin T. L. Brown, The University of Auckland



**EDUCATION AND
SOCIAL WORK**

1



**EDUCATION AND
SOCIAL WORK**

Educational Authentic Assessment



2

Assuring education in authentic assessment



EDUCATION AND
SOCIAL WORK

- | | |
|---|---|
| <ul style="list-style-type: none"> ▶ Learners; Not yet competent ▶ Purpose: prepare generically for beginner role ▶ Design: Analytic, sequenced, scaffolded, selected & controlled | <ul style="list-style-type: none"> ▶ Professionals; competent ▶ Purpose: achieve a specific work-related goal ▶ Design: Integrated, dynamic, simultaneous, unpredictable |
|---|---|



Learners who can fail
Instructors who support



Workers who must succeed
Informal instruction?

Remember we are teaching learners, not judging already completed professionals....

Educa

Real World

3

Problems in making things authentic



EDUCATION AND
SOCIAL WORK

- ▶ School is a simulation of real life
 - doing of the task does not have 'real' consequences, that is where the person being assessed is not fully accountable for the outcomes.
- ▶ No matter how realistic a performance-based assessment is, it is still a simulation, and examinees do not behave in the same way they would in real life.
 - Cumming, J. J., & Maxwell, G. S. (1999). Contextualising authentic assessment. *Assessment in Education: Principles, Policies and Practices*, 6(2), 177–194. <https://doi.org/10.1080/09695949992865>

4

Authenticity Exhibits Learning



EDUCATION AND
SOCIAL WORK

- ▶ A mental model that guides the program and all participants in deciding what to assess

- A model I am involved in
- *Assessment = Demonstration*
- *Show me what you understand, know, & can do with it to meet learning goals*



The 4Es Our findings Workshop materials News and views More information...

The 4Es

Developing employability capabilities of students requires transformational practices that move beyond teaching students to 'know more' to 'be more'. The 4Es of Employability model aims to develop understanding of transformational teaching practices that others can replicate. Over the course of this project, exemplar cases at the Programme, Course and Session levels will be selected for each 'E'.

Download full summary > <http://www.futurereadygrads.ac.nz/the-4es/>

Excite

Excite your learners.

Students will learn if they are motivated by the prospect of a course, find learning relevant, and feel they will be successful.

Explore

Have your learners explore new ideas.

Students are more likely to learn if they are given opportunities to thoroughly explore new ideas in terms of their current undertaking.

Extend

Challenge learners' thinking.

Students will develop a deep understanding and be better prepared for transfer if they have opportunities to test, apply and use and challenge new ideas in multiple authentic, relevant and meaningful contexts

Exhibit

Ask students to show the world what they can do.

Students will gain most from assessment that involve building authentic artifacts in realistic conditions.

5

Designing for Feedback



EDUCATION AND
SOCIAL WORK

- ▶ Every task we set is an opportunity to do 2 things
 - Judge quality, rank, report grades (Accountability)
 - Assess or diagnose success and needs, and prescribe solutions (Improvement)
- ▶ Feedback opportunities have to be designed into the sequence of assessed tasks
 - Task 1 must be simpler than following tasks
 - Task 1 must be essential for later tasks, so feedback is relevant
 - Task 1 diagnostic information must help improve Task 2 performance
 - In terms of knowledge, understanding, skills

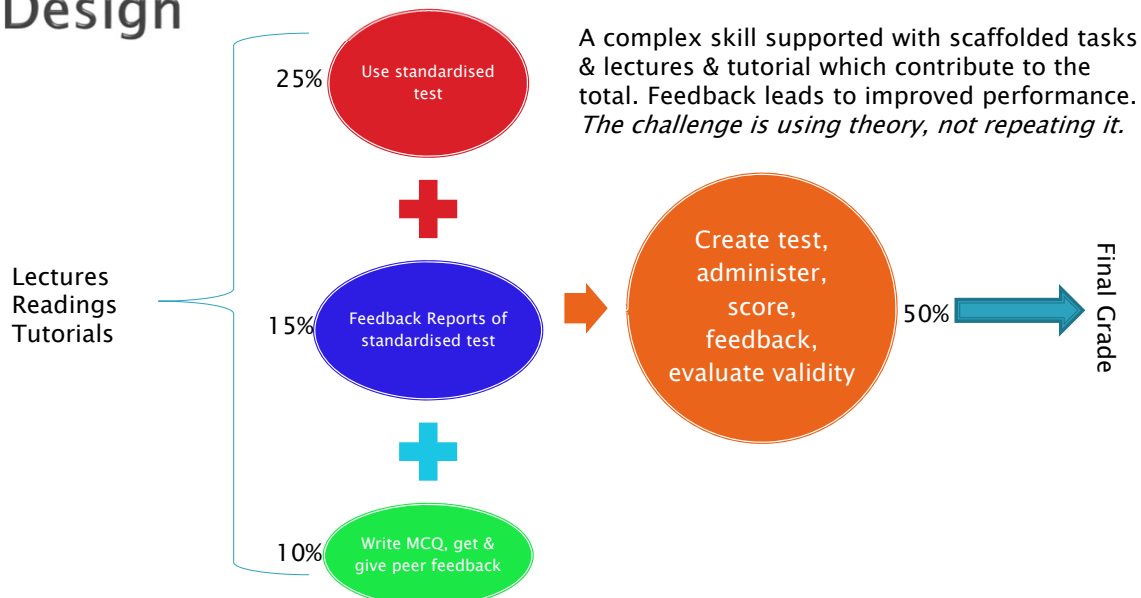
6

Authentic assessment exhibits integrated knowledge & skills

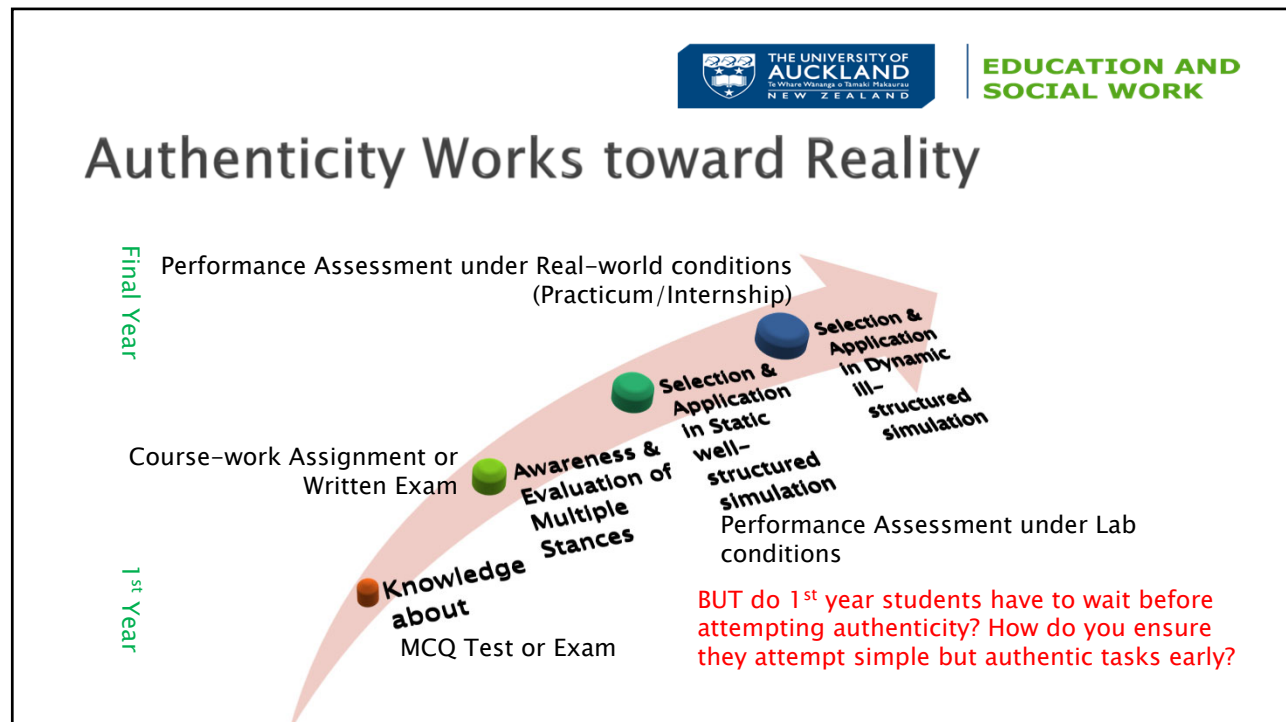
- ▶ I already know you can answer MCQ or write an essay....
- ▶ How can I get you to apply the that knowledge and skill in a novel way, so that I can see your ability to integrate and synchronise multiple objects, processes, and ideas simultaneously?
 - Because this is the way of the world...it's authentic.
- ▶ Example: My own course Educ 224 Assessment & Evaluation in Education within the BA degree

7


Educ 224 Assessment Design



8




9


EDUCATION AND SOCIAL WORK

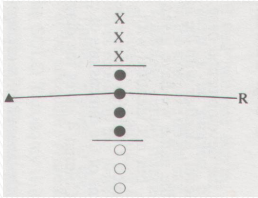
Authenticity not just difficulty

- ▶ Real learning occurs when learners do things that are achievable but challenging
 - Higher-order thinking skills,
 - *do more complex; not just "more in less time"*
- ▶ Cognitive challenge not just long and hard
- ▶ How can we ensure that we build challenge into all aspects of curriculum, course content, and assessment?
 - Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. New York: Academic Press.

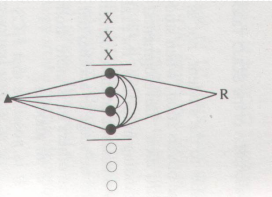
10


**EDUCATION AND
SOCIAL WORK**

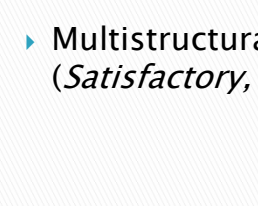
SOLO: Structure of Observed Learning Outcomes



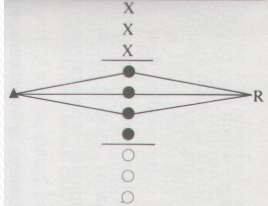
Unistructural (*Fail*)



Relational (*Good, B*)



Multistructural (*Satisfactory, C*)




Extended Abstract (*Excellent, A*)

Surface

Deep

11


**EDUCATION AND
SOCIAL WORK**

Assessing Deep Outcomes

- ▶ Ensure assessments require more than recall and recognition
- ▶ Tasks require analysis, synthesis, evaluation, insight, creativity
 - Understand relationships in material,
 - Transform material from one form to another,
 - Explain meaning of things for oneself,
 - Derive abstract principles from ideas or information,
 - Apply theory or knowledge to a different context,
 - Analyse and propose solution to a 'knotty' or 'wicked' problem,
 - Describe how student is developing or changing as an individual,
 - Create a novel angle or perception of things,
 - Integrate material with that from other disciplines...etc.

12

Table Talk

- ▶ *How formative and educational are your assessments?*
- ▶ Discuss with people in a zoom chat room ...
 - To what extent do your assessment practices ensure students gain meaningful educational benefit?
 - No support ☐ ☐ ☐ ☐ ☐ ☐ Fully supportive
 - What changes might you have to implement to ensure that the assessments are constructive to learning?

13

Reliability in Authentic Assessment



14

Quality Judgment



EDUCATION AND
SOCIAL WORK



Real world is not usually a score; rather a quality judgment
How well was this done?
accuracy, speed, sufficiency, relevance

15

Quality judgment



EDUCATION AND
SOCIAL WORK

- ▶ We can't give up making evaluative or diagnostic judgments about valued knowledge and skills because that is how it is done in the real world
 - It's not a quantity but a quality that we are meant to reward
 - How well, not how much
- ▶ Challenge of proving that our judgments are valid, trustworthy
 - If we say a person is excellent, would anyone in our institution or in the world outside agree?

16

Design for Learning, not just judgement

► Discuss

- How (or what) would you restructure the sequence and nature of assessments in your program so that students:
 - Are supported in developing component skills
 - Demonstrate that they can integrate knowledge with competencies
 - Receive feedback in such a way that performance provides insights into improvement
 - Receive sequences of assessments within courses that support developmental acquisition of knowledge and skills
- What systems do you want to ensure your program supports learning?

17

Ensuring valid consistent judgments



18

How do you score?



**EDUCATION AND
SOCIAL WORK**

- ▶ What criteria will be used to judge the performance?
 - Determine the Rich Ideas that your scoring, teaching, and reporting need to reflect.
- ▶ Are the criteria norm or criterion referenced?
 - Can you get past incremental quantity differences to a difference in quality?
- ▶ What performance/criteria fits each level / grade?

19

What is acceptable evidence in a performance?



**EDUCATION AND
SOCIAL WORK**

- ▶ Determine the amount and quality of evidence needed to determine the progress interpretations you want to make

20

Assuring Judgments

- ▶ Ideally the same work should be given the same score or grade no matter who marks it
 - But agreement & correlations between and within markers are poor.
 - Differences for the same essay (max 20-point) ranged from zero to 33%
 - Differences varied between 20% and 35%.
 - Scores (max. 25) for 10 end-of-year examination essays marked by 10 different tutors differed by 5 to 14 marks ($M=8.40$ or 34% of maximum score).
 - not one of 11 duplicated science practical reports and three copied social science essays was given the same mark.

21

Even panels of experts won't agree

Olympic scoring

- Range: 8.80 to 9.30—
 - actually pretty close, but these are the elite performers
- Only 3 times do 2 judges give the same mark.
- Remember how quickly this performance happens
- This is hard!

	Judges					
Athlete	A	B	C	D	E	F
1	9.30	9.10	9.10	8.90	9.00	8.90
2	9.00	9.30	9.00	8.80	9.00	9.10

22

Olympic Judging

- ▶ Highly skilled PANEL of judges who systematically observe multiple performances by elite athletes in-the-moment of performance
- ▶ Holistic rating on a multi-point scale
- ▶ Rules for deducting points for significant faults
- ▶ Bias reduction by removing highest and lowest judges per judgement
- ▶ Cumulative sum discriminates between performances
- ▶ Usually very reliable and valid

23

Evaluating Judgements

- ▶ Since error & inconsistency in marking is so commonplace we ought to check how we are doing
 - Between markers → moderation
- ▶ Cross-checking by having 2 qualified judges mark and compare scores for a common group of performances

24



EDUCATION AND
SOCIAL WORK

Evaluating Judgements

- ▶ If ratings are similar between judges or across times then we can trust the scores as reflecting consensus of an expert community
 - **Consensus**—the % of times the same score was awarded
 - Identical >70%
 - Approximate (within 1 mark) >90%
 - **Correlation**—tendency to give high and low marks to the same pieces of work
 - $r > .70$
- ▶ **So plan mechanism for checking degree of agreement**
 - **Not knowing is dangerous to our ability to persuade outsiders that we know what we are doing**

25



EDUCATION AND
SOCIAL WORK

Moderation & Training

- ▶ If you meet the expected targets you can use the scores defensibly to make decisions about learning needs and priorities and to report
- ▶ But if not.....
 - More training
 - Perhaps bring in a more experienced person (a Professor, the program manager?)
 - Perhaps include industry, profession people to comment on marking
 - Think about evaluating the work against the standard of “A beginner who would get a job today = A excellent”

26

Simple moderation



EDUCATION AND
SOCIAL WORK

- ▶ Are these close enough to use?
- ▶ What discussion is needed?

	Task		
marker	A	B	C
<u>1</u>	B+	C	D+
<u>2</u>	B	C+	C+
<u>3</u>	B+	B-	B-
Midpoint	B+	C+	C
Agreement			
Exact	0.67	0.00	0.00
Approximate	1.00	1.00	0.33

How far apart do we need to be before we stop trusting the marks?

27



EDUCATION AND
SOCIAL WORK

Table Talk

- ▶ *Moderating your scoring, rating, grading*
- ▶ Discuss with people in a zoom chat room ...
 - To what extent do you moderate or compare your grading with others?
 - Completely independent ☐ ☐ ☐ ☐ ☐ ☐ ☐ Fully Moderated
 - What changes might you have to implement to ensure that the assessments are reliable enough to use?

28