

Assessment in higher education: Challenges and solutions

Gavin T L Brown

Congresso Internacional sobre Avalicao no Ensino Superior,
University of Minho, Braga, Portugal, February 7–8, 2019



EDUCATION AND
SOCIAL WORK



EDUCATION AND
SOCIAL WORK

Challenges

- ▶ **Bologna Process & Assessment Reform Movements**
 - Student-centred; skill & ability, deep thinking;
 - Formative, feedback
- ▶ **Global Education Reform Movement**
 - Evaluating and ensuring quality of institutions through assessment
- ▶ **Workloads**
 - Large class sizes, much interactive feedback
- ▶ **Modern environments**
 - Students low attention & effort



Challenge: Traditional Approaches to Assessment

- ▶ Grand Traditions in Higher Education
 - Summative, terminal, end-of-course timing
 - How does student know if they have learned?
 - Multiple-choice questioning
 - Coverage but depth of understanding & application?
 - Essays
 - Depth but doubtful validity in marking, feedback comments
 - Marking (100% or A to E/F) and Ranking (1st to last)
 - No/little information on *what do I have to do next?*



Solution possibilities

- ▶ Assessment formats
 - Better item development
 - SOLO taxonomy to ensure depth
 - alternatives to written essays,
 - psychometric methods of eliminating bad items from tests or exams,
- ▶ Assessment design
 - within courses and programs for diversity of methods, meaningful feedback, and
 - rigor in making decisions about grading: standard setting.



**EDUCATION AND
SOCIAL WORK**

SOLO

- ▶ ensure Higher Order Thinking Skills are required in assessments
- ▶ But
 - 80%+ of all Teacher Questions (spoken or written) can be answered with Lower Order Thinking Skills
 - By recall or remembering
 - By knowledge
 - By simple handling of a restricted set of ideas, data, knowledge

The Structure of Observed Learning Outcomes (SOLO) Taxonomy



**EDUCATION AND
SOCIAL WORK**

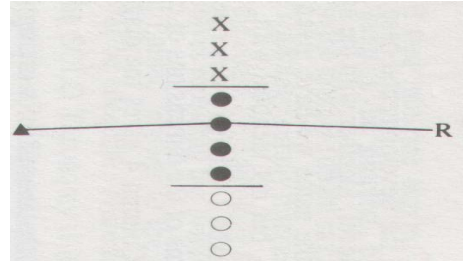
- ▶ A taxonomy developed by analysing the structure of university student responses to assessment tasks by JB Biggs & K Collis, 1982
- ▶ SURFACE (increase in quantity)
 - Unistructural,
 - Multistructural,
- ▶ DEEP (change of quality)
 - Relational,
 - Extended Abstract

Elements of SOLO



EDUCATION AND
SOCIAL WORK

- Given Facts, Ideas, Information
 - Black Dots
- Irrelevant, not Given
 - Xs
- Relevant, not Given
 - Os
- Respondent
 - Triangle
- Response
 - Rs



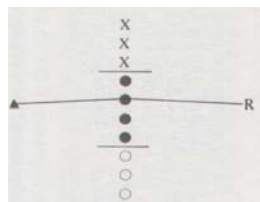
SOLO Diagrammed



EDUCATION AND
SOCIAL WORK

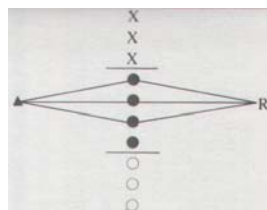
▶ Unistructural

- Use 1 fact or idea



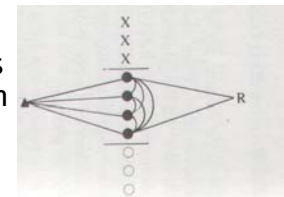
▶ Multistructural

- Use list (2 or more) of facts or ideas but not related to each other



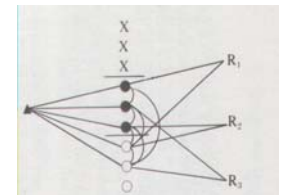
▶ Relational

- How facts or ideas are related to each other



▶ Extended Abstract

- The general underlying principle, rule for set of data, ideas, relationships that gives meaning to all





Using SOLO in item writing

► Achievement Objective: Recognise & Use Factors in a Variety of Practical Settings

UNISTRUCTURAL → Complete $(x + 2)(x + 2) = x^2 + 4x + ___$

MULTISTRUCTURAL → Factorise $x^2 + 2x - 8$.

RELATIONAL → The area of a rectangle in square centimetres is represented by the expression $x^2 + 2x - 8$. The length of the rectangle is $x + 4$ centimetres. What is an expression for the width of the rectangle in centimetres?

EXTENDED ABSTRACT → When the final term of a quadratic is a negative value, what is the pattern of operations within the two factors?

(A) both addition (B) both subtraction (C) addition & subtraction

Structured Superitem



Use this material to answer the set of questions

Pretest–Posttest Control Group Design

O_1	X	O_2
O_3		O_4

U: What does the letter X represent?

- a) The experimental treatment*
- b) The control group
- c) The experimental group
- d) An observation

M: Which symbols represent the experimental group pretest and the control group posttest?

- a) O_1, O_2
- b) O_3, O_4
- c) O_1, O_4^*
- d) O_3, O_2

R: What conclusion can we draw from the ability of this design to control for main effects of history, maturation, and testing?

- a) It is externally valid
- b) It is a true experimental design
- c) It is a quasi-experimental design
- d) It is internally valid*



**EDUCATION AND
SOCIAL WORK**

SOLO applied to holistic marking

- ▶ **A grade/excellence**
 - Sophisticated synthesis of material connecting the given material to abstractions or themes that integrate and give new meaning
- ▶ **B grade/good**
 - Integrated analysis of given material utilising patterns and themes within expected material
- ▶ **C grade/satisfactory**
 - A listing of relevant, given material without attention to patterns or themes

Essays



**EDUCATION AND
SOCIAL WORK**

- ▶ long time—some 2,000 years ago in Imperial China
- ▶ The core is the task, prompt, or question
 - instructs the student as to they type of writing they are expected to engage in.
 - cognitive task
 - discuss, compare, contrast, or analyse, etc.
 - content
 - the causes of World War I, the impact of the setting on a character's development, or the role of mutation in disease).
- ▶ usually can be crafted quite quickly
- ▶ clearly focuses attention on important cognitive and curriculum objectives.
- ▶ Exam use
 - a set period of time,
 - Task not previously seen, and
 - on-demand without time or resources
 - first draft piece of writing
 - generally not best writing

What is really scored in Essays?



EDUCATION AND
SOCIAL WORK

- ▶ Measure
 - Language, Style, Organisation, Preparation;
 - NOT Content, Thinking
- ▶ ELLIS PAGE study of 1,000 scripts, 6 judges
 - Computer Rating used:
 - Intrinsic variables: grammar accuracy, vocabulary
 - Approximations: length, ratio of active to passive voice
 - Average correlation of agreement between Computer and Humans higher than Humans with Humans
- ▶ Notoriously unreliable.
 - Consistency ratings for essays rarely exceed .75
 - Consistency ratings for other paper-and-pencil examinations regularly exceed .90
- ▶ Errors come from
 - the students themselves, the essay questions, and the markers.
 - each source has a different degree of impact on the interpretability of the scores

Solutions



EDUCATION AND
SOCIAL WORK

- ▶ Exclude language & structure features from the task and the scoring system.
- ▶ rethink the structure of the essay prompt or task.
 - provide a structural framework that all students must use
 - Don't use organisational characteristics as a proxy for ability or knowledge in the content area.
- ▶ Structured Marking
 - Rubrics
 - Aligned to curriculum
- ▶ Moderation of Marking
 - Cross-marking
 - Sample at least
 - Measure agreement
 - Correlation, effect size difference

Structured Sequence



EDUCATION AND
SOCIAL WORK

- ▶ “Self-concept and academic achievement are not related”. Write an essay outlining your views on this quotation and the subsequent implications for classroom teachers. Follow the order of points listed below.
 - The evidence on this topic, generally says ...
 - List four contrasting findings from the literature on this topic.
 - How do these studies aid in addressing the topic?
 - Why is it more beneficial to assess how self-concept relates to learning?
 - How does self-concept influence learning, and learning influence self-esteem?
 - What strategies do students use to maintain their ‘status quo’ sense of self-esteem?
 - Note some teaching procedures you, as a teacher, could use to redress these strategies.
- ▶ State whether you think the attached diet is more adequate, adequate, or inadequate in respect to nutrition. Defend your position as follows:
 - a) Identify any items in the diet that you think should be deleted or limited in quantity. Give reasons for your choice.
 - b) Identify any items that you think should be added to the diet or increased in quantity. Give reasons for your answers.
 - c) Make as many summary statements as you feel are necessary to describe the overall adequacy of the diet.

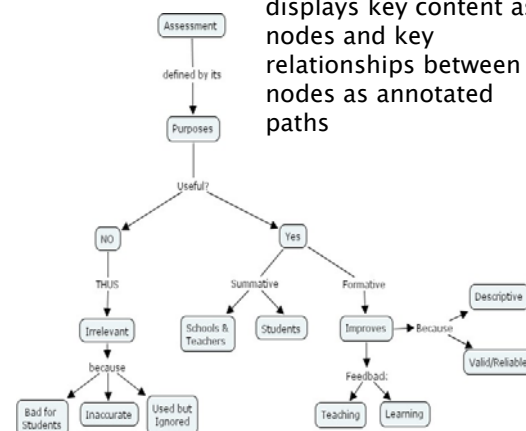
Concept mapping



EDUCATION AND
SOCIAL WORK

- ▶ requires knowledge–transformation and at least relational processing to be able to create connections between content learned and understood and the written essay
- ▶ Shavelson et al. (2005) showed that completing concept maps could be scored reliably to indicate quality of student learning.

A diagram which displays key content as nodes and key relationships between nodes as annotated paths



Moderation of scoring



EDUCATION AND
SOCIAL WORK

- ▶ Cross-checking by having 2 qualified judges mark and compare scores for a common group of essays
 - **Identical scores:** Target is 70% the same
 - **Approximately equal** (+/- 1 score point): Target is 90% the same if using A+ to F scale
- ▶ Debate and discussion and resolution is needed for any essay that differs by more than 1 letter grade or 3/20 or 10/100
 - Discussion must be linked to evidence in essay and criteria in scoring guide
 - If agreement can't be reached need 3rd judge who should be MORE experienced than both markers
- ▶ If you meet the expected targets you can use the scores defensibly to make decisions about learning needs and priorities and to report

	work		
marker	A	B	C
1	B+	C	D+
2	B	C+	C+
3	B+	B-	B-
Midpoint	B+	C+	C
Agreement			
Exact	0.67	0.00	0.00
Approximate	1.00	1.00	0.33

MCQ Item Testing



EDUCATION AND
SOCIAL WORK

- | | |
|---|--|
| <ul style="list-style-type: none"> ▶ Easy to sample widely across domain of interest ▶ Highly manageable ▶ Raise mean achievement – fewer missing data responses ▶ Students like them ▶ Efficiency | <ul style="list-style-type: none"> ▶ Hard to write quality items ▶ Believed to test surface level processing, usually because of poor construction ▶ Guessing factor ▶ May require good reading level ▶ Recognition of answer & test wiseness ▶ Invalid? |
|---|--|

Advantages

Disadvantages

Test of Objective Evidence



EDUCATION AND
SOCIAL WORK

- ▶ Each of the questions in the following set has a logical or “best” answer from its corresponding multiple choice answer set. Best answer means the answer has the highest probability of being the correct one in accordance with the information at your disposal. There is no particular clue in the spelling of the words and there are no hidden meanings. Please record your eight answers.
- ▶ TIP
 - *Think like a ‘test-wise’ student and you should get a high score*

Questions 1--2



EDUCATION AND
SOCIAL WORK

1. *The purpose of the cluss infurmpaling is to remove*
 - a. cluss-prags
 - b. treamails
 - c. cloughs
 - d. pluomots ()
2. *Trassig is true when*
 - a. clump trasses the von
 - b. the viskal flans, if the viskal is donwil or zortil
 - c. the belgo fruls
 - d. dissels lisk easily ()

Questions 3--4



EDUCATION AND
SOCIAL WORK

3. *The sigia frequently overfesks the trelsum because*
- all sigias are mellious
 - all sigias are always votial
 - the trelsum is usually tarious
 - no trelsa are feskable ()
4. *The fribbled breg will minter best with an*
- derst
 - morst
 - sortar
 - ignu ()

Questions 5--6



EDUCATION AND
SOCIAL WORK

5. *The reasons for trystal doss are*
- the sabs foped and the doths tinzed
 - the dredges roted with the crets
 - few rakobs were accepted in sluth
 - most of the polats were thonced ()
6. *Which of the following is/are always present when trossels are being graven?*
- rint and yost
 - Yost
 - shum and Yost
 - yost and plone ()

Questions 7--8

EDUCATION AND
SOCIAL WORK

7. *The mintering function of the ignu is most effectively carried out in connection with*
- a. arazmatoi
 - b. the groshing stantol
 - c. the fribbled breg
 - d. a frailly sush ()
8. _____
- a.
 - b.
 - c.
 - d. ()

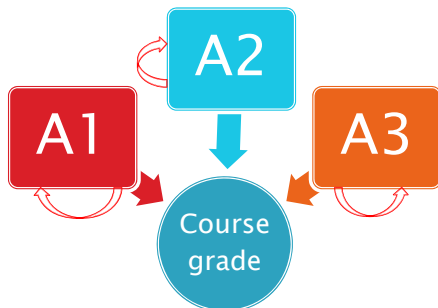
EDUCATION AND
SOCIAL WORK

Current innovations: technology helps

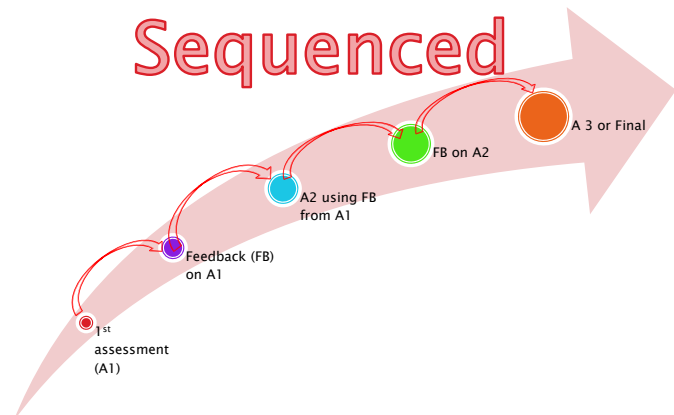
- ▶ Discrete Option MCQ
 - Online, each option scored separately until wrong or all options seen
- ▶ Personalisation
 - Each test has different distractors from a pool of wrong answers
- ▶ Computer adaptive testing
 - Difficulty changes as student gets items right/wrong
- ▶ PeerWise
 - Students right items to test each other, instructor uses best items in future test
- ▶ Statistical item analysis
 - Detect & delete poor quality items

Assessment in Course Design

Conventional



Sequenced

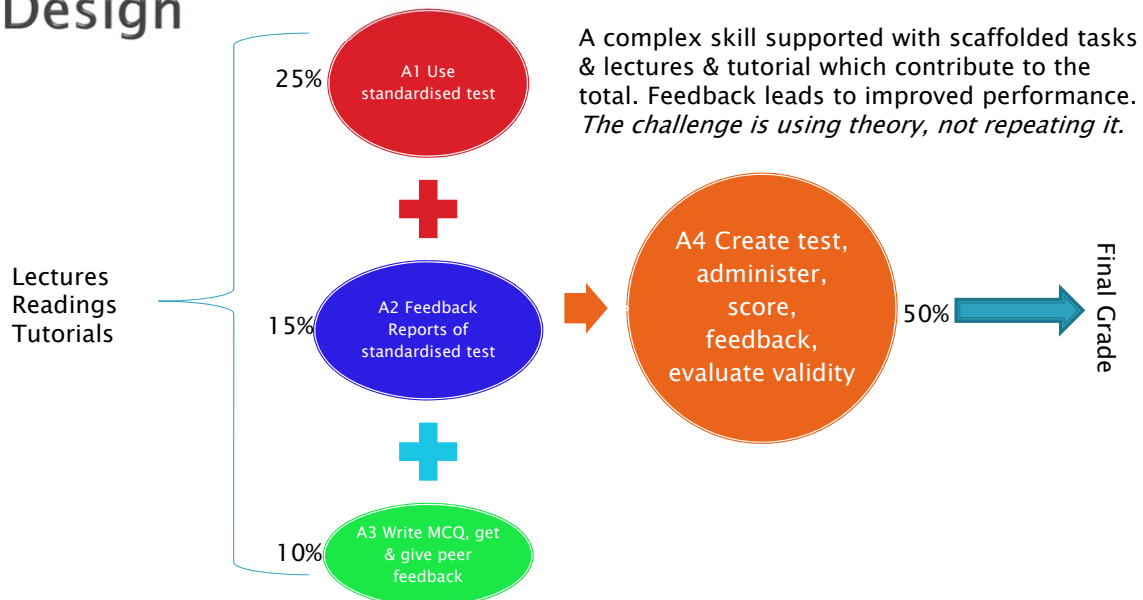


Why would we expect students to use feedback if it isn't related to what they have to do better on subsequent tasks?

Sequenced assessment

- ▶ Depends on deep curricular insights
- ▶ For feedback to be useful:
 - Task 1 must be simpler than subsequent tasks
 - Task 1 skills/knowledge must be useful for course goals & essential for later tasks, so feedback is relevant
 - Task 1 lead to Feedback (diagnostic information) that helps improve Task 2 performance
 - In terms of knowledge, understanding, skills

Educ 224 Assessment Design

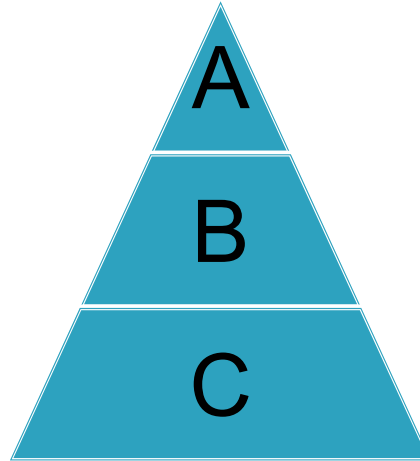


What do grades mean?



EDUCATION AND
SOCIAL WORK

1. Place in rank order
 - Top few
 - Middle some
 - Bottom many
2. Amount answered correctly
 - 85%+ correct
 - 65–84% correct
 - 50–64% correct
3. Quality of performance
 - Excellent;
 - Good;
 - Acceptable;



Are these interchangeable notions?

What's a standard?



EDUCATION AND
SOCIAL WORK

- ▶ The communal judgement of a group of peers
 - Peer review of publications
 - Peer review of graduate theses
 - Peer review of essays, assignments, course-work
- ▶ People who use our grades (employers, parents, students) **MUST** have confidence that we are competent to judge and that we judge competently



Grades & Scores



EDUCATION AND
SOCIAL WORK



High scores do not necessarily indicate high quality performance.
Tests can be made too easy, errors can be made
Grades reflect quality standards

Grades as Academic Standards



EDUCATION AND
SOCIAL WORK

A: Excellent, Highly Accomplished

B: Good, Competent

C: Adequate, Minimally Competent

D: Unsatisfactory, Inadequate

One approach to Standard Setting (modified Angoff)



EDUCATION AND
SOCIAL WORK

- ▶ Decide the level of performance expected of course takers
 - Example: Stage 2, introductory course
 - 'A' means highly competent/excellent for this level of learner
 - What an 'A' beginner can do is not the same as what an 'A' advanced practitioner can do
 - Level of expectation matters
- ▶ Imagine the Minimally competent person (gets C-)
 - What %, grade, score would such a person get on this task?
 - Repeat for **Good** student and **Excellent** student
 - Compare your scores with another tutor/lecturer
 - Set cut scores for grades and award grades based on cut scores not arbitrary scores

Standard Setting on end of course exam



EDUCATION AND
SOCIAL WORK

- ▶ **Goal:** to ensure that all students deemed to be minimally competent received a final grade of C-.
- ▶ **Procedure:**
 - The senior tutor and course lecturer both reviewed the exam paper to determine the minimum marks required to be judged to have met the minimum expectation of the course.
 - The exam had a maximum of 70 marks and a score of 28 was judged to indicate minimum competency having been reached (instead of 35/70).
 - Thus, 7 raw marks or 10% were added to all students' examination scores.
 - The course marks of students who were close to the grade boundaries for A, B, C grades were examined
 - examination scores were further adjusted to ensure that all those who were deemed as meriting the grade received it.

Standard Setting Effect



EDUCATION AND
SOCIAL WORK

- ▶ Coursework: Max=75; M=66% (SD=13%)
- ▶ Examination: Max=25; M=55% (SD=12%)
 - After standard setting M=65% (SD=12%)
- ▶ Final Course Distribution of Grades

<u>Grade</u>	<u>Before</u>	<u>After</u>
◦ Fail	$n=9$	$n=5$
◦ C	$n=31$	$n=19$ Minimally competent beginner
◦ B	$n=8$	$n=17$
◦ A	$n=0$	$n=7$ Highly competent beginner

This is not scaling or norming. It is standard setting. Using professional judgement to determine which work should best be described according to the standards.



EDUCATION AND
SOCIAL WORK

Solutions to challenges

- ▶ You need to ensure that assessment is
 - Appropriate
 - The right things are assessed in the right way
 - Assessments are aligned to curriculum & diverse
 - Educational
 - The sequence and structure of courses & assessments supports learning
 - That feedback supports the development of knowledge & skill
 - Robust, Reliable
 - That quality is being prioritised when judging work
 - End-users are assured that the performances have been awarded the right scores or grades