MCQ Workshop

Lester Gilbert

Associate Professor and Emeritus Fellow,
Electronics and Computer Science

Welcome and introduction

- Competences for workshop

  Attendees will be able to:

  - Maximise the advantages and minimise the limitations of MCQs for summative assessments.
  - Evaluate item and test quality.

- Workshop activities and discussions:
  - MCQ test on pre-reading.
  - Statistical and psychometric quality indicators.
  - MCQ item construction.
  - MCQ effort and cost-benefit.

Assessment: Formative & summative

- Formative: Assessment for learning.
- Summative: Assessment of learning.

- We are going to focus on summative.

Activity 1: Take an MCQ in student mode

- 5 MCQ items based on pre-reading.
- Work individually.
- Some of the questions may be poor, some may be better examples.
- Make a note of your response on the question sheet but also complete the answer sheet so we can analyse the results.
- You have 7 minutes.

MCQ test indicative answers

Q1: 1 or 2

Q2: 3

Q3: 1

Q4: 4

Q5: 1

Q1

Q1 You want to suggest the purchase of a licence for a Departmental item bank with over 30,000 computing and electronics MCQ items at a one-off cost of £15,000. The benefits of such an investment would be greatest if the item bank supports the Department's requirements:

1 For formative rather than summative assessment. Items will have feedback for each distractor which takes more time and effort to produce, this is beneficial.

2 For analytic and evaluative learning outcomes rather than recall and comprehension outcomes. Good analytic and evaluative items take more time and effort to produce, this is beneficial.

3 For items with demonstrated item-test correlations above 0.2. Item-test correlations depend on context, so this is not so useful.

4 For items with demonstrated difficulties between 0.4 and 0.6. Item difficulty depends on context, so this is not so useful.

Both 1 and 2 are the best answers.

Q2

Q2 The head of the School of Acting approaches you and asks where or how MCQ tests might work for their School. Assume that the underlined clauses starting "given" are true. You reply,

1 Given the School emphasises acting practice rather than acting theory, it will need to choose appropriate topics carefully because MCQs work best in the cognitive rather than the psychomotor or affective domains. It is true that MCQs are better suited to the cognitive domain, so while the School will have the opportunity to test acting theory with MCQs, it is true that using MCQs in practical settings needs careful consideration. This means MCQs will not work quite so well for the School which emphasises acting practice.

2 Given the usually subjective and often contentious judgement in staff assessments of student acting ability, the perfect inter-rater reliability of MCQs means that second, shadow, or double-marking examiners would not be required. While it is true that MCQs do not require second marking of student answer sheets as such, additional examiners would nevertheless be required for other quality assurance tasks in MCQ testing, for example, confirming that the options regarded as correct are indeed safely so. Perfect inter-rater reliability is not very helpful in this case.

3 Given that the principles of good acting have been generally articulated and accepted for some time, the module content of the School is particularly stable, which means that the front-loaded investment in MCQs will receive excellent pay-back subsequently. Stable and long-lasting content allows the curation of stable, long-lasting item banks, making this the best answer to give the School.

4 Given that School staff are not particularly numerate, MCQs provide a welcome ability to mark and re-mark an assessment until its internal reliability and consistency reaches an acceptable standard. The ability to iterate an assessment to achieve better quality certainly has some value for the School, but not because its staff are not very numerate.

Q3

Q3 The head of the Department of Psychoanalysis approaches you and asks you about the general **disadvantages** of MCQ testing. You reply that successful MCQ testing requires:

1 Significant front-loaded time and effort compared to essay-type assessment. True, with no "buts".

2 Significant familiarity and comfort with statistical concepts. Somewhat true, but not a disadvantage for motivated teachers.

3 Well-articulated intended learning outcomes. Somewhat true, but in practice a well-articulated syllabus, curriculum, or content outline mitigates the lack of good competences.

4 Summative rather than formative assessment. Not true, but providing effective feedback does requires more effort.

Q4

Q4 Your first topic is "Effective MCQ items for summative assessment". The key point(s) to make in this case is/are the need to:

1 Construct feedback for every distractor option. Not really, feedback is needed far more for formative items.

2 Tie every item to the syllabus. It is always best practice to be able to link an item with a syllabus entry, but failure to be able to do so does not invalidate a summative assessment.

3 Ensure items sample the full range of Bloom's cognitive capabilities from remember through comprehension, analysis, evaluation, and synthesis. Not really, the capabilities to be summatively tested should be given by consideration of the competences being assessed and not by some general requirement to assess the full range of Bloom's taxonomy.

4 Ensure the range of item difficulties are centred on p = 0.5. An effective summative test is one which discriminates the students with the required knowledge and skill from those without, and this is best done with items whose difficulty values vary around 0.5. You will scale afterwards….

Q5

Q5 Your second topic is "Effective MCQ items for formative assessment". The key point(s) to make in this case is/are the need to:

1 Construct feedback for every distractor option. True. Feedback is what characterises an effective formative test.

2 Tie every item to the syllabus. It is always best practice to be able to link an item with a syllabus entry, but failure to be able to do so does not invalidate a formative assessment.

3 Ensure items sample the full range of Bloom's cognitive capabilities from remember through comprehension, analysis, evaluation, and synthesis. Not really, the capabilities to be formatively tested should be given by consideration of the competences being assessed and not by some general requirement to assess the full range of Bloom's taxonomy.

4 Ensure the range of item difficulties are centred on p = 0.5. An effective formative test is one which maximises the opportunity of students to learn. Interestingly, this is best done with items whose difficulty values are well below 0.5, perhaps p around 0.1.

Analytic evaluation

- Characteristics of effective MCQ items and tests.
- Statistical and psychometric analysis of participant answers:
  - Distractor attractiveness, χ^2.
  - Difficulty level, p.
  - Item-test correlation, r.
  - Test reliability, KR-20.

Some MCQ test statistics

Summary of key points

- "Good" summative and formative MCQ items have quite different characteristics.
- A good summative item:
  - Maximises the discrimination of students, difficulty levels centred on p = .5.
  - Distractors are attractive to students without the desired knowledge or skill.
  - Feedback for distractors not particularly needed for summative items.
- A good formative item:
  - Maximises the learning opportunity for students, difficulty levels p << .5.
  - Distractors encapsulate typical errors in desired knowledge or skill.
  - Feedback for distractors the *raison d'être* of formative items.
- Psychometric and other statistical quality indicators suit summative tests and are essential for high-stakes (exam) MCQs, but not so much for formative tests.

Break

- Xx:xx to xx:xx

Guessing

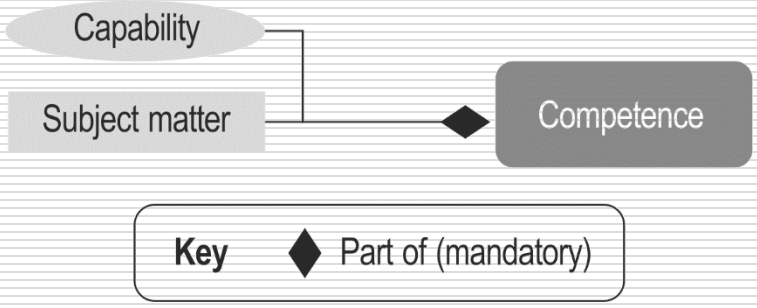
- You can get 25% with random guesses. Sure.
- You can also get 25% with a well-written essay devoid of content.
- Correcting for a wrong answer, presumed to be a guess, introduces personality factors of risk aversion and risk acceptance.
- Simply direct the student to attempt every question. We always bring something to the test in addition to random guessing.
- If “knowing when you don’t know” is a stated aim of the module (eg in medicine) then reward an unattempted question with a mark of 1, a wrong answer with a mark of 0, and a correct answer with a mark of 4.

Scaling

- There is no such thing as a student's "real" mark, since an MCQ assessment could have predominantly easy items or predominantly difficult items.
- Scale the marks to obtain an appropriate average and spread.

Constructing MCQ items

- Item construction:
  - It is essential to have the intended student competence in mind (aka the intended learning outcome, ILO) in mind, and specifically the desired student capability – what should the student be able to do? – in order to construct effective MCQ items.

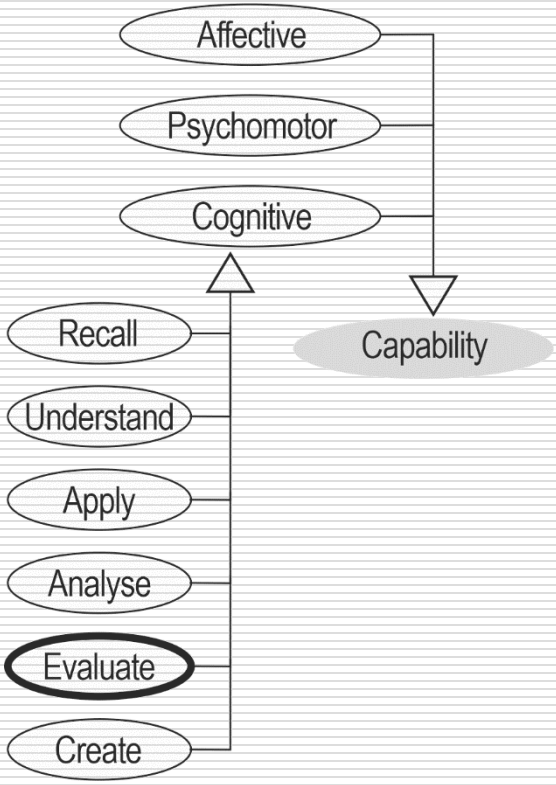


Using competences

- For example:
  - Competence: The student should be able to identify the appropriate agile method of software development for a particular situation.
  - So, we just ask the student to do exactly that.
  - MCQ item:
    "You have been recruited to lead a team of three other developers in a well-established manufacturer of 3D printer accessories. This is your first project, customising the existing Web-based application used by business customers to obtain product information, and place, pay for, and track orders. You decide the most appropriate approach for this project is:
    1. Scrum.
    2. Handcuff.
    3. Kanban.
    4. Extreme programming."

Higher levels of Bloom's taxonomy

- Item construction:
  - While it is quite straightforward to construct MCQ items for lower levels of Bloom's cognitive domain,
    - In principle, effective items can be constructed for higher levels (and for psychomotor and affective domains).
    - In practice, it takes more time and effort, and for e-delivery of psychomotor items, some expense…



Activity 2a: Construct a test item

- Work individually.
- This is the aim:
  "The student should be able to appreciate the value of a structured approach to systems development."
- Construct an appropriate MCQ item.
- You have 1 minute.

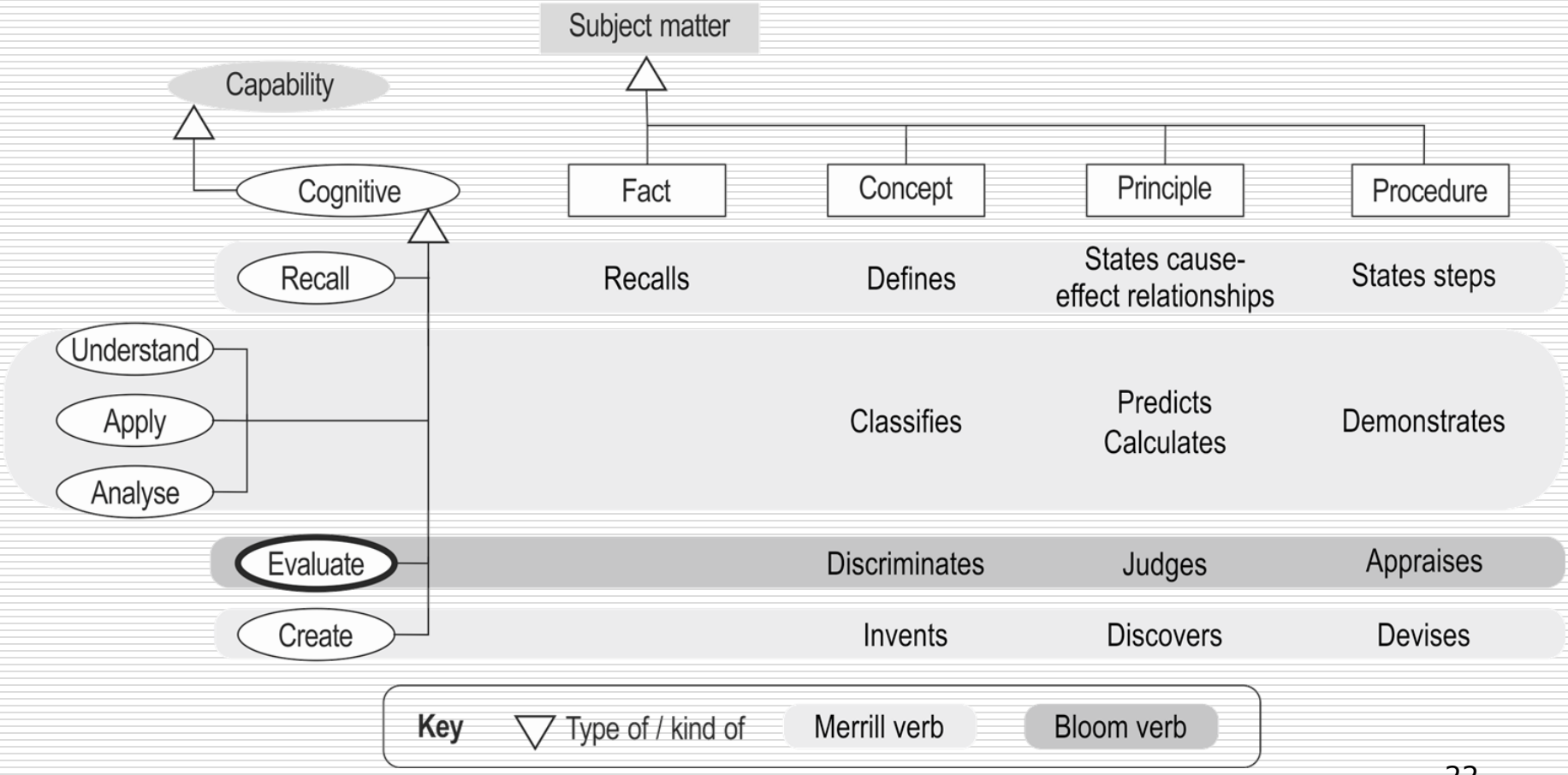
Discussion: Competence, not aim

- Anything starting, “The student should be able to appreciate …” or “The student should be able to understand …” is sadly not a competence and is of little use in constructing MCQ items.
- A competence requires the statement of the student’s behavioural capability, so that the MCQ item simply asks the student to do it.
- There are many lists of capability verbs to suit Bloom’s taxonomy.

	Cognitive ability	Observable capability verb
Remember	Knows terms, specific facts, rules, trends, categories, criteria, methods, procedures, principles, concepts, theories.	Name, label, define, state, recognise, list, recall, identify
Understand	Translates and paraphrases communications; interprets, summarises, and explains relationships; extrapolates from given data.	Explain, classify, summarise, extrapolate, interpret, convert, rehearse
Apply	Applies concepts, principles, rules, procedures.	Calculate, solve, construct, use, prepare, predict, demonstrate
Analyse	Analyses elements, relationships, or organisational principles; analyses connections, relationships, or arrangements.	Compare, contrast, infer, explain
Evaluate	Judges on the basis of criteria and evidence.	Appraise, argue, evaluate, criticise, assess, discriminate
Create	Produces new arrangement or new result.	Compose, originate, design, create, synthesise

Discussion: Competence, not aim

- We like Merrill's taxonomy, a defined vocabulary of 14 capabilities cross-referenced to their relevant subject matter:



Activity 2b: Construct a test item

- Work individually.
- This is the competence:
  "The student should be able to discriminate between structured approaches (such as agile, waterfall, prototyping, and incremental) to systems development."
- Construct an appropriate MCQ item.
- You have 4 minutes.

Discussion: That was easy…

- Competence: The student should be able to discriminate between structured approaches (such as agile, waterfall, prototyping, and incremental) to systems development.
- So, we just ask the student to do exactly that.
- MCQ item:
  "A well-established manufacturer of 3D printer accessories asks for your advice on customising their existing Web-based application for product information and orders. You suggest the following approach to development:
  1. Agile.
  2. Waterfall.
  3. Prototyping.
  4. Incremental."

Summary of key points

- Item construction:
  - It is essential to have the intended or desired student competence in mind, and specifically the desired student capability – what should the student be able to do? – in order to construct effective MCQ items.
  - Effective items can be constructed for all levels of Bloom's taxonomy. It just takes more time and good competence statements…

Activity 3: Is it worth it?

- What are the indicators which make MCQs more, rather than less, cost effective?
- Work in a group of 3.
- Identify 6 key indicators.
- List the indicators in order of significance.
- You have 7 minutes.

Discussion: Is it worth it?

- Factors influencing viability and sustainability of MCQ testing:
  - Large class size.
  - Repeated module delivery.
  - Module (content) stability.
  - Third party module delivery.
  - Competences mainly focus on lower cognitive capabilities.
  - Mastery of content required.
  - Assured quality or consistency of achievement required.
  - Statutory or professional licensing or certification involved.
  - Hazard, risk, or cost in subject matter, procedures, or assessments.

Item banks

- MCQ items are an investment.
- They need curating, called item banking.
- New items should be piloted.
- After each use, their statistics should be noted:
  - Distractor attractiveness, χ^2.
  - Difficulty level, p.
  - Item-test correlation, r.
- After each use, they may need to be tweaked, and possibly discarded.

Mock tests, item pilots

- For summative assessment, a class usually appreciates some examples of the MCQ items they will see. Provide a mock test of perhaps 5 or 6 items which are representative.
- A good opportunity to pilot new test items.

Summary of key points

- Effort and value:
  - Trade-offs of class size, content stability, personal commitment, etc, may or may not make MCQs viable for you.
  - Item banking, required for the committed, still an art form and not a science.
- MCQ high-stakes summative assessment still relatively rare in the Russell Group. Take care!
  - Pilot new items.
  - Provide mock tests.

Thank you!

Questions, comments?

NUS Principles of Effective Assessment

- Should be for learning, not simply of learning
- Should be reliable, valid, fair and consistent
- Should provide effective and constructive feedback
- Should be innovative and have the capacity to inspire and motivate
- Should measure understanding and application, rather than technique and memory
- Should be conducted throughout the course, rather than being positioned as a final event
- Should develop key skills such as peer and reflective assessment
- Should be central to staff development and teaching strategies, and frequently reviewed
- Should be of a manageable amount for both tutors and students
- Should encourage dialogue between students and their tutors and students and their peers

From: Times Higher Education, 29th January 2009.

Also see: https://www.nus.org.uk/en/advice/course-reps/feedback-what-you-can-expect-/

Observable capabilities

	Cognitive ability	Observable capability verb
Remember	Knows terms, specific facts, rules, trends, categories, criteria, methods, procedures, principles, concepts, theories.	Name, label, define, state, recognise, list, recall, identify
Understand	Translates and paraphrases communications; interprets, summarises, and explains relationships; extrapolates from given data.	Explain, classify, summarise, extrapolate, interpret, convert, rehearse
Apply	Applies concepts, principles, rules, procedures.	Calculate, solve, construct, use, prepare, predict, demonstrate
Analyse	Analyses elements, relationships, or organisational principles; analyses connections, relationships, or arrangements.	Compare, contrast, infer, explain
Evaluate	Judges on the basis of criteria and evidence.	Appraise, argue, evaluate, criticise, assess, discriminate
Create	Produces new arrangement or new result.	Compose, originate, design, create, synthesise