Making better use of existing multiple-choice test questions

Lucia Otoyo
London South Bank University

Apt Conference 2019, Monday 1 July, UCL
Contents

• Overview of Traditional MCQs
• A better test format for formative assessment
• Better test formats for summative assessment
• Results of a recently published study
Example of a ‘4 choice’ MCQ

Which one of these animals is a mammal?

A

B

C

D
Traditional MCQs - Pros

• Familiarity – one of the most widely used assessment types
• Educators already have a lot of experience in writing MCQs
• Reuse existing banks of MCQs
• Easy to administer to large cohorts
• Clear rubric for assessment
• Efficient to mark
Traditional MCQs - Cons

- Can only test superficial knowledge? (next slide)
- No immediate feedback = no new learning
- Unable to assess partial knowledge = guesswork
- Guesswork = unreliability
Questions *can* test thinking

Checkmate in 3 moves?
A “good” test

Requirements:
- validity
- reliability
A traditional 4-choice MCQ

Average mark

1 mark

'a pure guess'

0.25

'educated' guesses

distracters identified

0 1 2 3
...now with +3/-1 negative marking

Average mark

100%

distracters identified
How we think about m-c test questions

What’s in our minds when we try to answer an m-c question?

We can’t always identify a clear first choice answer option, so we make some guesses! hence unreliable test scores!

Maybe a Likert scale test format would be better...
Likert scale test format

This would give test takers full freedom to express their thoughts about each answer.

No more guesswork!

But how to assign a score?
Answer Until Correct

- approximation to Likert
- immediate feedback - shown to aid learning
- partial knowledge expression

“IF-AT” card

Image of a card with a scoring system for immediate feedback assessment technique (IF AT®). The card includes columns labeled A, B, C, D, and a column for scoring. The card is marked with stars and crosses to indicate correct answers. The card is held by a hand, suggesting it is used for feedback assessment.
In this recent journal paper...


...the authors state:

"Why does the scratch-card IF-AT works better than the computer-based IF-AT in language arts and social sciences? One possibility is that the scratch-card IF-AT can let students scratch off his/her answer as if scratching a lottery ticket. It is a scratch-and-win game, which sustains student curiosity and interest. To simulate an instant lottery ticket, the answer keys in the computer-based IF-AT may be re-designed to appear similar to a lottery scratched box with an opaque coating."
Let's learn about chocolate!

http://tiny.cc/aptconf19
Subset Selection

Another approximation to Likert is “subset selection”. This is the marking scheme we use at LSBU:

• correct answer only selected => 1 mark
• correct answer plus one wrong answer => 0.5
• correct answer plus two wrong answers => 0.33
• all answers (or no answers) selected => 0.25
• any other response => 0

Here is a live demo on QuizSlides: https://quizslides.co.uk/quiz/0ZLXp
Elimination Testing

It’s essentially a mirror image of subset selection; test takers are asked to select the \textit{wrong} answers. This is the usual marking scheme:

- each \textit{wrong} answer selected $\Rightarrow$ 1 mark
- each correct answer selected $\Rightarrow$ -3 marks!

Empirical Study - purpose

• Compare reliability of Trad vs Subset selection tests

• Analyse the extent of student perceived and actual guesswork

• Evaluate student preference of Trad vs Subset
Study background + about

- Answer until correct for formative assessment + subset selection tests for summative assessment

- 75 x 1st year students on a Computer science module: Media, Computers and Networks

- 3 ‘dual’ tests: traditional and subset selection (quantitative)

- Optional questionnaire after the final test (qualitative)
Study results…

“In the test that you have just taken, to what extent did you consciously pick one answer at random between two or more of the answer options when you were unsure which answer was correct?”

- I always guessed one answer. 44.2%
- I mostly guessed one answer. 23.3%
- I guessed about half the time. 11.6%
- I occasionally guessed one answer. 11.6%
- I never guessed; I always selected two or more of the answer options or I skipped the question. 9.3%
Study results – perceived guessing

• 2/3 always or mostly select more than 1 option when unsure

• Students want to express their partial knowledge and make use of subset selection rather than guessing
Study results – use of subset selection

How many answers per question were selected on average when taking subset selection tests? Our findings:

- For over 20% of questions the students expressed their partial knowledge and didn’t engage in guesswork!
Study results – reliability

Reliability of Subset Selection vs Traditional tests

Cronbach's alpha, with 95% confidence intervals shown

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>Trad.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>Test 2</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>Test 3</td>
<td>0.8</td>
<td>0.7</td>
</tr>
</tbody>
</table>
Study results – student preference

How did our students feel about subset selection tests versus traditional multiple-choice tests? Our findings:

- More students found subset selection tests less stressful (22 students) compared to traditional tests (12 students).
- More students overall preferred subset selection tests (28 students) compared to traditional tests (16 students).

Legend:
- Blue: Subset selection
- Orange: Traditional
- Gray: No difference
- Green: No response
Study results – further details…

Conclusion/discussion

• Subset selection & IF-AT better alternatives to Trad MCQs
• At LSBU subset in continual use for over 20 years
• Supported by students and EEs (reliability + less stress)

• What are your barriers to using these alternative test formats?
Any questions?

lucia.otoyo@lsbu.ac.uk