Framework for Undergraduate Mathematical Digital Testing

Alisa Lochner
University of Twente
a.j.Lochner@student.utwente.nl
Why?

The need
- Larger student numbers\(^1\)
- Reduced Resources\(^1\)

The advantages
- Can be as reliable as constructed-response tests\(^2\)
- Possibility of covering a wider variety of topics\(^2\)

MCQ example

Calculate the indefinite integral \( \int \ln(3x) \, dx \)

(A) \( \frac{x}{3} (\ln(3x) - 1) + C, \quad C \in \mathbb{R} \)

*(B) \( x(\ln(3x) - 1) + C, \quad C \in \mathbb{R} \)

(C) \( 3x(\ln(3x) - 1) + C, \quad C \in \mathbb{R} \)

(D) \( \frac{1}{x} + C, \quad C \in \mathbb{R} \)

Torres, Lopes, Babo and Azevedo (2009)
Multiple Choice Questions (MCQ) Concerns

• Choosing instead of making\(^1\,^2\)

• Difficulty

• Guessing\(^3\)

MCQ Writing Framework

Content
• Every item should concern one specific mental behaviour
• Use novel material to limit simple recall

Style
• Minimise the amount of reading in each item

Choices
• Research suggests three is adequate
• Make sure none of the choices overlap
• Keep choices homogeneous

Haladyna, Downing, & Rodriguez (2002)
(weak) MCQ example

Weak example:  

Calculate the indefinite integral \( \int \ln(3x) \, dx \)

(A) \( \frac{x}{3} (\ln(3x) - 1) + C, \ C \in \mathbb{R} \)

*(B) \( x (\ln(3x) - 1) + C, \ C \in \mathbb{R} \)

(C) \( 3x (\ln(3x) - 1) + C, \ C \in \mathbb{R} \)

(D) \( \frac{1}{x} + C, \ C \in \mathbb{R} \)

Fig. 4. Weak example of MC with no homogeneous alternatives

Also a weak example due to content, as options can be differentiated to get back to the question, which is a different learning goal than intended

1: Torres, Lopes, Babo and Azevedo (2009)
Quality Testing - Statistics

**Difficulty** (P-value)
- The proportion of examinees who selected the correct option.
- A low P-value \((p<0.30)\) indicates a difficult question whilst a high \(P-value\) indicates an easy question \((p>0.80)\)

**Item Discrimination Index**
- Item Discrimination Index\(^1\).
- Between -1.0 and +1.0.
- Above 0.30 is considered good.

**Distractors**
- For it to be a good distractor, at least 5% should choose it\(^2\).

1:Johnson (1951)  2:DiBattista and Kurzawa (2011)
### Pilot at the University of Twente

**June 2017**

**Mathematics 1D**

N = 494

Pass percentage of exam: 72%

Harry Aarts, Steffen Posthuma, Karen Slotman, Bernard Veldkamp, Jan van der Veen, Jan Willem Polderman

Hybrid exam consisting of:

- Questions 2, 3, 4, 7, 8 and 11b were multiple choice.
- Questions 1, 5, 9, 11a and 12 were final answer based.
- Question 6 and 10 were paper-based hand written.

This table can also be seen in the SUTQ project by Harry Aarts, called: “A Hybrid Test for Mathematics”, submitted 5 March 2018. For his analysis on the same data, he can be contacted on h.f.m.aarts@utwente.nl for a copy of the project contains his findings.
Choose the image under the transformation enclosed by the hyperbolas $xy = 1$ and $xy = 4$ and the lines $y = x$ and $y = 4x$

\[
y = x \quad \text{subs: } x = \frac{u}{v} \text{ and } y = v
\]

\[
v = \frac{u}{v}
\]

\[
v^2 = u
\]

\[
v = \sqrt{u}
\]

\[
v = \frac{4u}{v}
\]

\[
v^2 = 4u
\]

\[
v = 2\sqrt{u}
\]

---

<table>
<thead>
<tr>
<th>Question</th>
<th>P-value</th>
<th>Upper group</th>
<th>Lower group</th>
<th>Item Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>11b</td>
<td>0.63</td>
<td>0.90</td>
<td>0.40</td>
<td>0.50</td>
</tr>
</tbody>
</table>
Digital Testing Question 2

2. [3 pt]
Let \( f(x, y) = f(x(u, v), y(u, v)) \) and \( x(u, v) \) and \( y(u, v) \) be differentiable functions.

Use Tables 1 and 2 to determine \( \frac{\partial f}{\partial v}(u, v) \) in \( (u, v) = (1, 2) \).

1. \((u, v) = (1,2)
   \[x(1,2) = 3 \quad ; \quad y(1,2) = 0 \]
   \((x, y) = (3,0)\)

2. \[
   \frac{df}{dx} \cdot \frac{dx}{dv} + \frac{df}{dy} \cdot \frac{dy}{dv}
   \]

3. \[
   = -5 \times 4 + 9 \times 5
   = -20 + 45
   = 25
   \]

Choose from the alternatives below and fill in your answer on the answer sheet:

(a) \( 31 \)  (b) \( -21 \)  (c) \( 0 \)  (d) \( 25 \)
(e) \( 58 \)  (f) \( -32 \)  (g) \( 52 \)  (h) \( 26 \)

(26%)

(55%)

(6%)

<table>
<thead>
<tr>
<th>Question</th>
<th>P-value</th>
<th>Upper Group</th>
<th>Lower Group</th>
<th>Item Discrimination Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.26</td>
<td>0.50</td>
<td>0.17</td>
<td>0.33</td>
</tr>
</tbody>
</table>
Digital Testing Question 3

3. [2 pt]
Consider the function \( f(x, y) = xe^{-y} + 3y \) and the point \( P(1, 0) \).
Determine the unit direction \( u \) for which \( D_u f(P) \) is maximal.

Choose from the alternatives below:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>( \frac{1}{\sqrt{17}} ) ( i ) + ( \frac{4}{\sqrt{17}} ) ( j )</td>
<td>(b)</td>
<td>( -j )</td>
<td>(c)</td>
</tr>
<tr>
<td>(39%)</td>
<td>(5%)</td>
<td>(8%)</td>
<td>(19%)</td>
<td></td>
</tr>
<tr>
<td>(e)</td>
<td>( \frac{1}{\sqrt{5}} ) ( i ) + ( \frac{2}{\sqrt{5}} ) ( j )</td>
<td>(f)</td>
<td>( i )</td>
<td>(g)</td>
</tr>
<tr>
<td>(6%)</td>
<td>(3%)</td>
<td>(16%)</td>
<td>(5%)</td>
<td></td>
</tr>
</tbody>
</table>

1. \( \nabla f(x, y) = (e^{-y}, -xe^{-y} + 3) \)
2. \( \nabla f(x, y) = (1, 2) \) at \( P(1,0) \)
3. \( u = \sqrt{x^2 + y^2} \)
   \( u = \sqrt{1^2 + 2^2} \)
   \( u = \sqrt{5} \)
4. Therefore, \( \frac{1}{\sqrt{5}} \) \( i \) + \( \frac{2}{\sqrt{5}} \) \( j \)

---

<table>
<thead>
<tr>
<th>Question</th>
<th>P-value</th>
<th>Upper Group</th>
<th>Lower Group</th>
<th>Item Discrimination Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.39</td>
<td>0.68</td>
<td>0.19</td>
<td>0.49</td>
</tr>
</tbody>
</table>
Suggestions towards Framework for Undergraduate Mathematics

From Haladyna et al., 2002.

Content.

• Every item should concern one specific mental behaviour → Limit the question to testing one concept.
• Use novel material to limit simple recall → A must, to avoid trivial questions

Style.

• Minimise the amount of reading in each item.

Choices.

• Research suggests three is adequate → As many as five distractors can still be effective.
• Make sure none of the choices overlap → Choices should also not be mathematically equivalent.
• Keep choices homogeneous → Especially in questions regarding graphs.

In addition

• Good distractors that catch very bad misconceptions are key in item difficulty.
• Avoid distractors which are the result of small calculation errors.
• Limit question to 3 - 4 reasoning procedures, which contain simple arithmetic.

1: Jonassen, 2000
Conclusions

• MCQ can test difficult content.
• With unfamiliar context MCQ increases difficulty.
• With many options, guessing success is reduced.
• With well written distractors, students have to work out their answers in more detail to get to the right answer.

However

Multiple choice is still limited in measuring long-chains of reasoning.

100% digital testing exams for summative assessment might be possible with further research.
References

• Aarts, H.F.M. (2018). A Hybrid Test for Mathematics. SUTQ Project. Contact: h.f.m.aarts@utwente.nl for a copy.