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]

1: Nicol (2007) 2: DiBattista and Kurzawa (2011)

MCQ example

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

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

*(B) $x\left(\ln(3x)-1\right)+C,\ C\in\mathbb{R}$ ← Correct Answer

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

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

Item: the question and options (A)–(D)

Distractors: (A), (C), (D)

Torres, Lopes, Babo and Azevedo (2009)

Multiple Choice Questions (MCQ) Concerns

- Choosing instead of making[1,2]
- Difficulty
- Guessing[3]

1: Nicol (2007)   2: DiBattista and Kurzawa (2011)   3: Quaigrain and Arhin (2017)

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:[1]

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

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

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

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

(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

Item	Max Score	P
1	2	0.62
2	3	0.26
3	2	0.39
4	2	0.60
5	3	0.78
6	6	0.62
7	1	0.82
8	2	0.87
9	3	0.68
10	6	0.65
11a	1	0.78
11b	2	0.63
12	3	0.49

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.

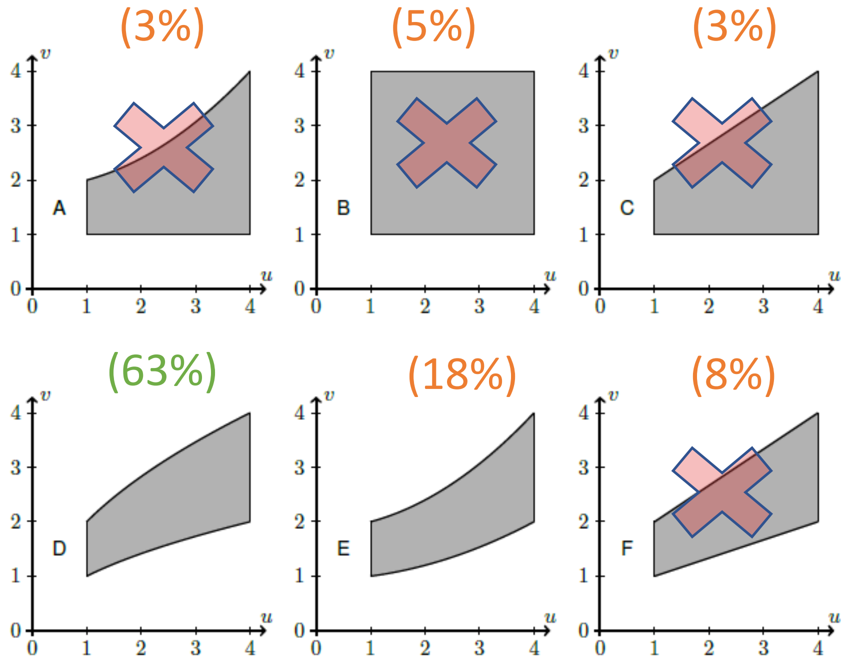
Digital Testing Question 11b

11. [3 pt]

Consider the transformation $x = \frac{u}{v}, \quad y = v \quad (u > 0, v > 0)$.

(a) [1 pt] Determine the Jacobian $J(u, v) = \frac{\partial(x, y)}{\partial(u, v)}$ of this transformation.

(b) [2 pt] Choose, from the six figures below, the correct sketch of the corresponding image under this transformation of the region in the first quadrant enclosed by the hyperbolas $xy = 1$ and $xy = 4$ and the lines $y = x$ and $y = 4x$.



Choose the image under the transformation enclosed by the hyperbolas $xy = 1$ and $xy = 4$ and the lines $y = x$ and $y = 4x$

$y = x \qquad\qquad y = 4x$

subs: $x = \frac{u}{v}$ and $y = v$

$$v = \frac{u}{v} \qquad\qquad v = \frac{4u}{v}$$

$$v^2 = u \qquad\qquad v^2 = 4u$$

$$v = \sqrt{u} \qquad\qquad v = 2\sqrt{u}$$

Question	P-value	Upper group	lower Group	Item Discrimination Index
11b	**0.63**	0.90	0.40	**0.50**

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 $(u,v) = (1,2)$.

(x,y)	$(0,0)$	$(1,2)$	$(3,0)$
$f(x,y)$	1	6	-2
$f_x(x,y)$	-1	3	-5
$f_y(x,y)$	-2	8	9

Table 1

(u,v)	$(0,0)$	$(1,2)$	$(3,0)$
$x(u,v)$	-2	3	7
$y(u,v)$	6	0	3
$x_u(u,v)$	1	2	6
$x_v(u,v)$	0	4	1
$y_u(u,v)$	3	-4	2
$y_v(u,v)$	0	5	-3

Table 2

1. $(u,v) = (1,2)$
   $x(1,2) = 3$; $y(1,2) = 0$
   $(x,y) = (3,0)$

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 (26%)

(e) 58 (f) -32 (g) 52 (55%) (h) 26 (6%)

Question	P-value	Upper group	lower Group	Item Discrimination Index
2	**0.26**	0.50	0.17	**0.33**

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 $D_{\mathbf{u}}f(P)$ is maximal.

Choose from the alternatives below:

(a) $\frac{1}{\sqrt{17}}\mathbf{i} + \frac{4}{\sqrt{17}}\mathbf{j}$ (5%)

(b) $-\mathbf{j}$

(c) $\frac{1}{\sqrt{10}}\mathbf{i} + \frac{3}{\sqrt{10}}\mathbf{j}$ (8%)

(d) $\mathbf{i} + 2\mathbf{j}$ (19%)

(e) $\frac{1}{\sqrt{5}}\mathbf{i} + \frac{2}{\sqrt{5}}\mathbf{j}$ (39%)

(f) $\mathbf{i}$ (6%)

(g) $\mathbf{i} + 4\mathbf{j}$ (16%)

(h) $\mathbf{j}$ (5%)

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, $\frac{1}{\sqrt{5}}\boldsymbol{i} + \frac{2}{\sqrt{5}}\boldsymbol{j}$

Question	P-value	Upper group	lower Group	Item Discrimination Index
3	**0.39**	0.68	0.19	**0.49**

Suggestions towards Framework for Undergraduate Mathematics

From Haladyna et al., 2002. | *Undergraduate Mathematics*

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[1]

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.
- Torres, C., Lopes, A. P., Babo, L., Azevedo, J. (2009). Developing Multiple-Choice Questions in mathematics. In *Proceedings of ICERI 2009 - International Conference of Education, Research and Innovation*. pp. 6218–6229.
- DiBattista, D., Kurzawa, L. (2011). Examination of the Quality of Multiple-choice Items on Classroom Tests. *The Canadian Journal for the Scholarship of Teaching and Learning*. *2011,* 2(2). doi:http://dx.doi.org/10.5206/cjsotl-rcacea.2011.2.4K.
- Haladyna, Downing, & Rodriguez (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied measurement in Education, 15*(3)*, 309-334.*
- Johnson, A. P. (1951). Notes on a suggested index of item validity: The U-L Index. *Journal of Educational Psychology, 42*(8), 499-504.
- Jonassen, D. H. (2000). Toward a Design Theory of Problem Solving. *Educational Technology Research and Development*, 48, 63-85. doi:10.1007/BF02300500
- Nicol, D. (2007). E-assessment by design: using multiple-choice tests to good effect. *Journal of Further and Higher Education*, *31*(1), 53–64
- Quaigrain, K., & Arhin, A. K., & King Fai Hui, S. (2017). Using reliability and item analysis to evaluate a teacher-development test in educational measurement and evaluation. *Cogent Education*. *4*(1). doi:10.1080/2331186X.2017.1301013.

Q & A