VIV ROLFE

SEVEN STEPS TO MCQ HEAVEN

vivrolfe.com

GETTING STARTED WITH INCLUSIVE AND EFFECTIVE MULTIPLE CHOICE QUESTION ASSESSMENT

BY DR VIVIEN ROLFE

AND THE WORKSHOP HELD AT THE UNIVERSITY OF THE WEST OF ENGLAND BRISTOL

3RD DECEMBER 2015

ABOUT THE AUTHOR

I am a UK National Teaching Fellow and alongside my university role I am invited to run workshops and sessions on learning, teaching and assessment across the UK. My specialist interests are in the application of learning technology and open education, and I like to explore the range of teaching practices and learning opportunities that have been made available through the use of open technology and open licensing – such as Creative Commons. In recent workshops at Thomson Rivers University in Canada I was involved in discussions about innovation and change in higher education, and also the state of play of educational research (TRU, November 2015).

As a Principal Fellow of the Higher Education Academy my roles within my institution involve supporting and mentoring aspiring fellows. I co-lead the Education Research Network (ERNie) that is an informal networking opportunity for over 100 colleagues to share education research techniques and to discuss the broader political contexts of the higher education landscape.

I run five blogs that share information and resources openly, and this eBook and further details about MCQ can be found on vivrolfe.com.

Just search for 'MCQ' in the search field or tag cloud.

I hope you enjoy this guide!

Viv
Twitter: @VivienRolfe

THE AIM OF THIS EBOOK

This document will hopefully help you develop a process for the design of good quality multiple choice questions (MCQs) for student assessments and evaluations. The aim is to provide a practical guide rather than a document underpinned with theory.

INDEX

If you are using this book as a workshop with colleagues, you might wish to consider the following activities.

WORKSHOP ACTIVITIES
True or false?

MCQ offer accessible assessment options
They are easy to write
They measure factual recall

MCQ BRAINSTORM

Get the group to list the pros and cons of MCQ

> *e.g.*
> *Automated marking (optical mark sheets, LMS)*
> *Questions often poorly written*
> *Difficult to question critical thinking (but possible)*
> *Easy to test the ability of the learner to perform the test rather than their knowledge.*
> *Online – Bb / Moodle – learner answers can be built in, multiple repeats – so can become an effective learning tool themselves.*

HANDS UP?
Who has ever received training on how to write good MCQ assessments?
Who uses MCQ assessments?

BACKGROUND CONTEXT

With the massification of higher education and advent of larger classes in many subjects, MCQ is a popular choice for assessment with academic teams. In Di Battita and Kurzawa's paper they reviewed a single Canadian institution, and 51% of instructors there used MCQ assessment, most commonly with 1st year classes (68%). Since MCQ holds the weight of such high proportions of course and programme assessment, it is essential we design these to the highest quality.

As also highlighted in this study, the majority of instructors had never received any formal training in how to write and validate tests, and I suspect this is also the case in the UK. In some US institutions, all summative tests – MCQs used for phase tests or end of module/course examinations – are scrutinised for quality and reliability. It would be useful for us to at least think through how these processes might work in our programme teams within our own institutions.

1. ASSESSMENT DESIGN

Q Where do you use multiple choice questions as part of your instruction and assessment?

MCQs can be used for many different purposes in education settings. The main terminology to understand here is the difference between formative [https://en.wikipedia.org/wiki/Formative_assessment] and summative assessment [https://en.wikipedia.org/wiki/Summative_assessment]. As you read through these notes you will begin to realise you can adopt different design approaches for both.

It is not unusual for new students to undergo some diagnostic testing, e.g. to test their background knowledge, their digital or numerical literacies for example. As they progress through their studies they might experience formative assessment – that is assessment and feedback that is developmental and 'informative'. MCQs are ideal for this purpose, and can be engineered in most software to incorporate a positive learning opportunity, working toward the notion of assessment FOR learning rather than assessment OF learning (Jisc 2015).

MCQ can be summative – that is, evaluate an individuals understanding of learning outcomes to generate marks as part of a final assessment. These might be regular 'phase tests' run through the duration of the module or course, or as well as a final MCQ assessment positioned at the end of a course or module.

APPLICATIONS OF MCQ TESTS

- Self-testing / diagnostic testing of knowledge or ability (formative)
- Phase tests (summative)
- End of module tests (summative)
- Evaluations (questionnaire)

Of course, we also use MCQ in order to evaluate our teaching practice, and the use of MCQ in qualitative research methodology follows some similar principles. More in-depth reading of the application of MCQ in questionnaires can be found in Rattray and Jones (2007).

PRE-DESIGN CONSIDERATIONS

Before you even get started writing questions, as tempting as this often is, some important principles apply and need thinking through with your teams.

1) What is the purpose of the MCQ test – formative or summative?
2) How do you intend to deploy it – there might be validation/reliability features built into online systems?
3) Are your questions challenging and discriminating?
4) Are you assessing the learner knowledge and understanding, or their ability to do the test?
5) Is your test inclusive?

2. QUESTION DESIGN

IT IS VERY EASY TO WRITE BAD QUESTIONS!

It is notoriously difficult to write consistently good quality MCQ. It is not uncommon to see questions on examination papers with different numbers of options (distractors or stems, see below), be worded so poorly that it is possible to guess the correct answer, and all manner of other easy pitfalls.

This learning resource from Phil Race is an excellent example of this and will open your eyes to the possibility of easy errors. I've used this with students in class to also help them understand some of the tricks behind answering MCQ. Have a go at Phil's test and see how you go (LINK).
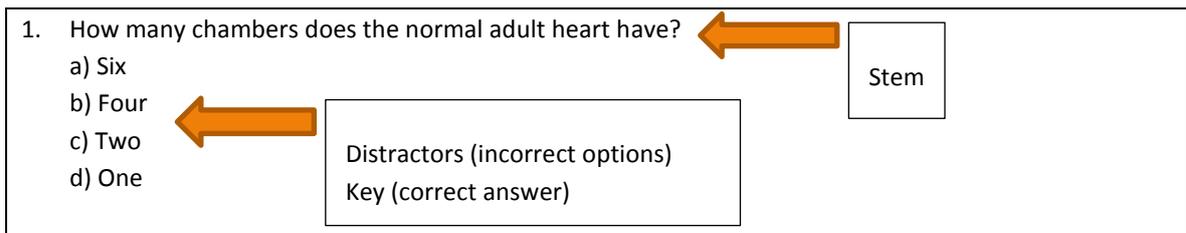
Common pitfalls in this example:

- Poor grammar that leads to the selection of the correct answer
- Elaborate choices that are the obvious correct answer
- A pattern to the answers – ABCD, ABCD or all B's that lead to the correct answer

Here are some points for consideration in no particular order?

ARCHITECTURE OF THE QUESTION

The question. This needs to be a clearly written and unambiguous statement. The writing of the distractors (or stems) is the hardest part of all and will make or break your MCQ test.

1. How many chambers does the normal adult heart have?
   a) Six
   b) Four
   c) Two
   d) One

   Stem

   Distractors (incorrect options)
   Key (correct answer)

ALIGN ASSESSMENT WITH LEARNING OUTCOMES

As with all good assessment practice, ensure it aligns to your learning outcomes. If you are working on a module with other colleagues, they will need to input into the question design and development. It is very easy to write an MCQ 'outside of the box' just because you think it looks to be an interesting question. Whilst not irretrievable, it is immensely inconvenient to have to alter the marks awarded in a summative test because of a poor question, and this will require transparency with external examiners and students. I have known it happen because mistakes simply do.

HOW MANY DISTRACTORS TO HAVE?

Most commonly, 4 or 5 distractors are used. Simply, if you have 4 options, the learner will have a 25% chance of guessing a correct answer. If you have 5, they will have a 20% chance. This obviously makes for a more challenging test, and you might be required to do this to comply with Professional Body standards, or you may decide as a programme team to produce more rigorous testing in more advanced years of study.

WHAT ABOUT NEGATIVE MARKING?

Negative marking has been more widely used in the past as it is now. Once the mainstay of Medical School assessment, particularly in the pre-clinical years, this is an area where there is a body of work that has made us think more deeply about applying a negative marking scheme. This generally means if a student gets a question wrong, they lose a mark. It makes an assumption about the reasons for getting wrong in an attempt to compensate for guessing. What research shows us now is that students try and be strategic knowing that they may lose marks, and therefore, the test provides a complicated picture of student strategy and levels of confidence rather than knowledge (Holsgrove 2001).

> *Negative marking have too many adverse characteristics. We can do better. (Holsgrove 2001).*

Negative marking throws up a number of other problems. Which student knows more? One who makes a bold attempt at answering all the questions but loses marks through the negative marking scheme? The other who gets the same mark by only responding to questions they know the answer to?

WRITING GOOD QUESTIONS AND DISTRACTORS

The biggest tip is not to rush writing questions and to involve as many people as possible. Section XX looks at students as co-creators which has benefits for a number of reasons.

Writing the question:

- Write a complete question not a vague phrase
- Keep questions in the test independent of each other
- Use very clear English (avoid colloquialisms, cultural references and complicated language)
- Avoid grammatical errors that might give rise to the answer (stem ending with 'an' leading to a distractor starting with a vowel, or errors within the use of singular and pleural items

Writing the distractor:

- Ensure there are consistent numbers throughout the test (e.g. 4 or 5 throughout is commonplace)
- Brainstorm them, research them, base them on common errors or misconceptions
- Avoid silly ones
- Make them mutually exclusive
- Avoid pitfalls and question styles that are not inclusive (see section 5)

Q1
The functional unit of the kidney is

a. The nephron (correct)
b. Contains the juxtaglomerular apparatus
c. The Crypt of Lieberkuhn
d. Donald Duck

Critique – clearly a better question is "Which of the following listed below is the functional unit of the kidney?"

Q2
Which scientists, including a researcher from the University of Nottingham, won the Nobel Prize for their "discoveries concerning magnetic resonance imaging" in 2003?

a. Paul C. Lauterbur and Sir Peter Mansfield (correct)
b. Richard Axel and Linda B. Buck
c. Barry J. Marshall and J. Robin Warren
d. Sir John B. Gurdon and Shinya Yamanaka

Critique – unnecessary reference to the University of Nottingham, and also a cultural reference that might advantage someone from the city.

Further reading:

Do read through these resources thoroughly as they go into far more detail of how to write good questions and distractors, and both give plenty of examples.

Brame C (2015). Vanderbilt University: Writing good multiple choice questions. Available: https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/

Maber J, Booth A, Hamburg L and Wassall T (2015). Leeds Beckett. MCQ help – writing good questions. Available:
https://my.leedsbeckett.ac.uk/bbcswebdav/institution/Health_and_Social_Sciences/ltu/areas/assessment/mcq/design.html

3. MCQ – DESIGN FOR LEARNING?

RECALL, INTERPRET, SOLVE

There is a misconception that MCQ can only test basic recall of knowledge. Carefully designed questions can evoke critical thinking and be aligned to Bloom's taxonomy. Questions can be designed for interpretation and

problem solving, and lend themselves well for example to mathematical or scientific problems that lead to one correct answer.

For advanced levels of study, a blend of recall and problem solving would make for a good test, bearing in mind that clearly, the longer and trickier questions would take longer for students to solve.

WHAT IS BLOOM'S TAXONOMY

Bloom derived a taxonomic framework [https://en.wikipedia.org/wiki/Bloom%27s_taxonomy] to provide a common education language. Those relevant to the design of assessments fall within the 'knowledge domain', and Bloom elaborates with a series of verbs that climb an intellectual pathway: knowledge, comprehension, application, analysis, synthesis and evaluation.

The following diagram expands this idea and provides question suggestions aligned to each part of the framework. When designing assessment, more simplistic models can be applied, that is, 1) questions that test basic recall of knowledge; 2) questions that require interpretation and analysis, and 3) questions that evoke problem solving and evaluation.

APPLICATION OF BLOOM'S TAXONOMY TO MCQ DESIGN

Application of Bloom's Taxonomy to MCQ Design

1 Knowledge Recall

Testing of basic recall. The weakness is that is encourages superficial learning. These are fine for formative tests and would make a good starting point for summative assessment. These are also useful for the labelling of diagrams.

Critique. Are the stem's plausible and internally consistent? Are we achieving the simple recall of factual information? I would say these were plausible as mostly functional units from other body organs.

Q1
What is the functional unit of the kidney?

e. Hepatocyte
f. Nephron (correct)
g. Juxtaglomerular apparatus
h. Crypt of Lieberkuhn

Q2
Which scientists won the Nobel Prize for their "discoveries concerning magnetic resonance imaging" in 2003? Critique. Are these plausible? I would say yes as they are Nobel winners from previous years.

e. Paul C. Lauterbur and Sir Peter Mansfield (correct)
f. Richard Axel and Linda B. Buck
g. Barry J. Marshall and J. Robin Warren
h. Sir John B. Gurdon and Shinya Yamanaka

Q3
According to the flow of urine formation through the kidneys, which of the sequences shown below is correct? Critique. Factual recall of the order of information. Useful for processes.

a. Glomerular filtration – collecting duct assimilation - tubular reabsorption – renal pelvis excretion – bladder storage
b. Tubular reabsorption – glomerular filtration - collecting duct assimilation – renal pelvis excretion – bladder storage
c. Glomerular filtration – tubular reabsorption – collecting duct assimilation – bladder storage - renal pelvis excretion
d. Glomerular filtration – tubular reabsorption – collecting duct assimilation – renal pelvis excretion – bladder storage (correct)

2   Comprehension

Aim – the explaining and comparing of knowledge rather than factual recall of items.

Q1 Which of the following answers describes the process that occurs at the Distal Tubule? Critique? Distractors must be feasible and must be of a similar length.

a. Wholesale reabsorption of water and electrolyte under the influence of hormones
b. The fine adjustment of water and electrolyte balance under the influence of antiduretic hormone
c. The fine adjustment of water and electrolyte balance under the influence of aldosterone (correct)
d. Filtration of the blood into the kidney tubule under the influence of antidiuretic hormone

3   Application

A useful question for testing calculations and the application of knowledge and comprehension of basic information related to the topic.

Q1
If heart rate is 100 BMP and stroke volume is 50ml, what is the cardiac output?

a. 5000 ml per minute (correct)
b. 200 ml per minute
c. 500 ml per minute
d. 50 ml per minute

Note, if I'd have chosen 70 BPM and SV of 60, this would have limited the usefulness of my distractors. This tests the student knowledge of the equation for Cardiac Output (CO = HR x SV) and comprehension of the units involved to derive the correct answer.

4   Analysis

When performing the haematoxylin and eosin stain (H&E), the initial step is dewaxing of the slide preparation. The specimen is then hydrated through a descending alcohol gradient. The haematoxylin is applied first followed by an alcohol –acid wash. Eosin is applied and the sample is immersed through an ascending alcohol gradient and mounted with a coverslip for viewing.

Q1
In the above process, what does the alcohol-acid wash represent?

a. Hydration
b. Dehydration
c. Illumination
d. Differentiation (correct)

Critique. Possibly, already using some weaker distractors here, it is important that all words end with the suffix *tion. We are testing the ability to analyse the process given in the above text and logically deduct the importance of the alcohol-acid wash.

Q2

Assuming that the blood vessels in the table below are the same length, which one has the greatest flow through it?

Answer	Pressure	Radius	Viscosity
A	100	1	10
b	50	2	5
c	25	4	2
D	10	6	1

Critique. A complicated question but relies on some basic calculations and knowledge of the Poiseuille equation relating blood flow to vessel diameter, pressure and viscosity.

$$\text{Flow} = \frac{\text{Change in pressure} \times \pi\, r^4}{8 \times L \times \eta}$$

Not requiring the exact calculation, but it would be reasonable to work out that D is correct where 6 to the factor of 4 multiplied by 10 is going to be the greatest answer of all.

5   Synthesis

6   Evaluation

These two latter categories are best suited to case-study type questions but I think these are very tricky to write as the answers become more subjective, and words like "most likely to occur" may introduce ambiguity. Here is a typical text book question.

Q1
A healthy 22 year old female has an exercise stress test at a local health club. A decrease in which of the following is most likely to occur in this woman's skeletal muscles during exercise?

a.   Blood flow
b.   Carbon dioxide concentration
c.   Arteriolar resistance
d.   Lactic acid concentration

Carbon dioxide and lactic acid increase. These dilate blood vessels and decrease arteriolar resistance thus increasing blood flow. Thus the only thing to decrease is arteriolar resistance.

TEST, TEST AND TEST AGAIN.

Validity and reliability testing are important components of the MCQ design process and equally apply to the design of questions to be used in surveys. For low-risk formative assessment I would not think these steps are necessary, but for high-risk and more advanced summative tests, it would be worth investing in understanding these approaches. If you are attempting more ambitions questions such as those 'evaluating' and 'synthesizing' in relation to Bloom's taxonomy, I personally would want to test the reliability of these questions in my bank of options solely because I find MCQ hard to write and I'd want to be confident I was doing the best job.

There are some very basic level questions you can ask yourselves of your MCQ test. Clearly if test questions are relatively easy, you will not have a discriminatory test. You are the best judgement of your class, and you may want to start of with less discriminatory formative tests to encourage new students for example, but introduce more rigorous formative assessment with time. Equally, a test set to a difficult standard will equally not be very discriminatory.

Central to all of this is the quality of the distractors. You want some options that will be selected – to act as a lure - otherwise your test will not discriminate. The paper by Di Battista and Kurzawa explore these ideas in detail.

- Are there questions that everyone is getting wrong? You can form a difficulty index where 1 = all wrong to 0 = all correct.
- Are there distractors that are either being selected by everyone or not at all?
- Do you think your test has been very discriminatory – do you have a good range of marks?

In their paper, the authors found of a sample of MCQ tests within their institution, 30% were not discriminating, expressed as a discrimination coefficient of <0.2 as their benchmark for their courses.

WHAT DOES VALIDITY MEAN?

How valid is the overall test in accomplishing the task? How valid is your survey question in understanding student engagement? How valid is the overall MCQ test in being a robust assessment for students?

You will test the validity by piloting the test with colleagues and/or students? Questions that are anomalies or that are not covered in the learning outcomes of the teaching sessions can be simply plucked out. You can judge whether the test overall is delivering what you expect – or has it been to easy? Have there been too may 'recall' questions?

WHAT DOES RELIABILITY MEAN?

How consistent are your questions in measuring the learning outcomes? How reliable are the distractors? A simple test called Cronbach's alpha can be applied to measure the reliability of the MCQ questions, or questions within surveys. It measures the internal consistency of an item within a group, and the calculation incorporates the variability of the responses.

A result less than 0.7 means there is an item that is poor quality within the survey or questions. In SPSS or Minitab you can conduct a factor analysis to get a further indication of which questions are inconsistent and need redesigning.

In the Battista and Kurzawa paper that analysed 16 MCQ tests within an institution, once adjusted for numbers of questions, a small number of tests were <0.7 , although not disasterously low in my opinion.

By simple calculations they also showed that over one third of all distractors were flawed because they were chosen by less than 5% of the examinees in total.

CALCULATIONS

Some software, e.g. automated optical marking systems, often provide reliability data as standard. Otherwise, a simple calculation test can be applied. Reliability and validity helps you understand how appropriate your questions were, and to identify any troublesome spots or questions that were not answered well, to easy, to hard, or that might reveal a flaw with the question.

Essentially you gather your question responses and code them, for example students answering distractor a = 1, b = 2, c = 3 etc. You input into SPSS or Minitab according to this video (https://www.youtube.com/watch?v=fmF0jZtZCRQ).

- Input your data according to the worksheet lower panel.
- Label columns Q1-Q6
- Statistics / multivariate / item analysis
- Select variables Q1-Q6

- Select 'Cronbach's alpha' and any other results you wish to display
- Run
- Results viewed in top panel
- Results display OVERALL value plus INDIVIDUAL ITEMS
- Any result <0.7 is problematic
- (As you see in the figure above, the entire survey is problematic!!)

Further reading can be found on the King's College website (KCL 2015)
[http://www.kcl.ac.uk/ioppn/depts/biostatistics/SAS/faqs9.aspx#a9_9].

In a further example in Allen's paper (2004) the authors are interested in how critical thinking skills are assessed in nursing students. Do tests of critical thinking correlate with actual ability? They have come up with a 17 dimension instrument to look at critical thinking and have assessed the robustness of this with a group of students. Their Cronbach alpha was between 0.7 and 0.8 showing a good level of correlation between the instrument and actual ability. This is a typical application of Cronbach alpha in the literature.

Further reading:
Using SPSS for a reliability test. Available: https://www.youtube.com/watch?v=fmF0jZtZCRQ

5. ACCESSIBILITY AND INCLUSIVE DESIGN

MCQ CAN BE ACCESSIBLE AND INCLUSIVE – BUT YOU HAVE TO DO THINGS.

We absolutely have to make sure that we comply with legal acts against discrimination, and also comply with our university regulations.

- The Disability Discrimination Act 1995 (as subsequently amended)
- The Disability Rights Commission Act 1999
- The Special Educational Needs and Disability Act 2001

Many university regulations specify circumstances by which a student can be supported, but my feeling is, this support does not go far enough to inform the academic teams of how to best put measures in place. With cuts and possible removal of the Disabled Students' Allowance (DSA) threatened, this will decentralise disability support, and it is possible that academic staff will become much more responsible for their own work.

So as academics, our challenge is to ensure we are inclusive in everything we teach, our learning resources, our use of technology, and certainly in our assessments. I find, once you understand and can apply a few basic principles, this can be extrapolated across all of our practice. In fact, this is just plane GOOD practice.

PRINCIPLES OF COGNITIVE LOADING

The educational psychology theories developed by William Mayer and others [https://en.wikipedia.org/wiki/Cognitive_load] explain the idea that our brains can only do so much at once. Cognitive loading is the ability to undertake a small number of sequential tasks at one go. In designing multimedia resources for example, it is advised that no more than 2 components exist – text and picture, image and sound, and as soon as you start applying additional layers, the learning ceases to be effective as the mental effort increases. Likewise, get it right, and you can have a more effective learning tool.

Similar principles are involved in assessment design, particularly MCQ. Imagine these scenarios.

- You have deployed a paper-based MCQ test and you require your students to apply their answers on an optical mark card.
- You have deployed a paper-based test that combines MCQ questions with questions such as filling in the gaps.
- Your MCQ test incorporates negative distractors or 'A, AB or none of the above' type question formats.

In each of these scenarios, you are building in additional and extraneous steps. Some students may find the transfer of answers from the exam paper to the optical mark card a challenge. In the second scenario, these additional processing steps can again be tricky. Filling in the gaps can doubly disadvantage dyslexic students who may not have accurate spelling, especially if your test is to be auto-marked by tests built into a LMS (learning management system) such as Blackboard or Moodle for example where you have inputted the answer and presumably 'correct' spelling to the system.

There has been a wealth of reliability testing that now provides a substantial evidence base that negative marking, negative distractors and that 'A, AB' type choice questions should be avoided.

CHECK OUT YOUR UNIVERSITY REGULATIONS

Many university regulations state that students must have assessment practice prior to the main test. You certainly would not want to launch a high stakes test on students without them practicing first and identifying any possible issues they might have. I personally think many education test and examination processes favour those with the ability to do the test. I hope as part of the DSA changes, that design becomes more inclusive at the level of the academic, and assessment options become more flexible.

UNIVERSAL DESIGN OF INSTRUCTION (UDI)

This is a set of principles or framework widely adopted in US schools and colleges, and also employed in UK higher education institutions. I find it a useful guide to think about how my academic practice can be accessible and inclusive to everyone. I think this is just good practice.

As Burgstahler explains (2015), you design in from the start a broad range of options for all learners. It is very difficult to fit this retrospectively. Imagine if you have spent hours digitizing learning materials and creating multimedia content and you have not thought through these principles. You will not be offering inclusive opportunities to learn.

Assessment and feedback principles

- Provide regular formative feedback (peer-peer, or self-evaluation)
- Provide clear expectations to students on what to expect and when
- Multiple ways to assess knowledge in a module – don't just rely on MCQ
- Provide the student with practice to gain familiarity with the test
- Offer online tests or printed versions
- Enable variable font format and size
- Ensure test is provided on variable / customisable colour background

Thompson et al (2002) provide another level of detail, and in the US, it is an obligation to ensure these standards of practice are met. These apply specifically to large-scale assessments. Clarity is key, as is cultural inclusion and maximum comprehensibility.

Large-scale assessment principles

- Inclusive assessment population
- Precisely defined constructs
- Accessible, non-biased items
- Amendable to accommodations
- Simple, clear, and intuitive instructions and procedures
- Maximum readability and comprehensibility
- Maximum legibility

For a UK perspective we would need to use the following resources from Jisc, CETIS and TechDis. Some of this information is now archived and unfortunately some of the hyperlinks may not be operational.

Further reading:

Jisc (2015) Making assessments accessible. Available: https://www.jisc.ac.uk/guides/making-assessments-accessible

CETIS (2015) Learning design and pedagogy. Available: http://wiki.cetis.ac.uk/Learning_Design_and_Pedagogy

TechDis (2015) Link to new site. Available: https://www.jisc.ac.uk/website/legacy/techdis

6. CREATIVE APPROACHES

WHY ARE YOU WRITING THE QUESTIONS?

MCQ are challenging and time consuming to produce. One approach is to produce a question pool or question bank within your teaching teams. Another approach is to engage students as co-producers. In my experience they are fantastic at coming up with 'distractors' and you get an insight into what they are thinking. This serves three purposes:

- You get insight into what students are thinking
- Students can use this as a reflective exercise – to provide an MCQ at the end of a lecture
- Students build their assessment literacies and learn about the MCQ process

So for several years I've got students to generate MCQ which I then edit and input as formative tests onto web software (HotPotatoes) or the university learning management system (LMS). A quick edit and construction of 20 questions takes about an hour. In the LMS such as Blackboard or Moodle, these can provide basic answers, give multiple attempts, or be more instructional and provide explanations of answers.

Other UK academics are using the student as co-producer approach really effectively, and PeerWise [https://peerwise.cs.auckland.ac.nz] is an Australian software system that enables students to entirely oversee the process themselves.

Debbie Bevitt and Nick Morriss, Newcastle

http://teaching.ncl.ac.uk/casestudies/storedcasestudies/name,120062,en.html

Students inputted their own MCQ onto a database and used the questions for revision. The submission of questions can become quite competitive, and therefore self-editing. As I recall, Debbie had minimum intervention other that checking the quality and accuracy of the questions at first.

Heather McQueen

http://www.research.ed.ac.uk/portal/en/publications/studentgenerated-content(2428b5b7-7b3a-4815-960d-5c66628cbe4d).html

Heather took a similar approach and conducted an evaluation across 3 UK institutions. The interventions had significant impact on student exam performance with minimal input from the academic teams.

7. ASSESSMENT DESIGN CYCLE

FROM MCQ HELL TO HEAVEN.

Here is an idea for an MCQ design cycle.

1) Gather your programme team, or faculty colleagues teaching similar subjects.
2) Discuss the concept of a question bank and brainstorm questions together. (Maybe have a 2-day off site question sprint).
3) Consider allowing students to generate MCQ as part of low stakes formative assessment.
4) Apply UDI principles – design for inclusivity up front.
5) Apply the principles for good quality questions – perfect your writing of questions plus distractors.
6) Evaluate your tests – apply some basic calculations and the statistical tests for reliability.
7) Work up to providing rigorous, discriminating, well-written and inclusive MCQ testing!

REFERENCES

Allen GD, Rubenfeld MG, Scheffer BK (2004). Reliability of assessment of critical thinking. J Prof Nurs. 20(1),15-22.

Brame C (2015). Vanderbilt University: Writing good multiple choice questions. Available: https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/

Burgstahler S (2015). Equal access – universal design instruction. Available: http://www.washington.edu/doit/equal-access-universal-design-instruction

CETIS (2015). Learning design and pedagogy. Available: http://wiki.cetis.ac.uk/Learning_Design_and_Pedagogy

Di Battista D & Kurzawa L (2011). Examination of the Quality of Multiple-Choice Items on Classroom Tests. Canadian Journal for the Scholarship of Teaching and Learning, 2(2), 4.

Holsgrove G (2001). THE article: What is the right choice? Available: https://www.timeshighereducation.com/news/what-is-the-right-choice/159581.article

Jisc (2015). Assessment for learning. Available:
http://jiscdesignstudio.pbworks.com/w/page/52947115/Assessment%20for%20Learning

Jisc (2015). Making assessments accessible. Available: https://www.jisc.ac.uk/guides/making-assessments-accessible

KCL Statistics Advice Centre (2015). Questionnaire development. Available:
http://www.kcl.ac.uk/ioppn/depts/biostatistics/SAS/faqs9.aspx#a9_9

Maber J, Booth A, Hamburg L and Wassall T (2015). Leeds Beckett. MCQ help – writing good questions.
Available:
https://my.leedsbeckett.ac.uk/bbcswebdav/institution/Health_and_Social_Sciences/ltu/areas/assessment/mcq/design.html

Rattray J & Jones M C (2007). Essential elements of questionnaire design and development. Journal of clinical nursing, 16(2), 234-243.

TechDis (2015). Link to new site. Available: https://www.jisc.ac.uk/website/legacy/techdis

Thompson SJ, Johnstone CJ and Thurlow ML (2002). Universal design applied to large scale assessments.
Available: http://www.cehd.umn.edu/nceo/onlinepubs/Synthesis44.html

Seven steps to MCQ Heaven
Dr Vivien Rolfe
Version 1: 3rd December 2015