

Multiple choice questions

Multiple choice questions (MCQs) perform a unique and important role in summative assessment¹ and make a valuable contribution to assessment programmes in a wide range of educational settings.

There is considerable support for the use of MCQ assessments both internationally and across different professions. For example, in Australia an MCQ test forms the first part of the [Special Tertiary Admissions Test](#) that is used alongside other sources of evidence to determine entry to certain undergraduate programmes, or for students without the national standard entry criteria. Medical schools across the world also use MCQ tests to assess student doctors' ability to apply their knowledge of clinical content and practice (see McHarg, *et al.*, 2005); and MCQ tests form an essential part of initial [pilot licensing and subsequent Type Rating assessments](#) for commercial pilots. There is also substantial research literature validating the use of MCQ tests (see for example, Haladyna, 2004) and supporting the reliability and validity of test outcomes (see for example, van der Vleuten *et al.*, 1996).

MCQs are considered to have many advantages, both in themselves and over other forms of assessment. These include:

- **Breadth:** MCQ assessments can test a broad range of content in a relatively short period of time, and are therefore more time-efficient than other means of assessment (as long as the focus is on valid subject content and not trivial content).
- **Less predictability:** Being able to sample widely across knowledge domains means that MCQ assessments are typically less predictable than other forms of assessment. This encourages students to learn the whole specification and gain a broader appreciation of the subject rather than just focusing on material considered likely to appear in a test. The literature relating to the use of MCQ assessments to measure professional competency (for example in medical education) suggests that broad sampling and a lack of predictability can foster a deeper, less superficial approach to learning (for example see Newble and Jaeger, 1983). This is particularly the case when MCQs are designed to assess higher order skills such as the application of knowledge.
- **Relevance:** Selecting questions from a representative sample of content areas and mapping the coverage of specification content over time is the basis of *content-valid test-score interpretations*; that is, knowing with confidence that a difference in test score reflects a difference in candidate ability in that knowledge area.
- **Adaptability:** Question writers can adapt MCQs to measure lower order processes such as knowledge recall and comprehension, and higher order processes such as application, interpretation, synthesis, and analysis. An MCQ test with varying demands can provide opportunities to stretch and challenge the most able students.
- **Reliability:** Objective marking of MCQ assessments typically produces highly reliable scores. Well constructed MCQ assessments can also lead to good test-retest reliability

¹ The goal of summative assessment is to *evaluate* student learning at the end of a component of study by comparing it to a known standard or benchmark. This is in contrast to formative assessment, the goal of which is to *monitor* student learning and provide ongoing feedback to improve teaching and learning.

(same test, different occasions) and parallel-forms reliability (alternative versions of a test, same occasion) (for example, see Nunnally and Bernstein, 1994).

- **Insight:** Classifying questions by topic area means diagnostic and meaningful sub-scores can be calculated. These can be used by teachers to inform their classroom strategies.
- **Differentiation between students:** MCQ assessments make use of the full mark range to a greater extent than other forms of assessment, such as essay questions. MCQ assessments that achieve a good spread of candidates along the score continuum enhance our ability to differentiate between students at various levels of attainment.
- **Performance evaluation:** MCQ tests lend themselves to a wide range of psychometric analyses that can be used to evaluate test performance and inform the development of future tests. This allows test writers to assess the quality of their questions and to understand how students respond to different questions. If more sophisticated methods, such as [Rasch analysis](#), are used, test writers can construct MCQ tests of known difficulty, which ensures comparability of standards over time.

The advantages outlined above are not inherent to MCQ assessments. The quality of the assessment lies in the quality of question writing, sampling and test construction processes. Inadvertent question errors (such as more than one correct response), poor quality distractors², and a focus on trivial subject content, are just some of the factors that can undermine the usefulness and reliability of MCQ tests. Indeed, writing good MCQs is not an easy task; the development of quality questions of varying demand is labour-intensive and challenging. It also requires a good understanding of a number of assessment concepts (such as question difficulty, demand, and construct irrelevance³). It is therefore essential that question writers receive training to ensure they adopt best practice in terms of question writing and test construction, which builds on the tacit understanding of question writing that examiners develop over the course of their professional experience.

The scholarly literature provides a wealth of evidence that the use of well-constructed MCQ assessments is defensible, fair, and beneficial to candidates and those who use test results for selection or other purposes. GCSEs and A-levels are designed around a collection of multiple sources of evidence that assess candidates' grasp of the content covered by each specification. MCQ assessments can provide one of these sources of evidence, and can do so in a way that is appropriate, robust and efficient.

December 2012

² Incorrect answers provided as options in a multiple choice tests, which must appear plausible to candidates.

³ The extent to which test scores are influenced by factors (such as mode of presentation or response) that are not relevant to the knowledge/skill the test is intended to measure.

References

- Haladyna, T.M. (2004) (3rd edition). *Developing and validating multiple-choice test items*. London: Routledge.
- McHarg, J., Bradley, P., Chamberlain, S., Ricketts, C., Searle, J. & McLachlan, J. (2005). [Assessment of Progress Tests](#). *Medical Education*, 39, 221-227.
- Newble, D.I. & Jaeger, K. (1983). [The effect of assessments and examinations on the learning of medical students](#). *Medical Education*, 17, 65–71.
- Nunnally, J. & Bernstein, I.H. (1994). *Psychometric Theory*. London: McGraw-Hill, Inc.
- Van der Vleuten, C.P.M., Verwijnen, G.M. & Wijnen, H.F.W. (1996). [Fifteen years of experience with progress testing in a problem-based learning curriculum](#). *Medical Teacher*, 18, 103–9.

Suggested reading

- Case, S.M. & Swanson, D.B. (3rd ed. 2002). *Constructing written test questions for the basic and clinical sciences*. Philadelphia: National Board of Medical Examiners. Available from: http://www.nbme.org/pdf/itemwriting_2003/2003iwqwhole.pdf
- Murphy, R., Greatbatch, D., Wilmut, D., Tolley, H., Mallia, C. & Hamer, J. (2008). *Investigating ways to improve written GCSE examinations: Final report to QCA from the Centre for Developing and Evaluating Lifelong Learning, University of Nottingham*. London: QCA. Available from: http://www2.ofqual.gov.uk/files/Improving_GCSEs_CDELL.pdf
- Pollitt, A., Ahmed, A., Baird, J., Tognolini, J. and Davidson, M. (2008). *Improving the quality of GCSE assessment*. London: QCA. Available from: http://www2.ofqual.gov.uk/files/Improving_the_Quality_of_GCSE_Assessment_final.pdf