Alternative assessment tools

Margaret V. Root Kustritz

University of Minnesota, College of Veterinary Medicine, St. Paul, MN

Introduction

Assessment is an essential component of teaching. Assessment may be formative, helping students identify gaps in their understanding and building competence, or may be summative, permitting a final evaluation of competence for assignment of a grade or to permit students to advance in a program or career. Assessment methods must accurately reflect the learning objectives and activities that preceded their implementation, must be a true test of student learning rather than an exercise in test-taking skills or ability to ascertain what the instructor considers most important, and must permit the instructor to provide consistent and timely feedback.

Keywords: Veterinary education, assessment, multiple choice questions

Multiple choice questions (MCQs) are written with a stem and several possible answers, or foils. The following are basic rules of MCQ construction:[1-6]

- Put the central idea in the stem and do not repeat it in the foils
- Avoid negative construction in the stem ("Which of the following is not correct…")
- Use proper grammar and ensure the correct foil is not obvious because of grammar clues
- Do not use the terms "always" or "never" in foils
- Place foils in logical or numerical order
- Make sure foils are parallel with respect to grammatical structure, length, and complexity
- Use "none of the above" with caution; do not use "all of the above" as a foil
- Three foils is the best number, based on research evaluating number of plausible foils
- To write plausible foils, consider most common student misconceptions or create a short answer question one year and use student responses to create foils for the next year
- Do not worry excessively about distribution of answers; research suggests that students pay no attention to pattern of answers as a test-taking technique
- Use humor sparingly

In the mid-20th century, Bloom and colleagues developed a hierarchy of educational objectives in the cognitive domain of teaching and learning that has come to be known as Bloom's Taxonomy. At the most basic level is knowledge, ability to remember previously learned material. In increasing order of complexity are comprehension, ability to grasp the meaning of material; application, ability to use information in new situations; analysis, ability to break material down into its component parts; synthesis, ability to re-order material in new ways; and evaluation, ability to compare value of material with other criteria. In veterinary medicine, some material clearly is best assessed at the knowledge level but most information must be assessed at the higher levels of analysis, synthesis, and evaluation.

Traditionally, people have used MCQs, true-false questions, and matching questions to test lower level learning and have used short answer and essay questions to assess higher level learning. In one study evaluating use of MCQs for assessment of learning across Bloom's taxonomy, of 2913 questions from a test databank 47.3% were written to test knowledge, 24.8% comprehension, 21.8% application, and only 6.5% analysis.[1] Multiple choice questions are easy to grade and a large databank can be created from which different questions can be drawn with each iteration of the examination to limit potential cheating. In one study MCQ testing was superior to assessment by peers, by working through cases or by assessment of a student portfolio, perhaps because it required students to put in significant effort committing the information to longer-term memory.[7] The primary concern expressed about MCQs is the perception that instructors purposefully write "trick" questions.[8] Examples of improper MCQ

construction that may be perceived as "trick" questions by students include introduction of trivial content, too little distinction between foils, multiple correct answers, and highly ambiguous foils.[9] Questions must be carefully worded to permit evaluation of higher order learning and student anxiety may prevent their showing their best performance with MCQ testing compared to more open-ended question types.[10,11]

Multiple choice questions can be written in a manner that addresses higher level learning. To create MCQs assessing comprehension, students can be asked to interpret information. Knowledge of the material is assumed and required for correct interpretation. To assess application, students are expected to know and understand background material and are asked to solve a problem or demonstrate a correct method or procedure. Analysis is assessed by requiring recognition of facts versus inferences or faults in reasoning and requires proof that the student understands and can work with the underlying structure of the material provided. Synthesis is difficult to assess in a multiple choice question because it requires the student to demonstrate creative ability. To assess the highest level of learning, evaluation, students are asked to judge the value or validity of material presented.[12]

Enhanced multiple choice questions are variations on traditional MCQs that more readily permit assessment of higher level learning. Enhanced multiple choice question types include content-dependent MCQs, demonstrated-reasoning MCQs, and confidence-level MCQs.[2,13-16] Many of these variations break the rules described above for multiple choice testing and are more difficult to grade quickly and consistently. However, they do permit assessment of higher level learning and assessment of other parameters, such as ability of the student to work through ambiguity and demonstrate confidence in their knowledge and skills.

Content-dependent MCQs are constructed with a body of information provided from which one or more MCQs are drawn. The information provided may be a case study, a chart or diagram, an image of a lesion, or any other information that students can analyze or evaluate.[2] Be cautious of unconsciously creating questions that are interdependent.

Demonstrated-reasoning MCQs are standard MCQs associated with an opportunity for students to demonstrate the reason behind their answer(s).[14-16] Students may be given points for explaining why their chosen answer is correct and why the other foils are not correct. One example of this is giving a given number of points for identifying the correct foil and fewer points for explanation by the student of the correct answer; this may permit instructors to give students credit for their knowledge and reasoning even if they do not choose the correct foil. A final variant of this type of question is conscious use of more than one correct foil with students getting credit for identifying all possible correct answers and getting credit for explaining the rationale for their answers.

Confidence-level MCQs are the most complex variant.[13] Six to eight foils are presented for each question, with several clear possibilities, several possibilities student may choose if unsure, and one option consciously entitled "I don't know." (Figure 1). The goal is to identify not just student learning but also how well informed that student felt in making their decision. Any student answering "b" ," c" or "e" obviously is confident and uninformed. Any student answering "d" is confident and misinformed as they clearly do not understand that the answer must be a single whole number. Any student answering "f" is not confident and not well informed. This provides interesting information for the instructor regarding student competence and confidence. One concern is that students must know they may get some credit for answering "I don't know" or they will never choose that answer. For this reason, this type of question is best used for formative, not summative, assessment.

Script concordance questions are a type of MCQ requiring students to demonstrate higher level learning. Information is provided and hypotheses generated. Students are offered diagnostic test results or other information and asked whether those results make the hypothesis more or less likely (Figure 2).[17-19] In human medicine, script concordance testing results have been shown to correlate well with performance on clinics and with board examination testing when used with medical students, residents, and practicing physicians.[20,21] It is recommended that students be permitted to explain the rationale behind their answer. More than one alternative may be considered correct, recognizing lack of confidence by students in absolutely ruling in or ruling out any alternative and variation in student

backgrounds and the information those students used to create the "scripts" they have stored in long-term memory about those clinical situations.

The use of these alternative testing techniques requires explanation of varying technique to the student so they can demonstrate their knowledge and skills without anxiety about the testing methodology. Students may benefit from use of the assessment technique in a formative manner before it is used for summative assessment, to make it clear to the instructor how well students understand the testing technique and to help the students understand how the questions will be scored. For example, when the author first used script concordance testing, she presented the students weekly with a question of this type with the correct answer available to them after their initial attempt. This gave them a comfort level with this type of question before the summative assessment late in the semester and had the added benefit of forcing them to routinely review the material well before the examination.

Learning objectives are the foundation of any course and should guide content provided, activities and assignments used in the course, and assessments. Student preference should not prevent instructors from using inventive assessment techniques to help them best capture assessment of their chosen learning objectives.

Figure 1. Confidence-level multiple choice question

The correct answer for the mathematical equation 1+2 is:
  a. 3
  b. 5
  c. 256
  d. a or c
  e. a and b
  f. I don't know

Figure 2. Script concordance question

This type of examination is called a script concordance test. It evaluates not just your knowledge but your ability to use that knowledge to make clinical decisions. Read the scenario, think about likely hypotheses and then, for each hypothesis and finding presented, choose likelihood using the scale below.

-2 = they hypothesis is virtually eliminated from consideration
-1 = the hypothesis becomes less likely
0 = the information has no effect on the hypothesis
+1 = the hypothesis is becoming more probable
+2 = it can only be this hypothesis

Scenario: You are presented with a 6-day-old West Highland White terrier pup for the complaint of persistent crying and inability to nurse.

IF YOU WERE THINKING OF…	AND THEN YOU FIND…	THE HYPOTHESIS BECOMES…				
Canine herpesvirus	Decreased red blood cell number	-2	-1	0	+1	+2
Dehydration	Pale pink muzzle and groin	-2	-1	0	+1	+2
Septicemia	Decreased blood glucose	-2	-1	0	+1	+2

References

1. Masters JC, Hulsmeyer BS, Pike ME, et al: Assessment of multiple-choice questions in selected test banks accompanying textbooks used in nursing education. J Nurs Educ 2001;40:25-32.
2. Haladyna TM, Downing SM, Rodriguez MC: A review of multiple-choice item-writing guidelines for classroom assessment. Appl Meas Educ 2002;15:309-334.
3. Briggs DC, Alonzo AC, Schwab C, et al:: Diagnostic assessment with ordered multiple-choice items. Educ Assess 2006;11:33-63.
4. Taylor AK: Violating conventional wisdom in multiple choice test construction. Coll Stud J 2005;39:141-148.
5. McKeachie WJ: Teaching tips: strategies, research, and theory for college and university teachers. Lexington (MA): DC Heath and Co.; 1994. p. 78.
6. McMillan JH: Classroom assessment: principles and practices for effective instruction. Boston: Allyn and Bacon; 2001. p. 150.
7. Struyven K, Dochy F, Janssens S, et al: The overall effects of end-of-course assessment on student performance: a comparison between multiple choice testing, peer assessment, case-based assessment and portfolio assessment. Stud Educ Eval 2006;32:202-222.
8. Dale VHM, Wieland B, Prikelbauer B, et al: Value and benefits of open-book examinations as assessment for deep learning in a post-graduate animal health course. J Vet Med Educ 2009;36:403-410.
9. Roberts DM: An empirical study on the nature of trick test questions. J Educ Meas 1993;30:331-344.
10. Turner H, Williams RL: Vocabulary development and performance on multiple-choice exams in large entry-level courses. J Coll Read Learn 2007;37:64-81.
11. Birenbaum M, Feldman RA: Relationships between learning patterns and attitudes towards two assessment formats. Educ Res 1998;40:90-98.
12. http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html#C2, accessed 02-28-10.
13. Swartz SM: Acceptance and accuracy of multiple choice, confidence-level, and essay question formats for graduate students. J Educ Business 2006;81:215-220.
14. Draper SW: Catalytic assessment: understanding how MCQs and EVS can foster deep learning. Brit J Educ Tech 2009;40:285-293.
15. Ennever F: More than multiple-choice. Science Teacher 2006;73:54-56.
16. Pomplun M, Omar MH. Multiple-mark items: an alternative objective item format? Educ Psych Meas 1997;57:949-962.
17. Smith CS: A developmental approach to evaluating competence in clinical reasoning. J Vet Med Educ 2008;35:375-381.
18. Charlin B, Tardif J, Boshuizen HPA: Scripts and medical diagnostic knowledge: theory and application for clinical reasoning instruction and research. Acad Med 2000;75:182-190.
19. Charlin B, van der Vleuten C: Standardized assessment of reasoning in contexts of uncertainty: the script concordance approach. Eval Health Prof 2004;27:304-319.
20. Norcini JJ, Shea JA, Day SC: The use of aggregate scoring for a recertification examination. Eval Health Prof 1990;13:241-251.
21. Silbert L, Darmoni SJ, Dahamna B, et al: On line clinical reasoning assessment with script concordance test in urology: results of a French pilot study. BMC Med Educ 2006;6:45-53.