

Assessing the Application of Knowledge in Athletic Training: Extended-Matching Questions

Jim F. Schilling, PhD, ATC, CSCS
Northern Arizona University, Phoenix

Context: The accuracy of summative assessment scoring and discriminating the level of knowledge in subject matter is critical in fairness to learners in health care professional programs and to ensure stakeholders of competent providers. An evidence-based approach to determine examination quality for the assessment of applied knowledge is imperative.

Objective: To present a written assessment question format that is simple to construct, can be scored electronically, and will accurately assess the application of knowledge in examinations during the classroom curriculum and for high-stakes examinations.

Background: Traditionally, free-response written assessments have been administered to assess the application of knowledge, for example, short answer or essay questions. Selected-response questions such as a carefully constructed multiple-choice questions or extended-matching questions (EMQ) are reliable and valid options to assess the depth of understanding in subject matter.

Description: Extended matching is a selected-response question format used to assess the application of knowledge. It consists of items consistent with a common theme and a large list of homogenous options providing distractors and a correct answer.

Clinical Advantage: Extended-matching questions have demonstrated good psychomotor measures and can assess the application of knowledge over a large number of themes in 1 examination. They are simple to create, minimize cueing, and are scored objectively.

Conclusions: Both well-written multiple-choice questions and EMQs can be used to assess the application of knowledge level of cognition. The EMQ is easier to write with less chance of cueing or guessing the correct option and is optimal for evaluation and treatment scenario type questions.

Key Words: Multiple choice, assessment, application

Dr Schilling is currently Associate Clinical Professor at Northern Arizona University. Please address all correspondence to Jim F. Schilling, PhD, ATC, CSCS, Northern Arizona University, 435 North 5th Street, Phoenix, AZ 85004. jim.schilling@nau.edu.

Full Citation:

Schilling JF. Assessing the application of knowledge in athletic training: extended-matching questions. *Athl Train Educ J*. 2019;14(1):80–89.

Assessing the Application of Knowledge in Athletic Training: Extended-Matching Questions

Jim F. Schilling, PhD, ATC, CSCS

KEY POINTS

- Depth of knowledge has traditionally been assessed using open-ended questions, but they are difficult to score and cover limited material. Multiple-choice questions could be used but they require extensive scrutiny to create and are vulnerable to cuing and assessment of factual recall.
- Extended-matching questions are created in a simple, systematic fashion, easily linked to learning objectives, and demonstrate strong reliability and validity.
- Curriculum and high-stakes examinations can benefit by using extended-matching questions to assess the application of knowledge in an accurate and efficient manner.

An accurate summative assessment of knowledge and skills in professional health care education programs is crucial to the fairness of the examination for learners and assurance of competent clinicians for future patients. In addition to classroom curricula, summative assessments are playing an increasing role with professional regulating bodies assessing the competence of graduates on high-stakes examinations such as licensure or certification assessments. There are a number of factors that can influence the quality of questions that formulate an examination. Measures such as reliability and validity are used to determine their accuracy. A reliable question will measure knowledge and skills in a precise and reproducible manner.¹ Learners vary in acquiring skills and learning subject matter, so there is a need for a broad sampling of questions and adequate examination length to ensure acceptable reliability for an assessment.² To be valid, the examination must not only demonstrate the learners have acquired appropriate content, but they must also have the ability to make distinctions among learners at differing levels of knowledge.³ Additional measures are used with selected-response questions (multiple choice and extended matching), such as difficulty index and discriminatory power, which demonstrate the degree of separation between levels of performance.⁴ *Difficulty index* is defined as the proportion of examinees who select the correct option, and the *discriminatory power* is a view of the extent a more knowledgeable learner will select the correct option over a less knowledgeable learner.⁴ If the difficulty index is either too high or too low, it will have a negative impact on the discriminatory power.⁴ Another factor to consider is *educational impact*, a concept where learners focus their study habits according to what they believe will be the type of question presented on the examination.⁵ This concept was demonstrated when students used a surface learning approach when preparing for a multiple-choice question (MCQ) examination and a deep learning approach for essay questions.⁶ Since the assessment often drives learning, the content of an examination should be constructed so that the assessment tools match the learning objectives or specific knowledge and skill competencies, a process referred to as *blueprinting*.² Even more factors include ease of construction or feasibility and resourcing needs, such as the materials, money, staff, and other assets necessary to conduct the examination effectively.⁷

A standard for the examination also needs to be established, meaning a minimum score to be considered competent to practice, referred to as *criterion referencing*.² There are multiple criteria to consider with question format when constructing written assessments, all of which contribute to their accuracy and effectiveness.

The depth of knowledge a specific question format is meant to assess is another area of consideration. Models were created that have categorized specific levels of knowledge. One such model is called Bloom's taxonomy model that was revised in 2001 to include the following 6 levels of cognition: remember, understand, apply, analyze, evaluate, and create.⁸ The apply level demands the ability to execute and implement knowledge, which is a reflection of critical thinking.⁸ This is a significant concept in that athletic training is a profession that provides a type of health care service requiring the ability to apply knowledge to be a competent provider. Similar to Bloom's taxonomy model, but specific to the assessment of clinical competence, is a model created by Dr George Miller which consists of 4 facets beginning with knowledge (knows), advancing to competence (knows how), then to performance (shows how), and finally action (does).⁹ The *knows how* level of Miller's model relates to a clinician's ability to gather information, analyze it, interpret it, and form diagnostic and treatment strategies.⁹ How to assess the application level of Bloom's taxonomy and the knows how level of Miller's model for clinical competencies most effectively and efficiently are the premise of this review. Examples of the application of knowledge are in the way of decision-making, problem-solving, and clinical reasoning skills. Clinical reasoning is a specific type of decision making used in clinical practice to evaluate and treat conditions. Assessment of this level of knowledge provides feedback as to the depth of understanding a learner has in a particular area of subject matter and not simply factual recall. The know how facet of clinical competence needs examinations that can assess skills in specific contexts, for example, the presentation of clinical scenarios or vignettes.⁹ Written examinations are being used in medical education to assess the knowledge and the ability to apply knowledge levels of cognition using questions presented as authentic clinical scenarios.¹⁰ In other words, they assess the know and know how facets of Miller's model or the remember, understand, and apply levels of Bloom's revised taxonomy model.

Written examination questions come in multiple formats with the intent of assessing various levels of knowledge. Examples of written examination question formats are true or false, multiple choice, short answer, essay, key feature, and extended-matching questions (EMQs).⁵ Free-response questions such as essay or short answer questions (SAQ) have historically been used to assess higher levels of cognition, such as applied knowledge.⁵ Several disadvantages have been experienced, such as requiring learners to assume what the instructor wants, what will be penalized, what will be rewarded, and the amount of detail expected, all of which

reduces the validity and reliability of scores.¹¹ Free-response questions are also time consuming for the instructor to review and subjective in scoring, which again reduces validity and reliability of the assessment.¹¹ They are also time consuming for the learner, which limits the number of questions, resulting in examinations with less content coverage and poor scoring reliability due to a low sample size.¹¹ A study¹² comparing the free-response format to MCQs did not find free-response questions to be superior in high-stakes summative assessments of higher cognitive processing, and they also lack the ability to be scored electronically. Selected-response questions are objectively scored and have the ability to be electronically marked. To assess a clinician's ability to analyze clinical situations, essay questions and oral examinations have been replaced by MCQs because of problems with consistency in grading, limited subject matter sampling, and overall cost.¹³ Quality MCQs are difficult to create due primarily to the need in constructing plausible distractors (options other than the correct answer) with the goal of minimizing cueing.¹ Cueing is an examination question writing flaw that tips the learner on what is or what is not the correct answer.¹⁴ With MCQs having limited options (5 or less), there is an increased risk of successfully guessing.¹ They are often used to assess the recall of isolated facts and not the ability to apply knowledge, such as in the way of problem-solving or clinical reasoning competency skills.¹⁵ Critiquing summative examinations in undergraduate nursing has exposed many flaws in question writing, such as increased cueing and the risk of guessing correctly.¹⁴ A MCQ can easily be transitioned to an EMQ by increasing the number of options.¹⁵

The EMQ format is being used in high-stakes licensing and certification examinations nationally and internationally,¹⁶ primarily in medical education.¹¹ Extended matching is a question that can be administered electronically, produces objective measureable outcomes, and whose format reduces writing flaws.¹¹ They were found to have good reliability^{3,17–22} and validity.^{17,18,23} They have also demonstrated good discriminatory power^{3,20,24,25} and an examination that minimizes cueing,³ shows good feasibility,^{3,17,18} cost effectiveness,¹⁸ and ability to assess applied knowledge.^{3,7,20,26,27} The purpose of this review is to share a type of written assessment question format that can be used in the classroom or on high-stakes examinations that effectively assesses the application of knowledge in athletic training subject matter in an efficient and accurate manner. The article will first describe the format of EMQs in general and then demonstrate how this type of question format can be used in athletic training education followed by a discussion supplying evidence to support its use in the assessment of applied knowledge.

EXTENDED-MATCHING QUESTION FORMAT

There are 2 types of EMQ formats; one is referred to as the R type, where only 1 option is the correct answer, and the other is the N type, where you have up to 5 correct answers.²⁸ The R type is most commonly used and will be the type emphasized in this review. With an EMQ, the number of options are expanded beyond the usual 5 for a MCQ to a possible 26 (letters in the alphabet), which presents a situation where eliminating the incorrect options becomes difficult.¹⁵ The EMQs used in medical education are short scenarios called *vignettes*, where a medical case consists of 3 to 6 sentences and describes a patient's symptoms along with laboratory

Table 1. Creating an Extended-Matching Question

Components	Content
Theme	Content area
Lead-in statement	Instructions for learner
Options	Possible answers (6–26; pictures, phrases, or words)
Items	Questions in a content area (few, breadth; many, depth)

examination results, and a learner chooses an option (diagnosis).¹⁵ The questions usually require evaluation or treatment skills and are evidence based rather than based on a philosophical approach to optimize the accuracy of the answers.¹⁵

Extended-matching questions contain 4 elements: a theme, a lead-in statement, an option list (>5), and 2 or more items or questions.¹¹ A systematic approach made up of the following steps is suggested (Table 1).¹¹ The first 2 steps consist of establishing the content area to be assessed and directions for the students. The first step is the theme or general topic for a set of questions.¹¹ For example, the diagnosis of a low back pain condition, treatment strategy for low back pain, class of drugs, or anatomical function. The next step is the lead-in statement, which provides clear directions for the set of questions.¹¹ For example, "...for each patient with shoulder pain, select the most likely diagnosis." Also, include a statement clarifying that each option could be used once, more than once, or not at all.

The third step is more time consuming and involves creating the option list or choices that apply to the questions in the set usually ranging from 6 to 25 or more in number.¹¹ One way of determining the number is by the level of expertise of the learners. For example, in a 2-year Master of Science Athletic Training Program, the first-year learners may be given 10 options and the second year 15 to 20 for questions under the same theme. The options should be consistent with the purpose of the theme and level of learner competence. They are usually single words, short phrases, or possibly pictures of the body with regions labeled as options, in the case of anatomy.²⁹ It is recommended that the options be listed in alphabetical order and be homogeneous or all relevant and created from the theme.¹¹ One list can be used for multiple questions, and 1 option could be the answer to multiple questions. There should be a single best answer for each item (R type), and the rest of the options are relevant distractors.

Once a list of options is formed, it is time to create the set of questions within a theme that are to be answered to complete the process.¹¹ For example, with a diagnosis theme, a question could include a complete list of signs and symptoms attained from a hypoductive gathering of data to assess a learner's evaluation process expertise, or the question could supply a brief set of signs and symptoms specific to a condition assessing pattern recognition (a script). The questions should not contain misleading information that would make distractors too attractive, unless the learners are at an advanced level.¹¹ To minimize cueing, questions should be parallel in structure,¹¹ for example, all patients being athletes with musculoskeletal injuries. In addition, the content being assessed in the questions should directly reflect the learning

Table 2. Extended-Matching Question Example for Evaluation in Athletic Training (Diagnosis of Musculoskeletal Conditions)

Lead-in: For each athlete with knee pain, select the 1 most likely condition from the list of options. Each option can be used once, more than once, or not at all.

Option list:

- A. Iliotibial band friction syndrome
- B. Lateral collateral ligament tear
- C. Common peroneal nerve compression
- D. Patellar tendinitis
- E. Meniscus tear
- F. Patellofemoral pain syndrome
- G. Stress fracture
- H. Plica syndrome

Items:

1. “A volleyball player has been wearing a functional knee brace for a previous knee reconstruction surgery. She has been complaining of sharp lateral knee pain that radiates into her lower leg. What is the likely condition?” Correct option: **C**
2. “A cross-country athlete has been complaining of anterior right knee pain that on occasion presents as posterior knee pain. His pain increases when going up and down steps and sometimes as he sits down. You also note that his pelvis tilts to the left during gait. What is the likely condition?” Correct option: **F**

objectives, referred to as blueprinting.¹¹ In other words, when creating an examination that covers material in a specific content area, construct EMQs that directly assess the established learning objectives for that area.

APPLICATION OF EXTENDED-MATCHING QUESTIONS IN ATHLETIC TRAINING

With EMQs being effectively used in medicine as vignettes,¹¹ they could also be used effectively in the areas of clinical reasoning for the evaluation and diagnosis of musculoskeletal injuries and treatment techniques (exercise, manual, and mechanical therapies) in athletic training. There are standards listed in the Commission on Accreditation of Athletic Training Education’s 2020 Standards for the Accreditation of Athletic Training Programs document that specifically ask for applied knowledge.³⁰ Several examples are located within the Examination, Diagnosis, and Intervention subheading under Section IV. One example is “Perform an examination to formulate a diagnosis and plan of care for patients with health conditions commonly seen in athletic training practice.”^{30(p13)} Another example is “Select and incorporate interventions (for pre-op patients, post-op patients, and patients with nonsurgical conditions) that align with the care plan.”^{30(p14)} The following are examples of using EMQs to assess skills and knowledge in athletic training.

Assessing Evaluation Skills

Vignettes regarding musculoskeletal injury evaluation (Table 2) and treatment (Table 3) content areas can be used as questions in the athletic training program curricula. A theme

Table 3. Extended-Matching Question Example for Treatment in Athletic Training (Musculoskeletal Conditions)

Lead-in: For each athlete with shoulder pain, select the 1 most likely form of treatment from the list of options. Each option can be used once, more than once, or not at all.

Option list:

- A. Ischemic compression
- B. Deep transverse friction massage
- C. Strain-counterstrain
- D. Percussion massage
- E. Joint mobilization
- F. Active release
- G. Joint traction
- H. Myofascial release

Items:

1. “Your patient complains of neck-shoulder pain with no known cause. It has persisted for approximately 7 months. You have palpated a small tender nodule in the midtrapezius region that elicits pain into the shoulder and upper arm region with pressure. What form of effective treatment might you employ for immediate pain relief?” Correct option: **A**
2. “A baseball player complains of a gradual onset of shoulder pain that has persisted for approximately 2 months. From your clinical examination, you find he has a positive Hawkins/Kennedy sign, and when reaching behind the back, he is unable to reach up his back as high on the affected side. Your goal is to address the cause of the condition, so what form of treatment would be most effective?” Correct option: **E**

for an evaluation and diagnosis scenario can be specific to a certain region of the body, for example, the hip, ankle, low back, or knee, as in Table 2. The list of options can be consistent with the musculoskeletal conditions commonly found in that region. For example, with low back pain, the list of options could be intervertebral disc lesion, muscle strain, intervertebral joint sprain, sciatic nerve compression, myofascial pain, trigger point, spondylolisthesis, or spondylolysis. This set of options would be homogenous or all relevant to possible conditions sustained in the low back region. The questions for a low back pain scenario could provide a grouping that fits a script for a condition such as pain symptoms with extension for a spondylolysis condition or with flexion for an intervertebral disc lesion. Themes for evaluation can also come in the area of acute care. For example, a homogenous list of options could all be relevant to emergency conditions such as dehydration, concussion, epileptic seizure, brain injury, heatstroke, diabetic coma, shock, or hypotension. The questions could involve scenarios that provide signs and symptoms consistent with 1 of these acute, catastrophic condition options, such as increased thirst, nausea, rapid heart rate, and fruity breath odor, which are consistent with a diabetic coma.

Assessing Treatment Skills

Knowledge and skills in the treatment of musculoskeletal conditions also need to be assessed in an athletic training

Table 4. Extended-Matching Question Example for Pharmacology in Athletic Training (Acute Care)

Lead-in: For each athlete experiencing a traumatic event, select the 1 most likely action from the list of options. Each option can be used once, more than once, or not at all.

Option list:

- A. Remove facemask
- B. Begin cardiopulmonary resuscitation
- C. Remove helmet
- D. Place on spine board
- E. Locate automated external defibrillator
- F. Stabilize C-spine
- G. Check pulse
- H. Activate emergency medical services

Items:

1. "You are covering a high school football game, and a player is down on the field. You survey the field, and it is safe to approach the athlete. Once there, you do not see any gross deformities or bleeding. The athlete is conscious, talking coherently, and not reporting any abnormal symptoms. Which is the most likely action to perform first?" Correct option: **G**
2. "You are covering a high school football game, and a player is down on the field. You survey the field, and it is safe to approach the athlete. Once there, you do not see any gross deformities or bleeding. The athlete is unconscious. Which is the most likely action to perform first?" Correct option: **F**

program. A theme for the treatment of a musculoskeletal condition could supply an option list homogenous in the way of manual therapy, such as that provided in Table 3 for a shoulder condition. Homogenous options could also consist of exercise therapy techniques, such as examples of stretching, strengthening, agility drills, functional training, neuromuscular control, proprioception, plyometrics, or other power training exercises. Another treatment option list could consist of mechanical devices, such as ultrasound, diathermy, electric stimulation, laser, cold pack, hot pack, TENS unit, or mechanical traction. Management of emergency situations, such as acute care injuries or illnesses, can be another area assessed with EMQs. Emergency scenario questions are created, and an option list of how to manage them is provided in Table 4.

Another area that would work well with EMQs is in pharmacology, either with drug adverse effects or with pharmacodynamics (mechanism of action) of drugs (Table 5). The list of options provided in Table 5 is homogenous in that they all fall under the same category of therapeutic medication (nonsteroidal anti-inflammatory drugs [NSAIDs]). Item 1 of Table 5 is a question addressing the adverse effect that can occur if aspirin is used with children (Reye's syndrome). The second question addresses the mechanism of action of a particular NSAID drug (Celebrex; Pfizer, New York, NY) as a selective COX-2 enzyme inhibitor. Although the option list is homogenous (all NSAIDs), the questions are not parallel. One option assesses the learner's understanding of the adverse effects of a particular NSAID medication, and the other assesses the

Table 5. Extended-Matching Question Example for Therapeutic Drugs in Athletic Training

Lead-in: For each of the scenarios, select the 1 most likely drug from the list of options. Each option can be used once, more than once, or not at all.

Option list:

- A. Bayer
- B. Advil
- C. Tolectin
- D. Celebrex
- E. Daypro
- F. Feldene
- G. Naprosyn
- H. Lodine

Items:

1. "A high school soccer player had a respiratory viral infection with an associated fever last week. She comes into the athletic training room seeming much more sluggish than usual and also came in late thinking it was earlier in the day. She admitted having episodes of vomiting the last couple days. Which of the medications could be the cause?" Correct option: **A**
2. "The mechanism of action of this therapeutic drug is to selectively inhibit the COX-2 enzyme. Which medication would have that effect?" Correct option: **D**

learner's understanding of the difference in mechanism of action of drugs within the NSAID category of therapeutic medications. It was suggested to provide a homogenous list of options for a particular theme and parallel set of questions that address specific learning objectives.¹¹ In this example, the items should both either assess the learner's understanding of the mechanism of actions of drugs or adverse effects.

Assessing Anatomy Function

Extended-matching questions in anatomy can also be constructed to assess the application of knowledge. The questions could supply information regarding a particular movement, and the learner chooses from a list of muscles as options, 1 of which could provide that movement (Table 6). This example would be addressing a learner's understanding of muscle actions and how they apply to the human body with limb movement as opposed to identifying a particular muscle from a photo or description that would be addressing factual recall. Addressing the role or function of a muscle, nerve, blood vessel, or organ as opposed to identifying a structure or position of structure assesses a higher order of cognitive processing.

The examples used in Tables 2–6 provide an 8-option, 2-question format for each topic. Limiting the number of options to 8 rather than 26 and presenting only 2 questions can provide an examination of greater breadth of subject matter by creating a large sampling of themes. If the goal is depth within specific topics, a greater number of options and questions per theme would be recommended. The same format could be used to assess the application of knowledge for other classroom learning objectives in athletic training.

Table 6. Extended-Matching Question Example for Anatomy

Lead-in: For each of the muscle action descriptions, select the 1 most likely muscle from the list of options. Each option can be used once, more than once, or not at all.

Option list:

- A. Teres minor
- B. Supraspinatus
- C. Pectoralis minor
- D. Teres major
- E. Deltoid
- F. Subscapularis
- G. Infraspinatus
- H. Latissimus dorsi

Items:

1. "You ask your athlete to elevate his or her arm overhead in the sagittal plane. Which muscle would you primarily be using?" Correct option: **E**
2. "You ask your athlete to throw a baseball. Which muscle will provide the greatest action concentrically?" Correct option: **F**

DISCUSSION

The goal in using EMQs is to minimize the cueing (recognition) effect and success in guessing that can occur with MCQs by providing many possible combinations of questions and options.⁵ Also, by using scenarios instead of facts, the questions provide an authentic approach that assesses the application of knowledge instead of simply factual recall.⁵ The EMQ format was first developed in medicine to assess clinical reasoning, specifically diagnostic pattern-recognition skills of physicians, and are now used extensively in medical licensing examinations.²⁴ The EMQ format can be used extensively in the athletic training program curricula and in the Board of Certification examination, especially assessing clinical reasoning skills.

Clinical Reasoning

Clinical reasoning skills are needed with many aspects of the athletic training profession, emphasizing their importance in becoming a competent provider. Clinical reasoning is a form of cognitive processing that has been mentioned synonymously with decision making, problem solving, or judgment, where data from a clinical scenario is analyzed and integrated by the clinician using their knowledge and experiences to evaluate and treat conditions.³¹ There are several clinical reasoning strategies that have been identified as ways to diagnose a condition. One is termed *hypothetico-deductive reasoning* or backward reasoning, which is a data-gathering strategy where the clinician performs a comprehensive data collection (signs and symptoms) for a condition and from this data deduces a diagnosis.³² Another strategy called *schema-inductive reasoning* or forward thinking, is a mental categorization of information.³² This strategy requires long-term memory to have an organized mental framework of subject matter for a topic area, referred to as schema. This will allow the clinician to recognize signs and symptoms as they are being gathered, since these are embedded, allowing for forward reasoning. A third strategy is called *pattern recognition*, which is a type of

schema where similar pieces of information form a specific grouping.³² Pattern recognition in clinical reasoning means a grouping of data specific to the evaluation or treatment of a condition is embedded in long-term memory. In medicine, this is referred to as an *illness script*. These scripts will develop cognitively as the clinician evaluates the same condition repeatedly. Illness scripts have been shown to improve diagnostic accuracy with nonexpert medical students and are more commonly used by expert clinicians than novice.³³

There is additional research with physicians that supports a strong relationship between forward reasoning (schema inductive or pattern recognition) and diagnostic accuracy, and a poor relationship between recall and diagnostic accuracy.³⁴ It is important to understand these clinical reasoning strategies to administer optimal examination question formats. Research was conducted to find if there was a difference in clinical reasoning strategies used by learners when confronted with different examination question formats.³² This study³² compared medical school learners with practicing experts taking examinations with MCQ and EMQ formats and found that, although both groups used schema-inductive and pattern-recognition strategies, the experts used them more frequently than learners. These results suggest that, as learners gain expertise, they begin to use forward reasoning to a greater degree when engaging clinical reasoning skills.³² It also suggests either question format can assess higher-order cognitive processing, if well-constructed.³² Consequently, when a MCQ does not assess higher levels of cognitive processing, it is not necessarily due to the format but rather the content of the question.¹¹

Multiple-Choice Questions

Although it is possible to promote critical thinking and assess applied knowledge with MCQs, they need special scrutiny to have that capability. The key characteristics of MCQs to promote a higher level of cognitive processing are difficulty index and discriminatory power, which is determined by the effectiveness of the distractors (incorrect options).⁴ Multiple-choice questions need high levels of discrimination power,⁸ giving them the ability to accurately distinguish between stronger and weaker performing learners.⁴ This is accomplished by creating quality, plausible distractors¹⁴ that will reduce cueing.¹² Examples of cues to the right answer are such things as options stating always, never, completely, absolutely, or options identical to or similar to words in the question.³⁵ Cueing does not only lead learners to the correct option, called *positive cueing*, but they can also lead learners to the wrong option, called *negative cueing*.¹⁰ It has been suggested that a properly functioning distractor can be determined objectively by being selected by at least 5% of the learners.⁴ That would mean, in a class of 25 to 30 learners, the distractor options should be selected by 1 to 2 learners.

A meticulous approach to the construction of MCQs is required to assess applied knowledge. A study that critiqued 10 summative examinations in undergraduate nursing courses for question-writing flaws found there were a significant number of unclear questions, unnecessary information, implausible distractors, and cues to learners.¹⁴ The examination questions also had low discriminating power, which benefited exam-smart learners by cueing them to correct answers and harmed the more capable learners.¹⁴ Another

study⁴ inquiring examinations taken by undergraduates from various disciplines found only 55% of the distractors functioned properly (5% of learners selecting a distractor). As hypothesized, when the number of plausible distractors increased in this study, the questions increased in difficulty, and the discriminatory power became greater.⁴ Other research¹⁰ investigating cueing discovered examinations using MCQs had significantly higher scores compared to free-response (short answer) formats when assessing the same material, suggesting cueing may have had an impact on the improved scores. Multiple-choice questions are an efficient form of written assessment with psychometric measures, such as reliability and validity demonstrating good results, primarily due to the breadth of material covered creating a large sampling.³⁶ They have the ability to assess higher levels of cognitive processing, but they are cumbersome to create, subject to cueing, and can easily fall to the assessment of factual recall.

Extended-Matching Questions

The extended-matching selected-response question format is an alternative to the MCQ that can consistently assess applied knowledge and avoids the difficulty in construction encountered with MCQs. A number of criteria was investigated to evaluate the quality of EMQs as an assessment tool for applied knowledge. Regarding examination acceptability, questionable face validity was found with medical school learners who were first introduced to the EMQ format.¹⁷ This outcome could be explained by their years of previous practice in eliminating distractors on MCQs.¹⁷ It has been suggested that, with a little practice, learners will become accustomed to the EMQ format and prepare accordingly.³⁷ A study²² with veterinary school learners found their attitudes toward the use of EMQs as positive and felt it was an appropriate tool to assess clinical reasoning and could relate the scenarios asked to experiences in their clinical rotations. It appears poor face validity with EMQ assessment is due to the unfamiliarity of their format and should decline with experience.¹⁷

Research in medical education has attempted to interpret the quality of the EMQ format using psychometric property measures. Several studies using medical school learners^{17–21} and 1 with veterinary school learners²² have found acceptable reliability measures using various formulas when administering examinations with EMQs. In a comparison study³ with medical school learners, it was discovered that the EMQ format had greater reliability than the MCQ format. Several studies found there was adequate examination reliability (0.83–0.85) with question sampling sizes of 55,¹⁸ 52,¹⁹ and 50²² questions. It was also discovered that, if the number of questions²⁰ or themes²¹ increased, the reliability would correspondingly increase. These findings suggest that an adequate sampling of questions and topic areas along with close blueprinting to learning objectives²² will positively influence the reliability of an examination using the EMQ format. Good examination validity was demonstrated using EMQs, first with medical school learners, who were given an examination consisting of 100 EMQs.¹⁷ The examinations were grouped into 9 series of 100 questions each and administered for 7 years. A moderate correlation (0.45) was found between the examination score per series and the total examination score of each of the 7 years, providing evidence of construct validity.¹⁷ Another study²³ using medical school

learners found EMQs to have a strong correlation (0.65) with MCQs, suggesting having a strong correlation with MCQs demonstrates validity. The MCQ is an established valid assessment tool. Additional research¹⁸ with medical school learners found the scores of examinations consisting of EMQs increased with greater learner experience, again demonstrating a valid examination

With EMQ formats having a greater number of options than MCQs, they have the ability of being more difficult and having greater discriminatory power. A study²⁰ with medical school learners found that, with most questions, the EMQs were able to discriminate between strong and poor performers. Another study²⁴ with medical school learners found that increasing the number of options per 2-question set from 9 to 23 substantially improved the discriminatory power. Other studies using medical school learners that compared the discriminatory power between MCQs and EMQs found EMQs were better at detecting excellent performing learners²⁵ and excellent from marginal.³

The explanation for good discriminatory power with an EMQ format may come from the larger number of options and the questions being authentic scenarios. Studies have also investigated the EMQs' ability to assess applied knowledge in the way of problem solving and clinical reasoning. Research²⁰ using medical school learners being assessed by examinations consisting of 25 EMQs, each having 9 options, found the questions had the ability to assess higher-order knowledge, specifically problem-solving skills. Another study³ using medical school learners compared MCQ, true or false, EMQ, and free-response question formats and concluded both the extended-matching and free-response formats were superior in assessing problem-solving skills. Additional research⁷ with medical school learners used an examination consisting of questions with 4 formats, EMQ, MCQ, SAQ, objective structured clinical examination (OSCE), that were given to medical school learners. Six themes of EMQs (total of 25 questions) were used to assess problem solving.⁷ The EMQs seemed to measure clinical problem solving because they had a strong correlation with OSCE (0.77) and SAQs (0.72).⁷ The correlation between MCQs and OSCE was lower (0.56). Consequently, from these results, the EMQs are more appropriate for assessing the application of knowledge than MCQs. It was also recommended they be conducted during the final stages of the curriculum, where challenging a higher level of cognitive processing is appropriate.⁷

Research inquiring about EMQs and the assessment of clinical reasoning skills has been conducted. One study²⁶ found that internal medicine residents provided more correct answers than medical school learners and used more forward than backward reasoning strategies. This would suggest that EMQs seem to allow for the detection of differences in clinical reasoning skills between novice and advanced learners.²⁶ There is evidence to support the use of forward reasoning skills in the accuracy of diagnosis,³³ and their use tends to increase with expertise.³² An instrument called the Diagnostic Thinking Inventory is a reliable and valid examination for measuring diagnostic thinking.³⁸ The Diagnostic Thinking Inventory had a significant correlation with EMQ examination scores with medical school learners, suggesting EMQs can measure the diagnostic aspect of learner clinical reasoning.²⁷

The ability of EMQs to distinguish a difference in clinical reasoning skills between advanced and novice clinicians demonstrates adequate assessment tool validity and the ability to assess applied knowledge.

There are factors other than psychometric measurements worth considering with an EMQ format that provide support for its choice in written assessments. Because of their ease of construction by using a systematic approach, EMQs were determined to be feasible.^{3,17,18} Unlike examinations consisting of free-response questions that assess applied knowledge, EMQs can be conducted electronically with minimal resources needed, making them very cost effective.¹⁸ Extended-matching questions use a large number of options per set, which minimizes cues and likelihood of guessing the correct options.³ In addition to being an effective tool for the assessment of applied knowledge, it is an efficient question format to construct at a minimal expense. Extended-matching questions are an efficient and effective way to assess learning objectives (blueprinting) that have been determined for a particular course.

Research³⁹ was conducted comparing anatomy class learners taking weekly quizzes consisting of either MCQs or EMQs and their performance on interim and final examinations. Learners from the EMQ group had the highest average scores on all interim examinations (consisting of MCQs), but the difference was not statistically significant due to a small group size. The grades on the final examinations (oral and practical) were significantly higher in the EMQ group. The innovation of the EMQ assessment on a weekly basis may direct the learners' approach to examination preparation, and a higher level of knowledge was acquired. In other words, learners will engage in the weekly chunks of material, knowing the assessment will demand an in-depth understanding. It was determined that the EMQs were easier to construct than MCQs, and the use of EMQs as a means of monitoring the learners' progress was recommended.³⁹ When comparing EMQs with other written examination formats, there is a consistent demonstration of higher psychometric measures (reliability: 0.83–0.85; correlation: 0.45–0.77) with less difficulty in their creation and a more valid measure of applied knowledge.

Theme, Option, and Question Quantities

Studies inquiring about the optimal quantities of themes, options, and questions were also conducted. One such study²¹ analyzed the variance between learner-topic and learner-question interaction on EMQ examinations. The results found the variance attributed to questions was greater than topics. The topics or themes used seemed to be less important than the individual clinical scenarios used as questions. The participants being medical residents may have had existing illness scripts embedded in long-term memory, allowing them to recognize the answers to scenarios, effectively explaining the variance.²¹ These results suggest the reliability of EMQ examinations may improve from an increase in topics rather than questions per topic; unfortunately, this would also significantly increase examination time.²¹

Research¹⁶ analyzing the US Medical Licensing Examination (USMLE, sponsored by the Federation of State Medical Boards and National Board of Medical Examiners) found

that questions with larger numbers of options are more difficult than fewer numbers. However, there was little difference in question discrimination as a function of question format or number of options when the question order was randomized.¹⁶ The time necessary to complete the examination was also significantly longer with a greater number of options. Consequently, decreasing the number of options may not have a negative impact on discrimination and significantly reduces examination time if you randomize the order of questions. A common range of options is 5 to 26 for EMQs, but 8 options were found to be adequate and make efficient use of examination time.¹⁶ Keep in mind that fewer options place greater pressure on the effectiveness of the distractors. Previous research²⁴ that analyzed the results from the USMLE discovered that, when increasing the number of options, it significantly increased the question difficulty due to a lower probability of guessing the correct answer and in turn improved question discrimination. Consequently, these earlier results from the USMLEs suggest providing a large quantity of options.²⁴

Other research²⁰ exploring the quantity of EMQs used medical school learners who were given 25 EMQ items with 9 options each and found that most questions were good at discriminating high performers from poorer. It was also showed that, although the reliability was acceptable, to improve it, the total number of questions may need to increase significantly.²⁰ An EMQ examination was used to assess neurological localization with medical school learners and postgraduate trainees.¹⁸ The examination consisted of 55 questions in various clinical scenarios with 25 options. It took 1 hour to complete and resulted in positive psychometric outcomes.¹⁸ In research using 52-question examinations,¹⁹ it was suggested that increasing the difficulty of the questions would require fewer total questions to have comparable reliability. The number of items to include in a set depends on the purpose of the examination. Sets with few questions (2) are optimal for sampling the breadth of material.²⁴ In other words, with fewer questions per list of options, one could provide a large number of themes and cover a greater amount of material in an examination. Sets with many items (20) under the same theme are optimal for sampling depth.²⁴ For example, a large number of questions asked using the same list of options provides an opportunity for the examiner to approach the same subject matter from a number of scenarios using various contexts. The optimal number of themes, options, and questions to include in a written examination for acceptable psychometric measure outcomes has yet to be established. The studies provided assistance in creating general guidelines to use that will improve sampling and in turn reliability and accuracy of the examination.

CONCLUSIONS

In summary, the advantages of using the EMQ format can be looked at from both the developmental and psychometric measure standpoints. Regarding the development of an examination, creating an EMQ is accomplished in a feasible, systematic fashion, enabling blueprinting or linking assessments to learning objectives in a natural process. The questions consist of a homogeneous, extended list of relevant options derived from a theme followed by developing a set of parallel questions, all produced with minimal concern of

writing flaws. They are cost effective and can be marked electronically and statistically analyzed. The EMQ format is a valid and reliable examination used to assess the application of knowledge that drives learner comprehension in the direction of critical thinking and understanding subject matter. The EMQs provide an authentic presentation in the way of clinical cases, which adds to the validity of outcome measures and assessment of higher levels of cognitive processing. Learner performance usually suffers with the first exposure to EMQs, but learners adapt quickly and, along with faculty, have stated positive acceptability with their use in summative examinations.

Examinations using EMQs are most suited for summative assessments at the end of units, preferably later in the curriculum, but were also used in weekly quizzes for monitoring the learners understanding of material. The successful use of the EMQ format in high-stakes examinations such as the USMLE to assess clinical reasoning skills makes this type of question format suitable to assess the application of knowledge on the Board of Certification examination in athletic training.

REFERENCES

1. Towns MH. Guide to developing high-quality, reliable, and valid multiple-choice assessments. *J Chem Educ.* 2014;91(9):1426–1431.
2. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet.* 2001;357(9260):945–949.
3. Fenderson BA, Damjanov I, Robeson MR, Veloski JJ, Rubin E. The virtues of extended matching and uncued tests as alternatives to multiple choice questions. *Hum Pathol.* 1997;28(5):526–532.
4. DiBattista D, Kurzawa L. Examination of the quality of multiple-choice items on classroom tests. *Can J Scholarsh Teach Learn.* 2011;2(2):1–23.
5. Schuwirth LW, van der Vleuten CP. ABC of learning and teaching in medicine. *BMJ.* 2003;326(7390):643–645.
6. Scouller K. The influence of assessment method on student's learning approaches: multiple choice question examination versus assignment essay. *High Educ.* 1998;35(4):453–472.
7. Wass V, McGibbon D, Van der Vleuten C. Composite undergraduate clinical examinations: how should the components be combined to maximize reliability? *Med Educ.* 2001;35(4):326–330.
8. Krathwohl DR. A revision of bloom's taxonomy: an overview. *Theory Pract.* 2002;41(4):212–218.
9. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(suppl 9):S63–S67.
10. Desjardins I, Touchie C, Pugh D, Wood TJ, Humphrey-Mutro S. The impact of cueing on written examinations of clinical decision making: a case study. *Med Educ.* 2014;48(3):255–261.
11. Case SM, Swanson DB. Extended-matching items: a practical alternative to free-response questions. *Teach Learn Med.* 1993;5(2):107–115.
12. Hift RJ. Should essays and other open-ended-type questions retain a place in written summative assessment in clinical medicine? *BMC Med Educ.* 2014;14:249.
13. Elstein AS. Beyond multiple-choice questions and essays: the need for a new way to assess clinical competence. *Acad Med.* 1993;68(4):244–249.
14. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. *Med Educ.* 2008;42(2):198–206.
15. Wood EJ. What are extended matching questions? *Biosci Educ.* 2003;1(1):1–8.
16. Swanson DB, Holtzman KZ, Allbee K. Measurement characteristics of content-parallel single-best-answer and extended matching questions in relation to number and source options. *Acad Med.* 2008;83(suppl 10):S21–S24.
17. Beullens J, van Damme B, Jaspert H, Janssen PJ. Are extended-matching multiple-choice items appropriate for a final test in medical education? *Med Teach.* 2002;24(4):390–395.
18. Yau CWL, Chin HX, Tan K, Lim E, Ponnampuruma G, Tan N. Assessing the clinical skill of neurological localization using extended-matching questions [abstract]. *Neurol.* 2013;80(suppl 7).
19. Kreiter CD, Ferguson K, Gruppen LD. Evaluating the usefulness of computerized adaptive testing for medical in-course assessment. *Acad Med.* 1999;74(10):1125–1128.
20. Vuma S, Sa B. A descriptive analysis of extended matching questions among third year medical students. *Int J Res Med Sci.* 2017;5(5):1913–1920.
21. Dory V, Gagnon R, Charlin B. Is case-specificity content-specificity? An analysis of data from extended-matching questions. *Adv Health Sci Educ.* 2010;15(1):55–63.
22. Tomlin JL, Pead MJ, May SA. Veterinary students' attitudes toward the assessment of clinical reasoning using extended matching questions. *J Vet Educ Med.* 2008;35(4):612–621.
23. Slattery RM. Objective versus subjective methods to assess discipline-specific knowledge: a case for Extended Matching Questions (EMQs). Talk presented at: 3rd International Conference on Higher Education Advances (HEAd), 2017, Valencia, Spain. <http://dx.doi.org/10.4995/HEAd17.2017.5473>. Accessed February 8, 2019.
24. Case SM, Swanson DB, Ripkey DR. Comparison of items in five-option and extended-matching formats for assessment of diagnostic skills. *Acad Med.* 1994;69(suppl 10):S1–S3.
25. Eijsvogels TMH, van den Brand TL, Hopman MTE. Multiple choice questions are superior to extended matching questions to identify medicine and biomedical sciences students who perform poorly. *Perspect Med Educ.* 2013;2(5–6):252–263.
26. Buellens J, Struyf E, van Damme B. Do extended matching multiple-choice questions measure clinical reasoning? *Med Educ.* 2005;39(4):410–417.
27. Beullens J, Struyf E, van Damme B. Diagnostic ability in relation to clinical seminars and extended-matching questions examinations. *Med Educ.* 2006;40(12):1173–1179.
28. Samuels A. Extended matching questions and the Royal Australian and New Zealand College of Psychiatrists written examination: an overview. *Australas Psychol.* 2006;14(1):63–66.
29. Vorstenbosch MA, Bouter ST, van der Hurk MM, Kooloos JGM, Bolhuis SM, Laan RF. Exploring the validity of assessment in anatomy: do images influence cognitive processes used in answering extended matching questions? *Anat Sci Educ.* 2014;7(2):107–116.
30. 2020 standards for accreditation of athletic training programs. Commission on Accreditation of Athletic Training Education Web site. <https://caate.net/wp-content/uploads/2018/09/2020-Standards-for-Professional-Programs-copied-clean.pdf>. Updated January 9, 2018. Accessed January 28, 2019.
31. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ.* 2005;39(9):418–427.

-
32. Coderre SP, Harasym P, Mandin H, Fick G. The impact of two multiple-choice question formats on the problem-solving strategies used by novices and experts. *BMC Med Educ.* 2004;4:23.
 33. Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. *Med Educ.* 2003;37(8):695–703.
 34. Patel VL, Groen GJ, Arocha JF. Medical expertise as a function of task difficulty. *Mem Cognit.* 1990;18(4):394–406.
 35. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. *Appl Meas Educ.* 2002;15(3):309–334.
 36. McCoubrie P. Improving the fairness of multiple-choice questions: a literature review. *Med Teach.* 2004;26(8):709–712.
 37. Baird AS. The new extended matching question (EMQ) paper of the MFSRH examination. *J Fam Plann Reprod Health Care.* 2010;36(3):171–173.
 38. Round AP. Teaching clinical reasoning—a preliminary controlled study. *Med Educ.* 1999;33(7):480–483.
 39. Lukić IK, Glunčić V, Katavić V, Petanjek Z, Jalsovec D, Marusić A. Weekly quizzes in extended-matching format as a means of monitoring students' progress in gross anatomy. *Ann Anat.* 2001;183(6):575–579.