ASSESSING HIGHER-ORDER THINKING SKILLS THROUGH MULTIPLE-CHOICE QUESTIONS

L. Train [1]

Abstract:
Multiple-choice questions (MCQs) are favored by educators for their simplicity in both construction and grading, as well as for their utility in data collection and analysis. However, it is often assumed that MCQs can only measure lower-order thinking skills (LOTS). This paper challenges that assumption by examining three case studies - the Force Concept Inventory, a university biology exam developed by Jamie Jensen and colleagues, and the NCLEX-RN - to illustrate how scenario-based questions and carefully crafted distractors can successfully assess higher-order thinking skills (HOTS). The findings suggest that the deliberate design of MCQs around real-world scenarios and plausible distractors can elicit application, analysis, and evaluation, thereby expanding the potential of MCQs beyond the testing of rote memorization.

Key words: assessment, higher-order thinking skills, multiple-choice questions, Bloom's Taxonomy, analytical thinking, critical thinking.

Introduction

Multiple-choice questions (MCQs) are attractive to many educators because they are relatively simple to formulate, require minimal time to grade, and allow for the straightforward collection and analysis of test data. These features can inform other stages of the planning process by providing timely insights that help educators adapt instruction to students' needs. However, it is widely believed that MCQs mainly assess lower-order thinking skills (LOTS), such as remembering and understanding, as per Bloom's Taxonomy. While these skills are foundational for learning, education should also aim to develop higher-order thinking skills (HOTS), such as application, analysis, and evaluation. These more advanced skills are notoriously difficult to measure, and many educators assert that MCQs are inherently unable to do so. This paper challenges that assumption by examining three case studies in which MCQs were effectively used to assess higher-order thinking. It then highlights the design features common to all three, culminating in recommendations for educators who want to develop their own MCQs capable of evaluating deeper cognitive processes.

Case Study 1: The Force Concept Inventory

The "Force Concept Inventory" (FCI) is a multiple-choice test developed in the United States in the 1980s to assess understanding and application of Newtonian mechanics among post-secondary physics students. Although multiple versions of this test exist, most contain only multiple-choice questions (Hake, 2011). Crucially, the FCI is designed to measure conceptual reasoning rather than rote memorization. Its items employ plain-language descriptions of real-life scenarios - such as dropping an object or pushing a cart - and challenge students to apply fundamental physics principles to novel contexts (Hake, 2011). Because the questions focus on understanding forces and motion in realistic situations, success depends on one's ability to analyze and apply underlying principles, rather than merely remember or understand them.

One distinctive feature of the FCI is its use of carefully constructed distractors that reflect common misconceptions (Hake, 2011). For example, a question about falling objects may include a distractor which reflects the common misconception that heavier objects fall faster. When all answer choices seem plausible, students must reason more deeply about why one option is correct and another incorrect. By avoiding isolated factual or formula-based queries, the test compels learners to think conceptually. A sample question from the FCI is reproduced below:

Two steel balls, one of which weighs twice as much as the other, roll off of a horizontal table with the same speeds. In this situation:

A. both balls impact the floor at approximately the same horizontal distance from the base of the table.

B. the heavier ball impacts the floor at about half the horizontal distance from the base of the table than does the lighter.

C. the lighter ball impacts the floor at about half the horizontal distance from the base of the table than does the heavier.

D. the heavier ball hits considerably closer to the base of the table than the lighter, but not necessarily half the horizontal distance.

[1] *Train Louis*

E. the lighter ball hits considerably closer to the base of the table than the heavier, but not necessarily half the horizontal distance.

(Hestenes et al., 1992, p. 154)

Case Study 2: Jensen's Biology MCQ Exam

In 2014, Professor Jamie Jensen and her colleagues at an American university developed an MCQ-based exam that explicitly sought to assess higher-order thinking skills among biology students (Jensen et al., 2014). Rather than simply asking students to define key terms, this exam required learners to apply their knowledge to novel scenarios, predict outcomes of hypothetical experiments, and evaluate data.

The exam's distractors were often highly similar to each other and the correct answer (Jensen et al., 2014). Learners therefore had to think through each option rather than relying on superficial cues to dismiss obviously incorrect answers. This approach forced them to integrate and weigh multiple pieces of information, bringing them into the higher domains of Bloom's Taxonomy. A sample higher-order question from Jensen's exam is reproduced below:

The hydrogen car runs off the power of hydrogen created from water in the process of electrolysis (i.e., the splitting of water). In this process, water is split into hydrogen and oxygen gases by running an electrical current through the water. This is done by placing two metal rods in the water and attaching them to an electrical source. The cathode donates electrons to the solution; whereas the anode accepts electrons from the solution. The reactions at each node are shown below (keep in mind that these reactions are happening at the same time): At the Cathode, where hydrogen gas bubbles are produced: $2H^+ + 2e^- \rightarrow H_2$ (gas) At the Anode, where oxygen gas bubbles are produced: $2H_2O \rightarrow O_2$ (gas)$+4H^+ + 4e^-$ xviii. How many electrons must come from an outside source (like the anode or cathode) to make this reaction run?

a. Two electrons are being donated by the cathode
b. Two electrons are being donated by the anode
c. Four electrons are being donated by the anode
d. Eight electrons are being donated by the anode
e. No electrons are being donated, they are all contained within the original water molecule

(Jensen et al., 2014, p. 326)

Case Study 3: The NCLEX-RN Nursing Licensure Exam

The NCLEX-RN is taken by thousands of nursing graduates in the United States and Canada each year and is widely regarded as the gateway to a respected, often well-paid profession. In North America, nursing degrees can take up to four years to complete, and the median annual salary of a nurse in the USA is nearly $90,000, making it essential that the licensing exam be both valid and reliable (Bureau of Labor Statistics, 2024). This computer-adaptive test primarily uses scenario-based items that necessitate analysis of patient cases, interpretation of clinical data, and application of nursing knowledge to deliver safe, effective care (National Board of Medical Examiners, 2024).

Typical questions provide a concise clinical vignette describing a patient's symptoms, medical history, and current status, then ask for the best action or decision (National Board of Medical Examiners, 2024). Because several answer choices frequently appear valid, examinees must discriminate among them based on urgency, appropriateness, and the clinical priorities outlined in nursing practice. Distractors are chosen to be plausible, reflecting realistic alternatives a nurse might consider (Wendt et al., 2007). The NCLEX-RN further incorporates stimuli like lab results, vital signs charts, or audio clips of heart and lung sounds, requiring examinees to interpret multiple data sources. Its emphasis on genuine clinical reasoning translates into the measurement of higher-order thinking skills - application, analysis, and evaluation - in a complex, real-world context. A sample vignette-type question from the NCLEX-RN is reproduced below:

VIGNETTE:

A 32-year-old man comes to the office because of a 4-day history of progressive weakness in his extremities. He has been healthy except for an upper respiratory tract infection 10 days ago. His temperature is 37.8°C (100.0°F), pulse is 94/min, respirations are 42/min and shallow, and blood pressure is 130/80 mm Hg. He has symmetric weakness of both sides of the face and the proximal and distal muscles of the extremities. Sensation is intact. No deep tendon reflexes can be elicited. Babinski sign is absent.

LEAD-IN:

Which of the following is the most likely diagnosis?

OPTION SET:

A. Acute disseminated encephalomyelitis
B. Guillain-Barré syndrome
C. Myasthenia gravis
D. Poliomyelitis

E. Polymyositis

(National Board of Medical Examiners, 2024, p. 12)

Analysis and Synthesis

Two consistent features emerge from these three case studies: the use of scenario-based questions and the construction of plausible distractors. All three exams center their questions around realistic contexts. The FCI describes everyday motions (like falling objects) that force students to apply physics principles rather than recite formulas; the biology exam requires predictions about experiments and ecosystems; and the NCLEX-RN presents patient vignettes requiring clinical judgment. These scenarios prompt learners to generate explanations, propose solutions, or anticipate outcomes - tasks that go well beyond basic recall.

Scenario-based MCQs are essential for assessing HOTS because they require learners to integrate their knowledge in a context that mirrors real-world problem-solving. By forcing students to analyze the relationships between details in the prompt, such items demand cognitive processes at the application, analysis, or evaluation levels of Bloom's Taxonomy. They also encourage metacognition, as students must reflect on the rationale for selecting or discarding particular options.

Equally important are the plausible distractors. When distractors match common errors or misconceptions - such as the belief that heavier objects fall faster - the test-taker must engage in critical discrimination. This process exposes any weak points in understanding and compels students to justify their chosen response against alternatives that appear correct on the surface. Developing these kinds of distractors is therefore integral to designing MCQs that genuinely gauge deeper thinking, since guesswork becomes less effective if a superficial glance cannot easily eliminate incorrect responses.

From this analysis, a straightforward synthesis emerges. Educators seeking to assess higher-order thinking skills with MCQs should prioritize real or realistic scenarios demanding explanation, solution-finding, or data interpretation. They should also ensure that distractors are drawn from genuine mistakes or misconceptions. By doing so, multiple-choice tests can move beyond testing recall and engage learners in higher-level cognitive tasks.

Conclusion

MCQs are not intended to replace every other form of assessment. Performance-based evaluations remain indispensable for judging hands-on competencies and other skills that do not lend themselves well to multiple-choice formats. Nonetheless, it is misleading to dismiss MCQs solely for their perceived limitation to lower-order thinking. As shown by the Force Concept Inventory,

Jamie Jensen's biology exam, and the NCLEX-RN, MCQs that feature realistic scenarios and credible distractors can effectively measure application, analysis, and evaluation. Instructors should thus consider the demands of their educational context when deciding how best to assess their students. By thoughtfully crafting MCQs that reflect real-world situations and plausible alternative answers, educators can broaden the cognitive scope of their assessments and ensure that students are challenged to exercise higher-order thinking.

References:

[1]. Bureau of Labor Statistics. (2024, August 29). Registered nurses. Occupational Outlook Handbook. https://www.bls.gov/ooh/healthcare/registered-nurses.htm

[2]. Hake, R. R. (2011, August 8). The impact of concept inventories on physics education and its relevance for engineering education [Invited talk]. Second Annual NSF-Sponsored National Meeting on STEM Concept Inventories, Washington, DC.

[3]. Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. The Physics Teacher, 30(3), 141–158. https://doi.org/10.1119/1.2343497

[4]. Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test. . .or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. Educational Psychology Review, 26(2), 307–329. https://doi.org/10.1007/s10648-013-9248-9

[5]. National Board of Medical Examiners. (2024). Constructing written test questions for the health sciences. https://www.nbme.org/sites/default/files/2021-02/NBME_Item%20Writing%20Guide_R_6.pdf#:~:text=%5BPDF%5D%20NBME%C2%AE%20ITEM,series%20of%20option%20choices%2C

[6]. Wendt, A., Kenny, L. E., & Marks, C. (2007). Assessing critical thinking using a Talk-Aloud protocol. CLEAR Exam Review, XVIII(1), ISSN 1076-8025. https://www.ncsbn.org/Assessing_Critical_Thinking_Talk_Aloud_Protocol.pdf