Multiple Choice Questions in a Nutshell: Theory, Practice, and Postexam Item Analysis

Multiple choice questions (MCQs) are the most widely used assessment method in medical education. Although volumes have been written about MCQs, their use can be summarized succinctly. MCQs, which are typically machine scored, are objective since they eliminate human judgment. Interrater reliability is, therefore, 1.0.

Structure of an MCQ
1. When did the Spanish flu pandemic begin?

- **stem** (problem presentation—can be succinct or can be vignettes with patient demographics, history, lab data, images, etc.)
- **A. 1787**
- **B. 1892**
- **C. 1918**
- **D. 1986**

A, B, and D are distractors (wrong answers—intended to identify candidates with superficial or partial knowledge)

Conceptual Depiction of Easy, Medium, and Difficult MCQs Based on Options

- **Minimum performance level (MPL)**
  \[ \text{MPL} = \frac{1}{O_p - \Sigma P_o} \]
  where:
  - \( O_p \) = The number of options in the item;
  - \( P_o \) = The probability that a minimally knowledgeable candidate can eliminate that option as clearly incorrect.

Examples of easy, medium, and difficult items and their MPLs
- **Item 1 above, A = 1.0, B = 1.0, C = *, D = 1.0**
- **Item 2 below has a medium difficulty (MPL = .61).**
- **Item 3 below is very difficult, as all of the probabilities of elimination are 0 (MPL = .25).**

2. When did the Spanish flu pandemic begin?

- **A. 1910**
- **B. 1918**
- **C. 1926**
- **D. 1934**

A = .75, B = *, C = .75, D = .85

MPL = .61

3. When did the Spanish flu pandemic begin?

- **A. January 1918**
- **B. March 1918**
- **C. October 1918**
- **D. December 1918**

A = 0, B = 0, C = *, D = 0

MPL = .25

Discrimination and Improving MCQs

<table>
<thead>
<tr>
<th>Item</th>
<th>Proportion correct</th>
<th>Discrimination ( \text{Disc} )</th>
<th>A (%)</th>
<th>B (%)</th>
<th>C (%)</th>
<th>D (%)</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.98 (easy)</td>
<td>.00 (poor)</td>
<td>0</td>
<td>2</td>
<td>98*</td>
<td>0</td>
<td>Too easy unless for mastery</td>
</tr>
<tr>
<td>2</td>
<td>.72 (medium)</td>
<td>.48 (good)</td>
<td>16</td>
<td>72*</td>
<td>10</td>
<td>2</td>
<td>Ideal for difficulty &amp; discrimination</td>
</tr>
<tr>
<td>3</td>
<td>.34 (difficult)</td>
<td>.15 (poor)</td>
<td>25</td>
<td>20</td>
<td>34*</td>
<td>21</td>
<td>Ambiguous; random responses</td>
</tr>
</tbody>
</table>

*Discrimination refers to how well an item differentiates among students based on whether they know the material. Values > .40 are considered good. % refers to the proportion of the students selecting each item.

Educators may improve items by writing them at higher levels of Bloom’s taxonomy—beyond knowledge or recall to comprehension, application, analysis, or synthesis—which increases the level of cognitive measurement. MPLs should be similar to difficulty levels (proportion correct).

Iterative process to create and refine MCQs

Write MCQs and MPLs → Administer MCQs → Complete item analysis → Edit MCQs and MPLs

References:

Author contact: violato@ualberta.ca; Twitter: @EfremViolato