APPENDIX A: EXAMINATION QUESTION VALIDATION ALGORITHM

Ensuring the Quality of Multiple-Choice Tests: An Algorithm to Facilitate Decision Making for Difficult Questions

Processes for ensuring the quality of multiple-choice question (MCQ)-based tests occur before, during, and after test administration.

BEFORE THE TEST
Faculty development, policies, and item-vetting processes ensure the development of a good test.

DURING THE TEST
Policies and processes ensure smooth and fair administration of the test.

AFTER THE TEST
Policies, guidelines, and decision algorithms ensure good decisions, including regarding test rescaling.

Despite the existence of guidelines, posttest analysis requires judgment and can be particularly challenging. The Student Assessment Subcommittee (Undergraduate Medical Education, McGill University) developed user-friendly decision algorithms to support instructors in interpreting item difficulty and discrimination indices. Here we present our approach for what to do with difficult questions (i.e., items that less than 30% of students have answered correctly). In our experience these flowcharts—which can be adapted according to local needs and procedures—empower instructors to use item analysis reports judiciously to reach transparent and defensible decisions regarding examination scoring.

Decision Algorithm for Difficult Questions

1. Review the answer key.
   Make sure that the answer flagged as correct is really the correct answer. Ask several colleagues’ opinion.

   Did most students select the same distractor?
   - YES
   - NO

   Is that distractor actually correct?
   - YES
   - Request a change to the answer key.
   - NO

   Did another teacher give a conflicting message?
   - YES
   - Consider accepting both responses as correct.
   - NO

2. Review the clarity of the question.
   Make sure that the test takers are interpreting the question as intended.

   Did students contest the question on the basis of lack of clarity?
   - YES
   - Remove the question from the total score (recommend removing no more than 5% of questions) or consider accepting more than one correct response.
   - NO

   Do several colleagues agree that the question is clear?
   - YES
   - NO

3. Ensure that the proportion of difficult questions is appropriate (≤ 15%).
   Since each test needs a range of question difficulty, do not remove a question simply because it is challenging.

   Are there too many difficult questions in the test (≥ 20%)?
   - YES
   - NO

   Is the item discrimination (point-biserial correlation) ≥ 0.17?
   - YES
   - NO

   Do other questions test the same objective?
   - YES
   - NO

Acknowledgments: The authors would like to thank Courtney Kirkby, the graphic designer who assisted in creating the initial flowchart.

References:

Author contact: ralene.dory@gmail.com; Twitter: @ralene_dory

Page | 30