Scientific, educational, linguistic& formative (SELF) evaluation strategy of a sample of multiple choice questions from introductory research course

Tarek M El-gohary, Abdullah M Al-Shenqiti

ABSTRACT

Background: The purpose of this study was to apply the scientific, educational, linguistic& formative (SELF) evaluation strategy on a sample of multiple choice questions (MCQs) administered at introduction to research course

Purpose: The aim of this study was to apply the SELF evaluation strategy on a sample of multiple choice questions (MCQs) administered at introduction to research course

Method and Materials: A number of MCQs were carefully chosen from a pool of administered MCQs. All questions were discussed along with the justification of choosing one answer and excluding the other options. The justification was built of running the SELF evaluation strategy on all questions.

Results: Discrimination and difficulty indices were calculated for all questions included. The indices were settled in a template for convenience.

Conclusion: It is concluded that SELF evaluation strategy is feasible and practical, even among inexperienced academics, to assess the validity of MCQs within the field of research. It is recommended to use SELF evaluation strategy to have quality MCQs that lead to developing competency in research methodology.

Key words: Evaluation strategy, research, SELF evaluation, MCQ, competence

DOI: 10.5455/ijhrs.0000000113
INTRODUCTION

Research courses are axial core courses within physical therapy and rehabilitation programs.\textsuperscript{1,2} Research courses should be designed to match educational level of learners.\textsuperscript{3,4} Course designer must take learners gradually step by step from introductory courses to intermediate courses to advanced courses.\textsuperscript{4,5} The courses become more sophisticated when taught at the postgraduate level.\textsuperscript{4,6} Multiple choice questions (MCQs) are the gold standard and the most commonly used format for testing knowledge acquisition within the research courses; however its validity is still questionable especially in the absence of standardized assessment technique.\textsuperscript{1} The majority of academics reported that there is no specific framework and guidelines to follow when designing MCQs exams within the research courses. Additionally, there is a desperate need among academics to build a database of standardized MCQs within the field of research methodology. Recently, Dr. El-gohary has created and published a series of educational papers to amend the loophole in the body of knowledge. Dr. El-gohary created the scientific, educational, linguistic& formative (SELF) evaluation strategy to validate the quality of the MCQs.\textsuperscript{7-12} There is a gap in body of knowledge regarding having standardized assessment techniques for MCQs within research courses. There are also uncertainties among academics regarding relying solely on their own experience in designing MCQs exams. We hypothesize that SELF evaluation strategy is feasible to be used by academics and can be easily applied on the MCQs from the introduction to research course. The objective of the current educational paper is twofold: first, to apply the SELF evaluation strategy on a sample of MCQs given at the introductory research course, and second, to calculate the difficulty and discrimination indices of the discussed MCQs.

METHODS AND MATERIALS

Sampling

A miscellaneous sample of MCQs from the first midterm exam of the introduction to research methodology course were carefully chosen to represent different levels of thinking that reflects different educational categories.\textsuperscript{13,14} SELF evaluation strategy created by Dr. El-gohary was applied to all included questions to substantiate the scientific background, match its educational category, assure its linguistic soundness and its formative output.\textsuperscript{7} The discrimination & difficulty indices were computed for all discussed questions.\textsuperscript{15,16} Ethical approval was obtained from college of medical rehabilitation sciences. Approval # (CMR-PT-2017-012).

Q1: A physical therapist researcher would like to study children with rare genetic disease. The researcher is currently having only one child under study. The best approach to recruit more children is through talking to children’s families. This kind of sampling is called:

a) Simple random sampling
b) Snow ball sampling
c) Stratified random sampling
d) Purposive sampling

Comments:

The correct answer is (b) since rare cases recruitment belongs to non-probability sampling and the snow ball sampling is the best choice.\textsuperscript{1,17} The answer options “a”, “c”, and “d” are wrong since all options
have included inappropriate sampling technique. A quick run of the SELF evaluation strategy has established the scientific aspect of the question. Regarding the educational aspect, it was satisfied and showed that the question lies within the comprehension-analysis spectrum and that reflects early higher order thinking skills. The linguistic and the formative aspects were satisfied. The question number one showed difficulty and discrimination indices of 0.69 and 0.15 respectively. That can be interpreted as medium difficulty index and fair discrimination index. The question fairly managed to differentiate more capable from less capable learners. The question can be included with full faith in the future exams.

Q 2: The physical therapist (PT) instructor has supervised three PT students during mobility assessment of the elbow joint aiming to check for consistency of measurements. This approach is used to establish:

a) Interrater reliability
b) Intrarater reliability
c) Concurrent validity
d) Construct validity

Comments:

The correct answer is (a) since the checking for consistency belongs to reliability which is a prerequisite to validity but not a guarantee of it. Given that, three physical therapists have assessed the joint mobility then interrater reliability is the best choice. The options “b” is wrong because it is suitable for the reliability of the same tester. Options “c” and “d” are wrong because they are types of the validity. Academics should start by establishing the scientific aspect of the question. Regarding the educational aspect, this question belongs to comprehension-application category on the cognitive dimension and the procedural knowledge. The question and its answer options reflects higher critical thinking skills but within the frame of reliability and validity psychometric measures. The linguistic and formative aspects were satisfied. Cognizant academics should have the intended learning outcomes, course objectives and program objectives readily available. Experienced academics must make sure that the teaching strategies and classroom assessment methods are aligned with the intended learning outcomes. The interactive educational process should be under the umbrella of the vision and mission of the institution. The question number two showed difficulty and discrimination indices of 0.61 and 0.15 respectively. That can be interpreted as medium difficulty index and fair discrimination index. The question can be included with full faith in the future exams.

Q 3: If pediatric sample performed better on a dependent variable not as a result of the independent variable but because they are older, that will be considered threat to:

a) External validity in terms of setting
b) External validity in terms of interaction
c) Internal validity in terms of maturation

d) Internal validity in terms of selection

Comments:
The correct answer is (c) since maturation in pediatrics is considered as a threat to internal validity. The answer options “a” and “b” are wrong because they give external validity statements. The answer option “d” is wrong since getting older is not a selection threat to internal validity. A quick run of the SELF evaluation strategy has established the scientific aspect of the question. Regarding the educational aspect, it was satisfied and showed that the question lies within the analysis-evaluation spectrum and that reflects early higher order thinking skills. The linguistic and the formative aspects were satisfied. The question number three showed difficulty and discrimination indices of 0.61 and 0.3 respectively. That can be interpreted as medium difficulty index and good discrimination index. The question was excellent to differentiate more capable from less capable learners. The question can be included with full faith in the future exams.

The correct answer is (c) since both groups are examinee. The answer options “a” and “b” are wrong since they described interventional studies whereas patients received testing. The answer option “d” is wrong since double blind study needs the examiner also to be blinded. A quick run of the SELF evaluation strategy showed that the scientific aspect was established and the educational aspect was satisfied. The question lies within the analysis-evaluation spectrum. The question reflects higher order critical thinking skills needed by every learner to demonstrate distinguished research skills. The linguistic and formative aspects were satisfied. The question number two showed difficulty and discrimination indices of 0.07 and 0.07 respectively. That can be interpreted as hard difficulty index and poor discrimination index. The question must be revised and reformulated in the future exams due to its weak psychometric measures.

RESULTS:

Difficulty and discrimination indices:

Difficult index and discrimination index of the discussed questions were included. Difficulty index is a measure of item difficulty and it reflects the percentage of test takers who were capable of answering the given questions.
correctly. The difficulty index of 0.5 is desirable and is classified as a question of medium difficulty. The discrimination index is the best measure to differentiate between the performances of more capable against less capable students on a particular question. A discrimination index of 0.2 and above is desirable. The values of the discrimination and difficulty indices of the research MCQs were settled in the template created by Dr. El-gohary (Appendix 1).

Discussion:

SELF evaluation strategy proved to be feasible but robust when used by academics to create and formulate MCQs within the introductory research course. The majority of academics participated in the current study indicated that SELF evaluation strategy was self-explanatory and took them step-by-step when started to evaluate the accuracy of the created MCQs within the research course. There was a consensus among academics regarding the logical steps of having quality MCQs through establishing the scientific aspect, matching the educational level, and ensuring the linguistic and formative aspects of the questions. Inexperienced and novice academics reported having some difficulties when designing the MCQs particularly at the intermediate and advanced courses. They indicated that continuous education courses and workshops would be essential to develop their academic skills. Experienced academics confirmed the need for holding a series of training on using SELF evaluation strategy particularly it has been recently published in literature. The majority of academics participated agreed upon the benefit achieved from the series of academic papers published by the scholar Dr. El-gohary discussing the guiding principles for designing robust MCQs & SELF evaluation strategy for improving the quality of MCQs. The scholar has also included feasible blueprint matrix in addition to feasible difficulty and discrimination index template. Academics are in agreement with the great benefit of having these research tools. Academics added that the quality educational papers offered a paradigm shift in the way academics design their MCQs. Academics admit that they are currently having a framework that enable them to test their prospected MCQs. Academics praised the interactive format of the included research MCQs and its benefit for academic training. Academics indicated that the included MCQs concord with the content areas of the research course and are in harmony with the course and program objectives. Moreover, it aligns with the intended learning outcomes and teaching strategies. In essence, the research MCQs included should be adequate to test learners’ competence and capabilities within the research course.

Acknowledgment
The authors would like to thank all staff members for facilitating research work.

Conflict of Interest
The authors declare no conflict of interest.

REFERENCES

1) Portney LG, Watkins MP. Foundations of clinical research. Application to practice. 3rd ed. Upper Saddle River, NJ: Prentice Hall, Inc.;2009

2) Thomas J, Nelson J, Silverman SJ. Research methods in physical activity. 7th ed. Human Champaign, IL: Kinetics Publishers, Inc;2015.

3) El-gohary TM. Multiple choice questions exams: Guiding principles to develop
exam designing skills among novice& inexperienced academics. IJTRR. 2017;6:42-48.

4) Meyers NM, Nulty DD. How to use (five) curriculum design principles to align authentic learning environments, assessment, students’ approaches to thinking and learning outcomes. Assess Eval High Edu. 2009; 34:565-577.

5) Vuma S, Sa B. A Comparison of clinical-scenario (case cluster) versus stand-alone multiple choice questions in a problem based learning environment in undergraduate medicine. J Taibah Univ Med Sc. 2017;12:14-26.

6) El-gohary TM. Physical Therapy Profession: The vision of 2020 from illusions to delusions. IJTRR. 2017;6 (1):1-4.

7) El-gohary TM. An introduction to scientific, educational, linguistic& Formative (SELF) evaluation strategies for a subject-specific multiple choice questions exams. Int J Chem Tech Research. 2017; 10:650-657.

8) El-gohary TM. Practical examples of using different levels of Bloom’s taxonomy on a sample of multiple choice questions chosen from biomechanics& kinesiology exams. Int J Chem Tech Research. 2018;11:234-241.

9) El-gohary TM, Al-shenqiti AM, Abdelkader SM. Scientific, educational, linguistic& formative (SELF) evaluation strategy of a sample of intermediate musculoskeletal multiple choice questions. Int J Chem Tech Research. 2018;11:341-347.

10) El-gohary TM, Awadallah MF, Abdelkader SM. Scientific, educational, linguistic& formative (SELF) evaluation of a sample of pulmonary multiple choice questions. Int J Chem Tech Research. 2018;11:355-361.

11) El-gohary TM. Dr. El-gohary blueprint matrix: A feasible template purported to be a quick and simple guide for academics to design robust multiple choice questions exams. IJTRR. 2017; 6: 38-44.

12) Al-Rukban MO. Guidelines for the construction of multiple choice question tests. J Family Community Med. 2006; 13: 125–133.

13) Bloom’s taxonomy. Available at: https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/ Accessed on January 20th- 2018.

14) Bloom's Taxonomy of Educational Objectives. Available at: http://teaching.uncc.edu/learning-resources/articles-books/best-practice/goals-objectives/blooms-educational-objectives. Accessed on January 13th, 2018.

15) Pande SS, Pande SR, Parate VR, Nikam A, Agrekar SH. Correlation between difficulty& discrimination indices of MCQs in formative exam in physiology. SEAJME. 2013;7:45-50.

16) Taib F, Yusoff MSB. Difficulty index, discrimination index, sensitivity and specificity of long case and simple choice questions to predict medical students’ examination performance. J Taibah Univ Med Sc. 2014;9:110-114.

17) Hogan TM, Olade TO, Carpenter CR. A Profile of Acute Care in an Aging America: Snowball Sample Identification and Characterization of United States Geriatric Emergency Departments in 2013. Acad Emerg Med. 2014;21:337-346.

18) Ten Critical thinking and clinical reasoning - Pearson Higher. Available at: https://www.pearsonhighered.com/conten
19) Critical thinking skills. Available at: https://www.skillsyouneed.com/learn/critical-thinking.html. Accessed on January 3rd, 2018.

20) Lucarelli PR, Lima MO, Lima FPS, JG de Almeida JG, Brech GC and D’Andréa Greve JM. Gait analysis following treadmill training with body weight support versus conventional physical therapy: a prospective randomized controlled single blind study. Spinal Cord 2011; 49, 1001–1007.

21) Rethans JJ, Norcini JJ, Baron-Maldonado M, Blackmore D, Jolly BC, LaDuca T, et al. The relationship between competence and performance: Implications for assessing practice performance. Med Educ. 2002; 36:901-909.
Appendix 1:
Examiner’s Name: ___________________ Date: ___________________

**Difficulty Index & Discrimination Index**

Subject Title: Introduction to Research
Subject #: Xxxxx

| Discrimination Index | Difficulty Index |
|----------------------|------------------|
|                      | HARD (0-0.29)    |
|                      | MEDIUM (0.30-0.79) |
|                      | EASY (0.80-1)    |

**Question Numbers**

| Poor < 0.1 | Q1 | Q4 |
|------------|----|----|
|            |    |    |

| Fair 0.1 to 0.29 | Q2 | Q3 |
|-----------------|----|----|
|                 |    |    |

| Good > 0.30     | Q1 | Q4 |
|-----------------|----|----|
|                 |    |    |

To be avoided | Acceptable | Good
---|---|---

NB. Discrimination index of ≥ 0.2 is desirable and difficulty index around 0.5 is also desirable.