International Journal of ChemTech Research
CODEN (USA): IJCRGG, ISSN: 0974-4290, ISSN(Online):2455-9555
Vol.11 No.05, pp 348-352, 2018

Difficulty and discrimination indices of sample of typical multiple choice questions administered to students at different levels showed atypical values

Tarek M. El-gohary[1,2]*, Sameh R Ibrahim[2,3], Samiha M. Abdelkader[4]

[1)] Biomechanics Department, Faculty of Physical Therapy, Cairo University, Egypt
[1)] Board Certified Orthopedic Clinical Specialist, USA
[1)] Mechanical Diagnosis& Therapy, McKenzie Institute, USA
[2)] College of Medical Rehabilitation Sciences, Taibah University, Saudi Arabia
[3)] Department of basic science, Faculty of physical therapy, Cairo University, Egypt
[4)] Physical Therapy Department, College of Applied Medical Science, King Saud University, Saudi Arabia

Abstract : Objectives: To compare the statistical indices obtained from exposing students at different academic levels to the same multiple choice questions (MCQs). **Methods**: Samples of MCQs administered simultaneously to students at level one and level five within the physical therapy program. Questions were classified as high quality MCQs. A difficulty and discrimination indices were achieved for all questions. The MCQs were analyzed for their reliability using Kuder- Richardson-20 (KR-20). The results of MCQs extracted from level one and level five were compared. **Results**: The analysis showed disparity in the difficulty and discrimination indices for most of MCQs. The difficulty levels were classified based on Bloom's taxonomy of cognitive processes dimension. The discrimination indices were classified on a spectrum from poor to fair to good with having a template specifically designed to easily settle the MCQs at its corresponding cells. **Conclusions**: It is concluded that typical MCQs are more likely to give atypical values due to different backgrounds, experience, retention abilities and academic course priorities among students.
Keywords : Difficulty index; discrimination index;high quality MCQs; question analysis.

Introduction

Multiple choice questions (MCQs) exams are the gold standard for testing knowledge acquisition. [1-3] MCQs are the most commonly format for academic and professional testing since they provide an objective assessment with the least likely bias from examiners.[2,4] Academics primarily ensure the validity of administered MCQs through having its discrimination and difficulty indices.[2,5,6] Academics use valid MCQs to gradually challenge test takers from lower order thinking skills to higher order thinking skills.[7-10] Academics need to repeat specific MCQs through courses taught at different levels to address different intended learning

International Journal of ChemTech Research, 2018,11(05): 348-352.

DOI= http://dx.doi.org/10.20902/IJCTR.2018.110538

objectives; however the students' response is undetermined.There is a gap in the body of knowledge regarding the consistency of the performance of test takers study at different levels exposed to the same MCQs. There is also uncertainty among academics regarding if students at advanced level should perform better than first level when exposed to the same MCQs. It is hypothesized that students study at advanced level will perform superior to students at first level when exposed to the same MCQs.The objective of the current educational paper is to compare the difference, if any, in the academic performance between test takers study at different levels but exposed to the same MCQs.

Materials and Methods

Sampling

A retrospective descriptive analysis of miscellaneous sample of MCQs extractedfrom introduction to research course and case study course were carefully chosen to represent different intended learning outcomes stated at both courses. MCQs will be briefly evaluated by SELF evaluation strategy to ensure its scientific& educational soundness in addition to linguistic and formative correctness.[2]After satisfyingthe inclusion of the MCQs, a comprehensive analysis of difficulty and discrimination indices of every question will be reported for both levels.[2,5,6]Kuder- Richardson-20, (KR-20) test was used to test item reliability.[11,12] The MCQ items were analyzed for level of difficulty, discrimination index, item distractors and test takers performances.The author will discuss a real sample of MCQs that were administered to level one and level five students at physical therapy program (Appendix I). Students at either level were blind that the same questions were given to the other level.The statistician who analyzed the questions was blind to the nature of the questions administered and to which group and at what level.[11,12]No ethical approval was needed since no human subjects were directly involved in the study.

Results

Difficulty and discrimination indices[2,5,6]*:*

Discrimination index and difficulty index of the discussed questions were included (Table I). The discrimination index is the best measure to differentiate between the performances of more capable against less capable test takers on a particular question. A discrimination index of 0.2 and above is desirable.Difficulty index is a measure of item difficulty and it reflects the percentage of students who were capable of successfully answering the given questions. The difficulty index of 0.5 is classified as a question of medium difficulty and is desirable. A negative discrimination index may indicate that certain answer option is measuring something other than what the rest of the options are measuring.An item discrimination of 0 means the same number of test takers from each group chose the answer, so the answer doesn't discriminate at all. MCQs where every student got the correct answer will always have a discrimination of zero.

Table I. Difficulty and discrimination indices of MCQs from two different courses

	Difficulty Index		Discrimination Index	
	Level One	Level Five	Level One	Level Five
Q_1	0.91	0.64	0.08	0.21
Q_2	0.5	0.5	0.25	0.21
Q_3	0.5	0.42	0.08	0.21
Q_4	0.91	0.21	0.08	0.21
Q_5	0.5	0.85	0.25	0.07
Q_6	0.75	0.85	0.16	0.14
Q_7	0.5	0.57	0.16	0.07
Q_8	0.08	0.42	- 0.08	0.21

The values of the discrimination& difficulty indices of the included MCQs were settled in the template created by Dr.El-gohary (Appendix II). The upper and lower case of the "q" letter were used to represent the values of the indices from introduction to research and case study courses respectively.

	***Appendix I*:** Multiple Choice Questions
	Q1: The *best research approach* has to be: a) Reliable and balance between internal and external validity b) Reliable with particular emphasis on internal validity c) Valid and emphasis on interrater reliability d) Valid and balance between internal and external reliability
	Q2: The best way to *overcome the order effect* as a threat to internal validity is through: a) Measuring more than one time b) Counterbalancing the order of measurement c) Measuring at different clinical setting **d)** Measuring by three different testers
	Q3: Measuring only elderly subjects between the age of 60 and 85 years old and community ambulant *is best to indicate*: a) The study validity and reliability b) The study internal and external validity c) The study inclusion and exclusion criteria **d)** The study confidentiality and randomization
	Q4: Participants' *complete understanding* of the procedures, risks and benefits, and demands that may be encountered is part of the: a) Confidentiality b) Randomization c) Informed consent **d)** Reliability
	Q5: Regarding five times sit to stand, leaving 3 *to 5 minutes to rest between 1st and 2nd outcome measurement is best* to be considered: a) When is being taken from published article in peer reviewed journal b) When is being taken from public sites on the internet c) When is being recommended by group of class students d) When is being recommended by friends or personal preference
	Q6: Regarding SLS balance testing, setting one minute as *ceiling effect* is best described as: a) The average time the participant is allowed to maintain b) The median time the participant is allowed to maintain c) The minimum time the participant is allowed to maintain d) The maximum time the participant is allowed to maintain Q7: Measuring only one subject and *falsifying the measurement for the other subjects* by intentionally putting any close time outcome is best described as: a) Correlation b) Regression c) Probability value d) Scientific misconduct
	Q8: A student managed to balance on right leg during SLS for 28 seconds, 30 seconds, and 32 seconds. The tester recorded 30±2.9 in the chart. The time recorded indicates: a) The mean± SD b) The median± SD c) The mode± SD **d)** The range± SD

Discussion:

The results of the study showed different values derived from the difficulty and discrimination analyses of the typical MCQs administered to students study at different levels of the same program. There was no specific pattern of response for most of questions despite exposing students to the same MCQs. Students had the same course activity which emphasized on having students to record some functional outcome measures

from individuals who satisfied the inclusion criteria. However, MCQs that tested students' competence regarding the course activity showed disparity in response.Regarding the difficulty index of question number four, it was easy for level one but was hard for level five. In reference to discrimination index, it had very poor index for level one but fair index for level five. The disparity in students' performance could be attributed to the way that students set their priorities when studying certain courses. In this scenario, students at level one focused on specific of research whereas level five students focused on specific of case scenario despite having both levels exposed to the same experience. Students at level five had the same difficulty indices for question five and six but poor and fair discrimination indices respectively. In reference to question number eight, students at level one showed hard difficulty index and very poor discrimination index of negative value. The negativeitem discrimination means a higher proportion of less capable students chose the answer. This would be expected for an incorrect answer. However; a negative discrimination on a correct answer may indicate something is wrong, as more capable students are choosing an incorrect answer. A double check of question number eight confirmed its correctness so we should not be alarmed by the negative value which could be attributed to the possibility of guessing the answer by more capable students more than choosing the answer based on cognitive recognition. To make a question a good discriminator, academics must make sure that the correct answers have a high positive discrimination, and the incorrect answers have a negative discrimination.[5,6]Regarding question number seven, students at level one showed medium difficulty index and fair discrimination index; however students at level five showed medium difficulty index and poor discrimination index. The disparity in discrimination indices could be attributed to the disparity of students' ability to retain basic knowledge of research. Students at level five studied research in details four semesters ago and are less likely to retain the specifics compared with students at level one. Regarding question number two, students at both levels showed the same medium difficulty indices and fair discrimination indices. The harmony in performance by students at bothlevels could be attributed to the nature of the question that test lower order thinking skills since the question lays at the comprehension level of Bloom's taxonomy of cognitive domain.[1,7,8]Academics should make sure that MCQs included at any exam have satisfied all aspects of being classified as MCQs of good quality. Recently, Dr. El-gohary has published a series of high quality educational papers that guide novice academics step by step to design robust MCQs exams.[1] Academics have benefited from the scientific, educational, linguistic& formative (SELF) evaluation strategy, blueprint and difficulty and discrimination index template.[2,13]Academics must include high order thinking skills that communicate critical thinking skills.[9,10]Khan and Aljarallah[14] reported that MCQs format is effective in assessing higher order cognitive skills of medical students. Moreover, academics should include stand-alone as well as case cluster MCQs with ensuring good alignment of intended learning outcomes with the teaching strategies and assessment techniques.[15,16] Academics need continuous training to polish their exam designing skills that reflect learners' competence and capabilities.[2,17]In essence, academics should rely on high quality MCQs that have medium difficulty and fair to good discrimination indices when designing their exams.[2,5] MCQs of poor quality must be revised and amended before being considered.

Conflict of interest: The authors declare no conflict of interest.

Acknowledgments:

The authors would like to thank coworkers for supporting research work.

Study Funding: None

References

1. El-gohary TM. Multiple choice questions exams: Guiding principles to develop exam designing skills among novice& inexperienced academics. IJTRR. 2017;6:42-8.
2. 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; 10:650-7.
3. Al-Rukban MO. Guidelines for the construction of multiple choice question tests. J Family Community Med. 2006; 13: 125–33.
4. Campbell DE. How to write good multiple-choice questions. J Paediatr Child Health. 2011;47:322-5.

5. 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;9:110-4.
6. 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.
7. Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, et al.A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.New York: Longman; 2001.
8. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. BMC Med Educ. 2007;7:49.
9. Maudsley G, Strivens J. Promoting professional knowledge, experiential learning and critical thinking for medical students. Med Educ. 2000;34:535-44.
10. Kuiper R A,&Pesut DJ. Promoting cognitive and metacognitive reflective clinical reasoning skills in nursing practice: self-regulated learning theory. J AdvNurs. 2004;45:381-91.
11. Portney LG, Watkins MP. Foundations of clinical research. Application to practice. 3rd ed. Upper Saddle River, NJ: Prentice Hall, Inc.;2009
12. Thomas J, Nelson J, Silverman SJ. Research methods in physical activity. 7th ed. Human Champaign, IL: Kinetics Publishers, Inc;2015.
13. 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.
14. Khan MZ, Aljarallah BM. Evaluation of modified essay questions (MEQ) and multiple choice questions (MCQ) as a tool of assessing the cognitive skills of undergraduate medical students. Int J Health Sci. 2011;5:39-43.
15. 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.
16. 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; 34:565-77.
17. 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; 36:901-9.

*****