# ERRORS IN THE CONSTRUCTION OF MULTI-CHOICE QUESTIONS: AN ANALYSIS

**Zarrin Seema Siddiqui[1]**
[1]MD Education Unit, The University of Western Australia,Perth, Australia.

## ABSTRACT

**Background:** This study investigated the frequency of errors in the construction of multiple- choice questions used for the assessment of para-clinical sciences in medical course at the University of Western Australia.

**Methods:** : In all there were two hundred and ninety (290) questions reviewed to examine if they adhere to the uniformly accepted guidelines for writing multiple – choice questions. Of these, one hundred and fifty-one questions were already administered in examination so performance of distracters in those questions was also reviewed in addition to the guidelines for the construction of the questions.

**Results:** 38% of the reviewed questions did not adhere to general guidelines for their construction and 12% had more than one error. A placement bias was also observed in placing correct choices.

**Conclusion:** A number of flaws were identified in the review process which raises doubt about the use of objective test items as a true indicator of student learning. This review has led to changes in the assessment policy at Faculty level and Pre-examination reviews have been introduced to identify any flawed items before their administration in the examination. Regular series of workshops on item writing and item analysis have been arranged and a Faculty Teaching and Learning Grant were provided to create online resources on assessment.

**KEYWORDS:** Educational Assessment, Reliability of Results, Student Performance, Multiple Choice Questions.

*Corresponding Author*
**Dr. Zarrin S Siddiqui,**
Associate Professor, MD Education Unit
The University of Western Australia
E-mail: zarrin.siddiqui@uwa.edu.au

## INTRODUCTION

Multiple-choice questions (MCQs) are frequently used in medical exams because of positive psychometric characteristics, a long history of research evidence, versatility in testing more cognitive knowledge as well as ease of scoring[1-3].At thesame time constructing a good MCQ is itself a laborious and lengthy exercise requiring an hour to write a good question[4].There are general guidelines on how to construct an MCQ (Table1) however all MCQ formats are prone to construction errors and studies have identified flaws in MCQs[5].

MCQs are frequently used in online formative assessments and as part of summative assessment across the Faculty of Medicine, Dentistry and Health Sciences (FMDHS) at the University of Western Australia. These MCQs and other test items are written by the faculty members involved in teaching and are occasionally reviewed within schools responsible for that content. The Education Centre at FMDHS provides support as and when required in terms of conducting workshops and reviewing assessment results. A review of MCQs developed for the assessment of medical students was undertaken to determine the quality of multiple-choice questions and their adherence to unanimously accepted guidelines for writing MCQs[6-9].One area of the medical course was reviewed for convenience, however, the information or method of review can be applied to other areas.

## METHODS

Two hundred and ninety MCQs (one best type) were reviewed to see if they follow the generally accepted guidelines suggested for their construction in the literature. The questions assessed were created for the online assessment of para-clinical sciences students in the medical course and include disciplines of Pharmacology, Pathology, Immunology, Microbiology and Infectious diseases. The review was undertaken by the author who has expertise in educational assessment and is also a qualified medical specialist. The flawed items were further discussed with a team at the Education Centre to validate the errors.

The items were classified as standard or flawed. A standard item was one that followed the guidelines for item writing while a flawed item was one that violated these guidelines. For this study, we used the guidelines validated and supported by published research studies[10]. (Table 1)

**Table 1: General guidelines for constructing a multiple choice question**

| | |
|---|---|
| 1. | Construct each item to assess a single written objective |
| 2. | State the stem in simple clear language |
| 3. | Use as much as the information in stem rather than options but avoid irrelevant information |
| 4. | State stem in a positive form. If negatives are to be used make them bold and/or in uppercase |
| 5. | Do not use all of the above as one of the options |
| 6. | Use none of the above option with caution |
| 7. | Do not use double negatives i.e. negative stem as well as option |
| 8. | Avoid verbal clues that may lead students to select the correct choice. |
| 9. | Follow the normal rules for grammar and punctuation |
| 10. | Place item choices alphabetically or in logical sequence |
| 11. | Keep length of options equal |
| 12. | Avoid use of the specific determiners such as always , never etc.. |

Of the total items reviewed, one hundred and fifty-one test items had been already administered in the examination, so we were also able to review the performance of distracters within each test item in addition to the uniform guidelines

## RESULTS

### PRE-EXAMINATION REVIEW

One hundred and ten questions (38%) did not follow the general guidelines for developing test items and required modification. Thirty-four (12%) test items had more than one error in their construction. The errors commonly observed in writing test items are listed in Table 2 in the order of their incidence from highest to lowest.

Table 2. Frequency of common errors found in the construction of multiple choice questions (n=290 questions)

| | | |
|---|---|---|
| 1. | Stem structure requires rephrasing | 22% |
| 2. | Unfocused stem | 18% |
| 3. | Extraneous material in the stem or the options | 17% |
| 4. | Negatively phrased questions, noemphasis on negatives and using double negatives | 14% |
| 5. | Using all of the above as one of the options | 12% |
| 6. | Only four plausible options, incorrect option standing out | 8% |
| 7. | Mutually exclusive options | 5% |
| 8. | Length of options not equal | 4% |

### POST EXAMINATION REVIEW

All the item (n=151) that had already been used in earlier examinations with students were also reviewed for the performance of distracters. The common flaws identified on distracter analysis were:

**i. Implausible Distracters**
Each of the reviewed itemshad five options, thus there were 755 options in all. One hundred and ten options were not selected by any of the students and 65% of these 110 options were the fifth option which even a student with minimal knowledge of the subject could have easily eliminated as a choice.

**ii. More than one correct choice**
Three items had more than one correct choice. In one question it was obvious as the same option was repeated twice, while in the other two items it became obvious by looking at the options which were identical to each other and that either of the two couldbe selected as correct choice.

**iii. Distribution of correct choice**
A placement bias was observed in the distribution of correct choices among five options offered (Fig:1). The correct choices are more commonly placed between options B – D. The observed frequency of correct choice in each cell was compared with the expected frequency (20%) using chi

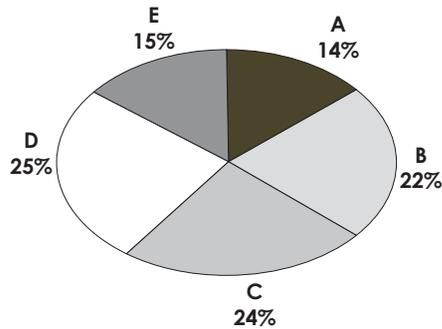square calculation. This yielded a value of 10.026(d-f=4) which was significant at 0.04 level.



**Figure 1. Distribution of correct answers among the five options for 290 multiple choice questions**

## DISCUSSION

A number of flaws have been identified in this review of test itemsspread equally between item stems and response options.

The errors identified in the stem construction can easily be corrected; however, when we look at the flaws in options, the scenario is different. Writing five options for the MCQs has been observed as a difficult task for the Faculty in this review. This compromises the effectiveness of the distracters as it is suggested that the response rate for each distracter should be around 20 – 30 %; a distracter with low response rate does not differentiate between students who have learnt the material from those who have not learnt.11 It is also observed from the post- examination analysis that students most of the time are choosing fromamong four options only because the fifth option is immediately eliminated. Literature suggests that having three good options is better than four options of which three are good and one option is so weak that it is obviously wrong[12-13].

Correct response on an MC test should be spread among all options positions. When guessing, students select the middle options more often than extreme ones. The academics tend to hide the correct response in the middle of the options and not first or last option[14].

Another error that was frequently observed and is relevant to be mentioned here was the actual format of the test item itself. Some of the questions do not seem appropriate for the format of MCQ (one best type). For example, some items are testing different outcomes within one item while in others there is a repetition of the same stem in a group of test items such as 'the mechanism of (drug name) is…. .'In such cases it is better to use a different form of objective test item like Multiple true/false type or Extended matching questions.

In late, item-writing templates called "item shells" were introduced. These facilitate the process of writing test items for novice writers[15].An item shell is a skeletal item containing a syntactic structure useful for writing sets of similar items[16].Writing high-quality test questions is a difficult task[17].A flawed MCQ interferes with accurate and reliable interpretation of test scores and also affects students' passing rate by adding construct-irrelevant variance (CIV) to the assessment[18].Construct-irrelevant variancemeans that the test has variables that are irrelevant to the measured construct thus affecting itemvalidity[19].This may help a student respond correctly because of extraneous clues or may make it more difficult for the studentsto answer correctly.

It is therefore important that Faculty is trained in item writing to improve the quality of test items[20 -21]. Once written, these should be critically reviewed. While pre-examination review focuses on the construction, relevance and validity of the test items, the post examination looks at how the test items have performed in the examination. Post examination assesses for items' level of difficulty (proportion of students who respond to the item correctly), discrimination (i.e. the ability to distinguish between students who are more versus less knowledgeable) and correlation of a student's correct response on the item with the total score.

An effective item review requires a team-based approach, which includes faculty member(s) involved in question design and selection as well as a member with expertise in test construction. The reviewed items may then be stored in item banks.An item bank is an efficient system for storage, maintenance and retrieval of test items with complete statistical information about each item and the corresponding learning outcome. With the advent of information technology,several software programshave been marketed to maintain item banks, and some institutions have developed their own. Few institutions have also collaborated jointly to share the item banks which would provide broader access to a larger pool of standardized items. It is especially feasible in health-related fields where the academics have responsibility of providing patient care along with the academic role.

## CONCLUSION

This study was a review of multiple-choice questions at a smaller scale and looking at only one aspect of assessment using convenience sampling. The analysis found that there are many flaws in the construc-

tion of test itemsin our faculty, which can make the questions ambiguous or ineffective for objective assessment. This interferes with making a valid assessment of students' learning. Therefore, in a high stakes examination, it is important to review and edit test items before and after administration.

This review has assisted the Faculty in recognising the importance of item review. It has now become common practice within the Faculty for questions of all types, not just MCQs, to be reviewed by both the discipline faculty and faculty of the Education Centre. Test Review Groups have also been established within other disciplines for interdisciplinary reviews. This not only allows the questions to be revised immediately but also nurtures a climate of peer support and a chance to look at the learning outcomes being assessed across disciplines. Regular workshops on item writing have also been introduced for faculty. These are typically scheduled during non-teaching weeks and take place at least once every semester. In addition, a lead academic in assessment provides one on one support to unit coordinators and is available when needed, which can help busy academics. These interventions have had a positive effect on the quality of test items used with our students, as evidenced by from subsequent reviews within the medical and other courses within the Faculty.

## ACKNOWLEDGMENT

## REFERENCES

1. Nendaz M, Tekian A. Assessment in problem - based learning medical schools: a literature review. Teaching and Learning in Medicine1999;11:232-243.doi:10.1207/s15328015tlm110408
2. Major CH. Assessing problem - based learning: a review and analysis of faculty-developed PBL based course portfolios. Journal on Excellence in College Teaching2000;11(2 & 3): 113 - 131.
3. Downing S. Assessment of knowledge with written test forms. In Norman G, van der Vleuten C,Newble.D. ed.International Handbook Of Research in Medical Education Dordrecht: Kluwer Academic Publishers. 2002; 656.
4. FarleyJ.Themultiple-choice question test: Writing the questions. Nurse Educator1989;14(6): 10 - 39.
5. Holsgrove G,Elzubeir M. Imprecise terms in UK medical multiple-choice questions: what examiners think they mean. Med Educ1989;32: 343 - 50.
6. Haladyna TM, Downing SM. Validity of a Taxonomy of Multiple-Choice Item-Writing Rules. Applied Measurement in Education1989;2(1): 51-78.
7. Gronulund N. and Linn R. In Measurement and Evaluation in Teaching New York, McMillan Publishing Company NY.1990.
8. Case S, Swanson D. In Constructing Written Test Questions for the Basic and Clinical Sciences Philadelphia, PA, National Board of Examiners.1998.
9. Masters J,Hulsmeyer B et al. Journal of Nursing Education2001;40: 25 - 32.
10. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-Choice item-writing guidelines for classroom assessment. Applied Measurement in Education2002;15(3):309 - 34.
11. Crisp GE, Palmer EJ. Engaging Academics with a Simplified Analysis of their Multiple-Choice Question (MCQ) Assessment Results. Journal of University Teaching and Learning Practice2007;4: 88 - 106. http://jutlp.uow.edu.au/2007_v04_i02/pdf/crisp.pdf.
12. King E. Constructing classroom achievement tests. Nurse Educator1978;3(5): 30 -36.
13. Rodriguez MC. Three options are optimal for Multiple-Choice items: A meta-analysis of 80 years of research. Educational Measurement: Issues and Practice2005;24(2): 3-13.
14. Cangelosi J. Designing Tests for Evaluating Student Achievement New York, Longman.1990.
15. Haladyna T. Developing and Validating Multiple-Choice Test Items New Jersey, Lawrence Erlbaum Associates, Inc.2004.
16. Haladyna TM,ShindollR. Item shells: A method for writing effective multiple-choice test items. Evaluation and the Health Professions1989;12: 97 - 104.
17. Schultheis N. Writing cognitive educational objectives and multiple-choice test questions. American Journal of Health Systems Pharmacology1998;55(22): 2397 - 401.
18. Stagnaro - Green AS. Downing SM.Use of flawed multiple-choice items by the New England Journal of Medicine for continuing medical education. Medical Teacher 2006;28(6): 566 - 68.DOI:10.1080/01421590600711153
19. BrualdiA.Traditional and modern concepts of validity. (1999). from http://www.ericdigests.org/2000-3/validity.htmAccessed on 24 July 2016
20. Naeem N, van der Vleuten C,Alfaris E. A. Faculty development on item writing substantially improves item quality. Adv. Health Sci. Educ. Theory Pract. 2012;17, 369–376.DOI: 10.1007/s10459-011-9315-2
21. Abdulghani HM, Ahmad F, Irshad M, et al. Faculty development programs improve the quality of Multiple Choice Questions items' writing. Scientific Reports2015;5:9556. PMC. DOI:10.1038/srep09556