JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH

ORIGINAL ARTICLE

Comparative Assessment In Pharmacology Multiple Choice Questions Versus Essay With Focus On Gender Differences

PAI M R S M *, SANJI N**, PAI P G ***, KOTIAN S ****

ABSTRACT

Background: Assessment is a driver of innovation and changes in education. In this study, medical student performances in pharmacology short essay questions (SEQ) and multiple choice questions (MCQ) were assessed and compared and gender differences were also explored.

Methods: A total of 712 second year students with records of scores in SEQ and MCQ papers were analyzed. The performances of students who scored > 60% marks in theory papers were compared with those who scored > 60% marks in MCQ and vice versa. Gender differences in performances were also studied. Statistical analysis was done using McNemar, Chi-Square and Pearson's co-efficient of correlation. P values < 0.05 were considered to be significant.

Results: The analysis showed that the probability of students scoring > 60% marks in SEQ papers was higher when they had scored > 60% marks in MCQ papers (P < 0.001). Their SEQ performance positively correlated with their MCQ performance (r= 0.768, r= 0.599 and r=0.767 for the first, second and third tests respectively). Females scored higher than males in all tests.

Conclusion: Literature indicates that MCQ as an assessment tool is debatable. This study indicates: performance in MCQ is an effective tool to predict SEQ paper performance; and correlates with global data on better performances by female counterparts.

Key Words: Pharmacology, assessment, essays, multiple choice questions

Key Messages:
1) Authentic assessment is inseparably linked to constructivist learning.The MCQ format of assessment is an effective, time saving tool of assessment in Pharmacology. The judicious use of MCQ can give a better insight into the student's understanding, application and the synthesis of knowledge.
2) The probability of students scoring > 60% marks in SEQ was higher when they had scored > 60% marks in the MCQ papers (P<0.001). The SEQ performance positively correlated with the MCQ performance.
3) The essay question format correlated poorly when compared with the MCQ pattern of assessment. Factors which were not in favour of essays included the idiosyncratic behaviours of examiners in the distribution of marks which were awarded for essays, language problems, the emotional maturity of students, poor, illegible handwriting, etc.
4) Female students consistently have outperformed their male counterparts and the data correlated with the data across the globe on gender differences in student performances.

_____

*(M.D.) (Pharmacology), **(M.D.) (Pharmacology)
***(M.D.) (Pharmacology)
*,***Department of Pharmacology, ****Department of Community Medicine, Kasturba Medical College, Mangalore, Manipal University.

 ****(M.Sc.)(Biostatistics),
**Department of Pharmacology, JJM Medical College, Davangere, Karnataka

Corresponding Author:

Dr. Preethi G Pai
Department of Pharmacology
Kasturba Medical College, P.B. No.53,
Light House Hill Road, Hampanakatta, Mangalore-
575001 Karnataka, (India)
E-mail: meddocpai2@yahoo.com;
Tele: Office: 91-0824-242227; **Resi-**91-9880750040
Fax: 91 824 2428183

Introduction

Review of the medical school curriculum and the evolutionary changes in curricula based on these reviews are compulsory requirements in medical schools across the globe. Some reasons which necessitate these innovations are: changing disease patterns; translation of medical education in health care professionals to nation building; exponential rise in medical student numbers globally and their optionless requirement of the highest quality of medical education; besides the desire for change as a primary requirement in man's endeavors. To stand the test of purposeful effective change that translates to better patient care, these educational innovations need to be assessed at the student level, to know if the service delivery component has been conveyed appropriately through the curriculum. Care has been taken that all faculty involved in delivery of the curriculum to students have been trained at the Medical Education Unit of the university, ensuring proper teaching skills.

Assessment is a powerful driver of innovative changes in education and defines goals for learners and teachers [1]. Student learning is driven by assessment and assessment is important to the student's experience [2]. In this century, momentum in every module of activity has hastened, leaving little time for answering essay questions meticulously as an assessment. Other reasons cited for its gradual disappearance include the lowest correlation with overall performance versus multiple choice question (MCQ) / clinical examination [3],[4] and low levels of reliability with generalizability [5]. The process of structuring and answering MCQ and objectively structured short essay questions (SEQ) has come of age in a majority of the medical schools in India. This is of relevance in developing countries as these modules are cost effective on overstretched staff time (not withstanding a requirement of breadth in their expertise) and institutional finances. Gender differences in student performances have been

reported. Studies show that in undergraduate medical examinations, female students performed significantly better than male students [6],[8].

The objective of this study was to evaluate student performances in MCQ and SEQ and to assess the predictive value of each one on their performance in the other component. Undergraduates who fetched more teaching time competing with lesser paid attention to parallel postgraduates formed the cohort of this study. The more powerful of these tools may then be used to predict later grades and its transition to clinical skills. Performances between genders have also been studied.

Methods

The study was an observational, retrospective study of written pen and paper type assessment at formative examinations that utilized a convenience sample of 200-250 second year medical students, at each of three consecutive academic years 2004-2005, 2005-2006, and 2006-2007. Student performance data (percentage marks) for each assessment module – MCQ and SEQ of these years were acquired from the central student records. Percentages for each component of these modules at summative examinations of the university were not available and hence, were not included in the study. The pharmacology course in Manipal University is offered during the second year of the students' MBBS curriculum whose duration spans a total of 18 months and is divided into three semesters; each of 24 weeks. The course contents include: general and clinical pharmacology, autonomic nervous system, cardiovascular system and diuretics in the first semester (hours allotted -40); central nervous system, gastrointestinal tract, blood, autacoids and respiratory system in the second semester (hours allotted -47); endocrine system, chemotherapy, immunomodulators, rational drug use, essential drug list and P drug in the third semester (hours allotted -45). The students received didactic lectures with teaching aides three hours a week and practical sessions inclusive of tutorials and seminars two hours a week.

The written theory paper consisted of two components: part 1 MCQ where four options

are given for a question and the single best answer was to be picked; part 2 theory paper of short essay questions (SEQ). The MCQ was timed at 30 minutes for 40 questions with 20 marks. (0.5 marks for the correct answer and 0.166 marks deducted for a wrong answer). The SEQ was timed at 150 minutes for 60 marks, which included questions for 2, 3, 5 and 9 marks. Every effort was made to eliminate examiner bias when the theory paper was evaluated by the faculty. The MCQ answer sheet was evaluated by an optical scanner.

Scores obtained in MCQ papers and SEQ papers were analyzed for: primary outcome – correlation between the high score group (> 60%) in SEQ, who were also in the high score group, (> 60%) for MCQ and vice versa; and secondary outcome - gender differences in written examinations.

All scores were analyzed using the Mc Nemar test. Correlation coefficients ('r' value) between the MCQ and SEQ components were estimated by using Pearson's product moment method. In all analyses, P-values of less than 0.05 were considered to be significant. Statistical package SPSS (version 11.5) was used for the analysis.

The University's principles and procedures on research ethics were adhered to throughout the study, which was approved by the Institutional Ethics Committee. Data on student performances were presented, such that identification of individual performances was impossible, complying with the requirements of the Data Protection Act.

Results
A total of 683, 653, and 707 second MBBS students who had appeared for the MCQ and SEQ components of the first, second and third formative examinations between the years 2004 to 2007 were included for the study. Pearson's correlation coefficient (r) in the high – score group (> 60%), in the SEQ, who were also in the high – score group (>60%) for MCQ and vice versa were studied for the level of significance.

The number and (percentage) of the students who scored more than 60% marks in the SEQ

and who had also scored more than 60% marks in MCQ at three formative examinations is given in [Table/Fig 1].The performances of the high score group in the SEQ positively correlated with the high score group for MCQ (r = 0.768, 0.599, 0.767). P values < 0.001 were indicative of very high statistical significance.

(Table/Fig 1) Theory performance of students who scored > 60% in MCQ

Sessional examinations	Scores in MCQ component > 60%		Scores in theory component >60%	
First	257/683	(49.8%)	225/257	(87.5%)*
Second	109/653	(37.6%)	27/109	(24.8%)
Third	352/707	(16.7%)	256/352	(72.7%)*

*P < 0.001is considered as highly statistically significant

The number and (percentage) of the students who scored more than 60% marks in the MCQ and who had also scored more than 60% marks in SEQ at three formative examinations is given in [Table/Fig 2]. The performances of the high score group in the MCQ positively correlated with the high score group for SEQ (r =0.294, 0.167, 0.445). P values < 0.001 were indicative of very high statistical significance. However, the r values were lower, indicating a lesser degree of positive correlation between the high score group of SEQ in their ability to be in the high score group for MCQ as well.

(Table/Fig 2) MCQ performance of students who scored > 60% in theory

Sessional examinations	Scores in theory component > 60%		Scores in MCQ component >60%	
First	352/683	(51.5%)*	225/352	(63.9%)
Second	205/653	(31.4%)	27/205	(13.2%)
Third	326/707	(46.1%)*	256/326	(78.5%)

*P < 0.001is considered as highly statistically significant

Gender differences were analyzed for 712 students (406 males and 306 females). In the first, second and third formative examinations, 57.35%, 58% and 57.5% of female students scored more than 60% marks in the MCQ and SEQ components (P<0.001) [Table/Fig 3] Female students consistently scored better than their male counterparts.

(Table/Fig 3) Differences in MCQ and theory performances: a gender comparison

Sessional examinations	Male students scoring > 60% in both MCQ and theory	Female students scoring > 60% in both MCQ and theory
First	42.65%*	57.35%
Second	42%	58%
Third	42.5%*	57.5%

*P < 0.001is considered as highly statistically significant

Discussion

Increasingly, academicians and members of national boards are confronted with issues associated with the assessment of large numbers, arising from a combination of factors including higher student / candidate enrolments and the introduction of a trimester of study in many universities. The resulting pressures on increased staff time in marking and cost effective measures are causing many academicians to search for alternative forms of assessment. University teachers are making more frequent use of multiple choice questions as a matter of expediency and in some cases, the quality of the assessment is being neglected [9].

Various assessment methods used in the curriculum studied included essays/short essays, multiple choice questions (MCQ), practical exercises and viva voce/oral examination. These assessments are done periodically at regular intervals at the end of each semester (formative) and at the end of the course (summative) examinations. In this study, we correlated the performance of two components of written examinations: MCQ and SEQ. Our analysis revealed an equal degree of statistical significance between the two comparisons and demonstrated an acceptable degree of concordance between the SEQ and MCQ components. However, better performances at MCQ correlated higher with better performances at SEQ than vice versa. The SEQ performance is positively correlated with the MCQ performance in a similar content area, but the magnitude of this correlation is lower in predicting student performance at MCQ (based on Pearson's coefficient of correlation 'r' value). Prout et al and Shitu et al also predicted a similar correlation in student performance during the Biochemistry and Anatomy examinations, respectively [10],[11].Their data and the data of this study are in good agreement with previously reported findings [12],[13],[14],[15].

The efficacy of multiple choice questions as a tool of assessment has attracted considerable debate over the past few decades [16].They are commonly used to test acquisition of knowledge and basic understanding. They have proved to be popular as a way of providing assessment for large cohorts of students, as much of the marking can be automated and so feedback and results can be provided quickly, even for the largest of classes [17].Moreover, assessment is not affected by a student's ability to write. The main advantage of MCQ testing is its versatility. There are significant cost savings and it is a format that can provide precision where other measurement options may be lacking (e.g. observing performance or interviewing). Criticisms at MCQs on the other hand, tend to centre upon unreliability due to random effects [18], on the inequity of the format in terms of its bias towards certain socioeconomic or ethnic groups [19] and also on the depth of learning the format engenders (or lack thereof) [20]. They are widely scorned as "multiple guess questions". They have been deemed as a means of surface approach to learning. They are less good/ not effective for higher order thinking skills; reasoning cannot be seen [17]. To further the efficiency of MCQ as a predictor of overall performance and also to change the students' leaning towards a strategic style, the construct of the MCQ needs to be made robust by the training faculty in the nature and outcomes of assessment; grades to be provided on absolute scores (when items are made easier) or on relative scores (when items are made difficult); subject content and skills required for writing items avoiding verbosity; reducing the number of test items and avoiding dysfunctional distracters could be some possible remedial measures [21].

Essay tests are not without problems. Assessment may be dependent on the neatness of the handwriting and can be influenced by the length of the essay. Grades depend on writing skills and are subjective. They are labour-intensive and time consuming. Despite all this, virtues of essay and short answer tests also exist. Students need practice formulating arguments; expressing thoughts clearly and integrating ideas. They are excellent for promoting integrative and synthetic thinking. The advantages of short answer questions over

MCQs [22] and the recommendations to make greater use of them in medical schools have been suggested in some reports [23],[24].

In undergraduate medical examinations in the UK, female students have been found to perform significantly better than men [7]. McManus et al (1996) showed that male medical students were 1.65 times more likely to fail in at least one of their final examinations than female students [25]. Likewise, Acheson (1997) found that male undergraduate students at Belfast were over three times more likely to fail in the final examinations than their female counterparts [26]. McDonough et al (2000) showed that female medical students at Dublin were significantly more likely to achieve the honours standard in the final psychiatry examinations than their male peers [27].The success rate in the Japanese national examination for medical practitioners is significantly higher for females than males [28]. In the USA, women medical students outperformed men in the obstetrics and gynaecology examinations [8].This study confirms the findings seen among the Indian undergraduate medical students of Pharmacology. The results were unlikely due to examiner bias, as the MCQ were corrected through computer software and were then correlated with the performance in theory. The reason for this consistent gender difference is not entirely clear. It has been suggested that women may be more diligent in their studies [86],[26].To add strength to this analysis, in our medical school, the number of male students taking more than one attempt at the summative examinations has been consistently more than their female counterparts over the last five years.

Thus, to conclude, the analysis of performances in the Pharmacology written examinations revealed that students scoring better in the MCQ paper invariably also scored better in the theory paper. The female students consistently outperformed their male counterparts. The implications and utility of this data are: students who were better performers in MCQ also performed better in SEQ with a very high correlation value. These better performers in MCQ can subsequently be subjected to the more time consuming, important assessment modules such as SEQ

and/or clinical examinations and /or viva voce, whichever is applicable or feasible.

Secondly, the successful candidates, while going through all assessment modules, will do so each time in decreasing numbers; as only the better performers (>60%) are picked for the subsequent assessment and they will be more completely assessed each time.

Thirdly, for those who eventually will come to the finals, this provides for an exposure to all modules; a holistic approach in assessment parameters, where particularly in the field of medical science, the value/importance of communication skills in all its forms with the patient is of utmost value.

Acknowledgements

References

[1] DFES (Department for education and skills) Towards a Unified e-Learning Strategy: consultation document, Nottingham: DFES Publications, 2003.

[2] Rust C. The impact of assessment on student learning: how can the research literature practically help to inform the development of departmental assessment strategies and learner-centered assessment practices? *Active Learning in Higher Education* 2002*;*3(2*)*:145-58.

[3] Robinowitz HK, Hojat M. A comparison of modified essay question and multiple choice question formats- the relationship to clinical performance. *Fam Med* 1989; 21 (5): 364-7.

[4] McCloskey DI, Holland RA. A comparison of student performances in answering essay- type and multiple choice questions. *Med Educ* 1976; 10 (5): 382-5.

[5] Tombleson P. Demise of the essay question. *Occas Pap R Coll Gen Pract* 1990; 46:23-4.

[6] Tyrer SP, Leung WC, Smalls J, Katona C. The relationship between medical school of training, age, gender and success in the MRCPsych examinations. *Psychiatric Bulletin* 2002; 26: 257-263.

[7] Pritchard, DJ. Effects of sex and alphabetical listing on examination performance of medical students. *Med Educ* 1988; 22(3): 205-210.

[8] Krueger PM. Do women medical students outperform men in obstetrics and gynecology? *Acad Med* 1998; 73(1): 101-102.

[9] Woodford K, Bancroft P. Multiple choice questions not considered harmful. ACM International Conference Proceeding Series; Vol. 106 Proceedings of the 7th Australasian conference on

Computing education- Volume 42. NewCastle, Australia;2005.p.109-116.

[10] Prout RES, Hoy TG. University examination performance of Dental students' correlation between entry qualifications and non clinical and clinical examinations .*Br Dent J* 1976; 141: 141 – 5.

[11] Shittu LA, Zachariah MP, Adesanya OA, Izegbu MC & Ashiru OA. The differential impact of various assessment parameters on the medical students' performance in the professional anatomy examination. *Scientific Research and Essay* 2006;1 (1):14-19.

[12] Jain S, Alkhawajah A, Larbi E, Al-Ghamdi M & Al-Mustafa Z.Evaluation of student performance in written examination in Medical Pharmacology. *Scientific J of King Faisal University (Basic and Applied sciences)* 2005; 6(1):1426-35.

[13] Lugman W, Ibrahim E. The outcome for some methods of evaluation of clinical students. *Mater Med Pol* 1987; 19: 60-1.

[14] Robinowitz HK. The modified essay question- An evaluation of its use in a family medicine clerkship. *Med educ* 1987; 21 (2): 114-8.

[15] Edelstein RA, Reid HM, Usatine R, Wilker MS. A comparative study of measures to evaluate medical students' performance. *Acad Med* 2000; 75 (8): 825-33.

[16] Williams JB. Assertion- reason multiple choice testing as a tool for deep learning: a qualitative analysis. *Assessment & Evaluation in higher education* 2006; 31(3):287-301.

[17] Epstein RM. Assessment in medical education. *N Engl J Med* 2007;356:387-96.

[18] Burton RF. Quantifying the effects of chance in multiple choice and true/false tests: question selection and guessing of answers, *Assessment and Evaluation in Higher Education* 2001; 26(1): 41–50.

[19] De Vita G. Cultural equivalence in the assessment of home and international business management students: a UK exploratory study. *Studies in Higher Education* 2002*;* 27(2): 221–231.

[20] Leamnson R. Thinking about teaching and learning. Sterling, VA, Stylus Publishing;1999

[21] Vyas R, Supe A. Multiple choice questions: A literature review on the optimal number of options. *Natl Med J India* 2008; 21 (3):130-33.

[22] Hettiaratchi ES. A comparison of student performance in two parallel physiology tests in multiple choice and short answer forms. *Med Educ* 1978; 12 (4): 290-6.

[23] Wakeford RE, Robert S. Short answer questions in an undergraduate qualifying examination: a study of examiner variability. *Med Educ* 1984; 18 (3): 168-73.

[24] Webber RH. Structured short answer questions: an alternative examination method. *Med Educ*1992; 26 (1): 58-62 .

[25] McManus I C, Richards P, Winder PC, Sproston KA. Final examination performance of medical students from ethnic minorities. *Med Educ* 1996;30**:** 195-200.

[26] Acheson AG. Do male students face prejudice? *Lancet*1997; 350**:** 964.

[27] McDonough CM, Horgan A, Codd MB, Casey. Gender differences in the results of the final medical examination of University College Dublin. *Med Educ* 2000; 34**:** 30-34.

[28] Matsubayashi K. Sex and examination results. *Lancet* 1997; 350**:** 1711.