RELATIONSHIP OF AWARDS IN MULTIPLE CHOICE QUESTIONS AND STRUCTURED ANSWER QUESTIONS IN THE UNDERGRADUATE YEARS AND THEIR EFFECTIVENESS IN EVALUATION

Junaid Sarfraz Khan, Osama Mukhtar, Saima Tabasum, Naveed Shaheen, M. Farooq, M. Abdul Irfan, Ajmal Sattar, M. Nabeel, M. Imran, Sadia Rafique, Maryam Iqbal, M. Sheraz Afzal, M. Shahbaz Hameed, Maryam Habib, Uzma Jabeen, *Malik Hussain Mubbashar

Department of Examination, University of Health Sciences, Lahore,*Vice Chancellor, University of Health Sciences, Lahore, Pakistan

Background: A number of evaluation tools for assessing the cognitive and affective domains in accordance with Bloom's taxonomy are available for summative assessment. At the University of Health Sciences, Lahore, Multiple Choice Questions (MCQs) and Structured Answer Questions (SAQs) are used for the evaluation of the cognitive domain at all six hierarch levels of taxonomy using the tables of specifications to ensure content validity. The rationale of having two evaluation tools seemingly similar in their evaluative competency yet differing in feasibility of construction, administration and marking is being challenged in this study. **Methods:** The MCQ and SAQ awards of the ten percent sample population amounting to 985 students in fifteen Medical and Dental Colleges across Punjab were entered into SPSS-15 and correlated according to the cognitive and affective level of assessment in relation to the Bloom's taxonomy and their grouping in the Tables of Specifications, using parametric tests. 3494 anonymously administered questionnaires were analyzed using ethnograph. **Results:** No statistically significant difference was found in the mean marks obtained by the students when MCQs and SAQs were compared according to their groupings in the Tables of Specifications at all levels of cognitive hierarchical testing. End-of-year cognitive level testing targets set were not met and more questions were set at the lower cognitive testing levels. Expenses incurred in setting MCQs and SAQs were comparable but conduct and assessment costs for MCQs and SAQs were 6% and 94% of the total respectively. In both MCQs and SAQs students performed better at higher cognitive testing levels whereas the SAQs and MCQs were able to marginally test the lower levels of affective domain only. Student's feedback showed that attempting MCQs required critical thinking, experience and practice. **Conclusion:** MCQs are more cost effective means at levels of cognitive domain assessment.

Keywords: Cognitive Domain, Affective Domain, Bloom's Taxonomy, Tables of Specifications, Evaluation

INTRODUCTION

The University of Health Sciences, Lahore (UHS) is strictly following the reforms introduced by the higher education commission Pakistan regarding examinations and as a result of this the university has developed a very high fidelity, high security, transparent and valid examination system which has been adjudged by international academicians as enviable.

Worldwide a number of modalities have been used for the end-of-year summative assessments.[1] In Pakistan, Long Essay Questions (LEQs) were the only means of gauging cognitive domain in the written examinations. In recent years LEQs are being widely replaced by Structured Answer Questions (SAQs) and Multiple Choice Questions (MCQs). It is generally believed that the latter two evaluation tools increase objectivity, allow greater coverage of curricular content and make it possible to construct elaborate yet precise Tables of Specifications.[2] However, it is also believed that these tools are more difficult to develop and administer.[2] Difficulty aside, these two tools are very similar in construction and evaluation process especially when the SAQs are constructed in a manner that each has multiple subcomponents.[3] Marking of SAQs is a more costly and subjective process when compared to the MCQs.[1,3] Therefore, the rationale of having two evaluative tools so similar in nature yet different in their requirements for marking and grading has previously been challenged.[4]

The purpose of the end-of-year summative assessment is to determine the level of acquisition of knowledge, attitudes and/or skills at the end of the academic year before promotion to the next academic session.[5,6] For this purpose, UHS administers written and oral/practical examinations. The written examination measures the effectiveness of the teaching/learning programmes of the University in introducing innovative changes in the system.

The education system in Pakistan is examination-centred. This means that mostly students study to pass the end-of-year examinations.[4] The onus of the University of Health Sciences, Lahore has

always been on encouraging students to gain knowledge, change attitudes and attain psychomotor skills necessary to become competent and safe professionals. To that end the University has developed Tables of Specifications, detailing the necessary levels of competency in all areas, in each subject. These Tables provide details of level of competency in each area related to Bloom's Taxonomy and their relative importance within the subject and to the student. Therefore, there might be topics that require assessment of factual knowledge yet within the same subject there might be topics that require analysis/comprehension and/or synthesis or assessment of changes in attitudes and acquisition of certain psychomotor skills. The rationale, is to tailor the learning and evaluation of students to the needs of the local, regional and international population and demands.[7,8]

Written examinations measuring knowledge and attitudes of students learnt over the previous year can be constructed using a number of assessment tools.[1,9] The three most commonly used tools by Universities in Pakistan are the Long Essay Questions (LEQ), the Structured Answer Questions (SAQ) and the Multiple Choice Questions (MCQ). University of Health Sciences now uses SAQs and MCQs in all Bachelor of Medicine; Bachelor of Surgery (MBBS) and Bachelor of Dental Surgery (BDS) Professional examinations. Before the University shifts emphasis from LEQs to objective tools of evaluation such as MCQs and SAQs in the postgraduate examinations as well, further research is warranted in determining the most effective and cost-effective evaluation modality in assessing the knowledge and attitudes keeping Bloom's Taxonomy as a guide.

The research project is aimed at determining the relationship of awards obtained by students in these two components and also the effectiveness of these methods as an evaluation tool in measuring cognitive and affective skills, two of the three components of Bloom's Taxonomy.

It is believed that MCQs and SAQS though different in construction and nature, measure the same entities quite similarly.[10] Whereas MCQs are entirely objective and can be evaluated using Optical Mark Reader Technology, SAQs require tedious manual 'subjective' assessment.[11]

The study will also provide an opportunity:

1. To determine the Affective and Cognitive level of each MCQ and SAQ constructed for Annual 2008 and 2009 examinations.
2. The response of students in relation to the level and difficulty index (DI) of each question.
3. The acceptability level of MCQ and/or SAQ components of theory examination amongst the student population.
4. To propose changes in the evaluation process based on the findings of the study.

There is no evidence of similar research at National level to date.

The objectives of study were:

1. To determine the correlation between the MCQ and SAQ component based on the Table of Specifications and Bloom's Taxonomy.
2. To determine the most cost-effective means of evaluation
3. To determine the effectiveness of the end-of-year evaluation process in relation to the Bloom's Taxonomy.
4. To evaluate the response of students in relation to the level and DI of each question.
5. To determine the Affective and Cognitive level of each MCQ and SAQ measured in Annual 2008 and 2009 examinations.
6. To gain student's perspective on the current system of evaluation and correlate it with the results of this study.
7. To identify areas requiring improvement in the current evaluation techniques and suggest solutions.

MATERIAL AND METHODS

The study was based on the results in the MCQ and SAQ components of written examinations in all Professional Annual 2008 and 2009 MBBS and BDS examinations. The MCQ and SAQ papers administered to students were analysed and compartmentalised in relation to the Tables of Specifications. Respective awards of each student in each subject of every Professional examination were entered into SPSS v. 15 and analyzed using Student's *t*-test.[12] The analysis included correlation of awards for each component of the Table of Specification in each subject. Next the degree of correlation between the MCQs and SAQs for each area of the Table of Specification in relation to Bloom's Taxonomy[13] was determined. For this, each MCQ or SAQ was graded depending on its level in the Taxonomy and compared with its counterpart. Interpretation depended on the analysis and relationship of each component of the MCQ and SAQ paper in relevance to the tables of specifications and Bloom's Taxonomy. A questionnaire to be filled anonymously was circulated to students sitting in these examinations, which was used to measure the acceptability and opinion of students regarding this system of examinations.

At the time of this study, there were 15 medical and dental colleges affiliated with UHS with

a total student population of 9,840. A total of 34,906 questionnaires were received (of all professional examinations in each of the eighteen subjects). Stratified random sampling was applied college, professional examinations and subject wise and 10% of the total population was included in the study. This meant that the sample population was 985 students. Since each subject is to be considered this means that a total of 3,494 questionnaires were analysed together with the individual awards of each SAQ and its compatible set of MCQs according to the Bloom's Taxonomy.

RESULTS

1. The correlation between the MCQs and SAQs component based on the Table of Specifications and Bloom's taxonomy:

Paired Sample *t*-test was applied on a sample of 3,494 and no significant difference was found in mean marks of SAQs and MCQs, (Table-1).

Table-1: Results of paired sample *t*-test of MCQs vs SAQs

		t-Statistics	*p*
Table of Specifications	SAQs vs MCQs based on Tables of Specifications	1.93	0.06
Bloom's Taxonomy	C1. Knowledge	1.92	0.06
	C2. Comprehension	1.82	0.69
	C3. Application	1.68	0.94
	C4. Analysis	3.25	0.07
	C5. Synthesis	1.75	0.08
	C6. Evaluation	1.44	0.15

2. The most cost-effective means of evaluation:

During the period of the study, the total cost of setting, conducting and assessing all MCQs and SAQs (from Paper setting to the final Paper assessment) was approximately Rs. 47 million (Cost of Paper Setting was approximately Rs. 0.14 million and Conduct and Assessment was approximately Rs. 46.86 million).

3. The effectiveness of end-of-year evaluation process in relation to Bloom's taxonomy:

Cognitive level targets set to assess the Cognitive Domain were not met as indicated in Table-2.

Table-2: Cognitive level targets met in the end-of-year Evaluation in Relation to Bloom's Taxonomy

Bloom's Taxonomy	Target to Set the Paper according to Bloom's Taxonomy	Paper set in 2008 and 2009	
		MCQS	SAQS
C1 & C2	33–34%	45.4%	51.3%
C3 & C4	33–34%	38.1%	37.2%
C5 & C6	33–34%	16.5%	11.5%

4. The response of students in relation to the level and difficulty index (DI) of each question:

Students' performed well in C5 and C6 category of Bloom's taxonomy. The results are shown in Figure-1.

5. The Affective and Cognitive level of each MCQ and SAQ measured in Annual 2008 and 2009 examination:

Total 3,514 MCQs were set during the study period, of which 816 (23.3%) were set at the level of C1 category of Bloom's taxonomy and 776 (22.1%), 780 (22.2%), 560 (15.9%), 300 (8.5%) and 280 (8.0%) were set at C2, C3, C4, C5 and C6 levels in the Bloom's Taxonomy, respectively. On the other hand, 888 SAQs were set, of which 250 (28.6%) were set at the level of C1 category of Bloom's taxonomy, 206 (23.2%), 196 (22.1%), 134 (15.1%), 80 (9.0%) and 22 (2.5%) were set at C2, C3, C4, C5 and C6 category of Bloom's Taxonomy, respectively. Only 350 (10 %) MCQs were set catering for A1 category in the Affective domain and 173(5 %) for the A2 category. No other category of Affective domain was covered by other questions. Similarly 10 % and 4 % SAQs were set at A1 and A2 levels of Affective domain respectively.

6. Students' perspective on the current system of evaluation:

Perception of the students on the current system of evaluation is given in Figure-2.

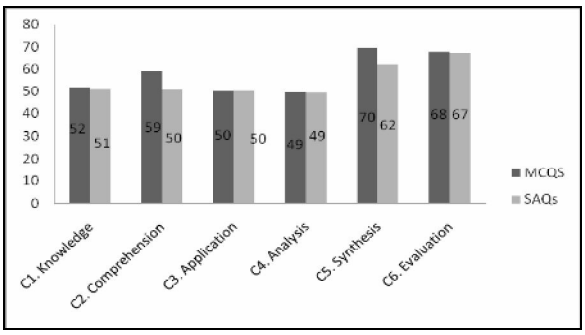

Figure-1: Marks obtained at various cognitive levels in accordance with Bloom's Taxonomy

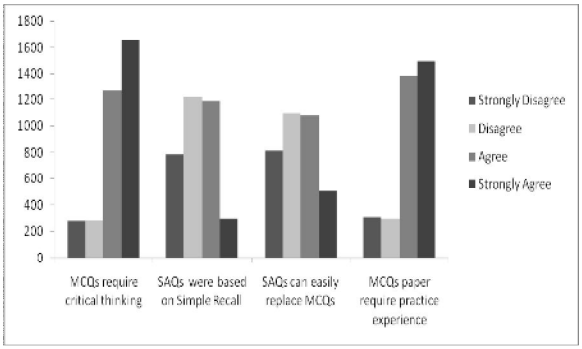

Figure-2: Students' perspective on the current system of evaluation

DISCUSSION

In the context of the present study, the Multiple Choice Questions (MCQs) and the Structured Answer Questions (SAQs) were to be compared

based on the tables of specifications and the Bloom's Taxonomy for effectiveness as a tool of evaluation including validity, reliability, cost effectiveness and feasibility; and acceptability amongst examiners and the administrators. Bloom published his taxonomy of cognitive learning based on a hierarchy of knowledge, comprehension, application, analysis, synthesis and evaluation.[14] In the Professional MBBS and BDS examinations held in 2008 and 2009 at UHS, each MCQ and SAQ in the evaluation instrument was designed to test at the same level of learning as the objective it was designed to assess in accordance to the Tables of Specifications (TOS) for any particular subject.

Our result showed that students were able to score equally well in both MCQs and SAQs at all six levels of Bloom's taxonomy. Therefore, a well constructed MCQ is just as good as an SAQ in testing higher levels of cognitive skills. The result is not dissimilar from that of a study by Mousumi Mukhopadhyay *et al*.[15] In this study Mousumi Mukhopadhyay *et al*. justified the use of MCQs for testing of higher cognitive skills in Medical Education. Moeen-uz-zafar and Badraljarallah[16] in their study described that constructing an SAQ was far more time and resource consuming than an MCQ, but a well-constructed MCQ was superior to an SAQ in testing higher cognitive skills of undergraduate medical students in a problem-based learning setup.

Our study, therefore, has confirmed the impression that well-constructed MCQs with high construct & context validity alone can be utilized to test knowledge as well as higher cognitive skills which previously were considered most suitably to be evaluated by SAQs and Long Essay Questions (LEQs).[17–21]

Moreover, our study has identified a huge gap in the feasibility in terms of cost & time of construction, administration and evaluation using the two evaluation tools. Whereas, the expenses incurred in constructing an MCQ and SAQ might be similar, the cost of administering & evaluating the SAQs far exceeds that for MCQs (Figure-3 and 4). The reduced effort in making and marking the MCQs might therefore be of benefit for institutions with limited faculty and monetary resources.[15, 22]

Table-2 shows that even though, the cost of administering and assessing SAQs was higher, comparatively fewer SAQs were constructed than MCQs for testing higher levels of cognitive skills. This result is in contrast to those of Webber[23] and Pai[24] who showed that the faculty found it easier to construct SAQs especially for testing higher levels of cognition than MCQs. Nevertheless, it was found that less MCQs were constructed to test higher cognitive levels than simple recall & knowledge. This trend was also identified by Epstein in their study published in 2007.[25]

The University of Health Sciences, since its inception in 2003 has focused on evaluation of higher cognitive skills. Assessment is a powerful driver of innovative changes in education and defines goals for learners and teachers.[26] Student learning is driven by assessment and assessment is important to the student's experience.[27] This, perhaps, is the reason for students performing much better in both the MCQs and SAQs when tested at higher cognitive levels as is evident from Figure-1. This trend was also identified by Pai.[24]

MCQs and SAQs were not very effective in evaluating most levels of affective domain. Studies by Moosa[28], Swanson[29], Newble[30] and Dauphinee[31] have reported similar results. Figure-2 gives the perspective of students on evaluation based on MCQs and SAQs and shows that the students believe that as an assessment tool MCQs are superior to SAQs provided students are trained in the art of attempting MCQs. Similar perceptives have been reported by Hettiaratchi[32] and Leamnson[33].

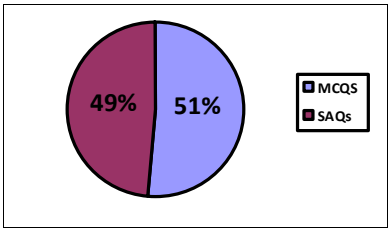

Figure-3: Expenses for paper setting

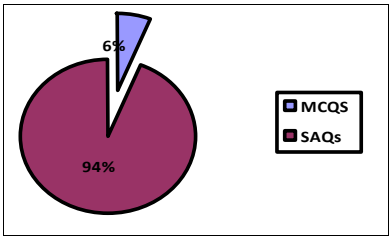

Figure-4: Expenses for conduct and assessment

CONCLUSION

Based on the results of this study, well constructed MCQs have been identified as a cost-effective, feasible, reliable and valid tool of assessment of all levels of cognition. MCQs have been found to challenge the students as much as SAQs. Therefore, in conclusion, we recommend that MCQs should replace SAQs as an assessment tool to reduce the time and monetary burden of evaluation in Punjab.

ACKNOWLEDGEMENT

The authors thank the Director Administration and Co-ordination of University of Health Sciences,

Lahore for his support in completing the project. This work was supported by grant of the Higher Education Commission, Pakistan.

REFERENCES

1. Schultheis NM. Writing cognitive educational objectives and multiple-choice test questions. Am J Health Syst Pharm 1998;55:2397–401
2. Writing Multiple-Choice Questions that Demand Critical Thinking. Available at: http://cit.necc.mass.edu/atlt/TestCritThink.htm#anchor1046303 [Accessed January 2nd 2008].
3. Dressel, PL, and Associates. Evaluation in higher education. Boston: Houghton Mifflin; 1961.
4. Airasian, PW. Classroom Assessment: Concepts and Applications (4th Ed.). New York: McGraw-Hill; 2001.
5. McMillan JH. (Ed.). New directions in teaching and learning: Assessing students' learning. No. 34. San Francisco: Jossey-Bass; 1988.
6. Gronlund, NE. How to write and use instructional objectives (6th Ed.) Upper Saddle River, NJ: Merril/ Prentice Hall; 2000.
7. Footlick JK. Truth and Consequences: How Colleges and Universities Meet Public Crises. Phoenix Ariz: Ammerica Council on Education; 1997.
8. Mayer D. Calamen DL. Gunderson A. Barach P. Telluride Interdisciplinary Roundtable. Designing a Patient Safety Undergraduate Medical Curriculum: The Telluride Interdisciplinary Roundtable Experience. Teach Learn Med 2009;21(1):52–8.
9. Tarrant M, Knierim A, Hayes SK, Ware J. The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. Nurse Educ Today 2006;26:662–71.
10. Shepard, LA. Evaluating test validity. In L. Darling-Hammond (Ed.),. Review of research in education (Vol. 19) Washington, DC: American Educational Research Association; 1993. p.405–50.
11. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions?: research paper. BMC Med Edu 2007;7(1):49.
12. Louis C, Lawrence M, Keith M. Research Methods in Education (6th Ed.) New York: Routledge; 2007.
13. Bloom Benjamin S, David R. Krathwohl. Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. Handbook I: Cognitive Domain. New York: Longmans, Green; 1956.
14. Bloom B, Englehart M, Furst E, Hill W, Krathwohl D. Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York, Toronto: Longmans, Green; 1956.
15. Mukhopadhyay M, Bhowmick K, Chakraborty S, Roy D, Sen PK, Chakraborty I. Evaluation of MCQs for Judgment of higher levels of Cognitive learning. Gomal J Med Sci 2010; 8(2):112–6.
16. Moeen-Uz-zafar, Aljarallah B. Evaluation of MEQ and MCQ as a tool for assessing the cognitive skills of undergraduate students at department of medicine, Qassim University. J Fam Community Med 2010;17(1):50–67.
17. Palmer EJ, Duggan P, Devitt PG, Russell R.. The modified essay question: its exit from the exit examination?
18. Marshall J. Assessment of problem-solving ability. Med Educ 1977;11:329–34.
19. Rabinowitz HK, Hojat MD. A comparison of the modified essay question and multiple choice question formats: Their relationships to clinical performance. Fam Med 1989;21:364–7.
20. Palmer EJ, Devitt P. Constructing multiple choice questions as a method for learning. Ann Acad Med Singapore 2006;35:604–8.
21. Collins J. Education techniques for lifelong learning: writing multiple-choice questions for continuing medical education activities and self-assessment modules. Radiographics 2006;26:543–51.
22. Rabinowitz HK. The modified essay question: an evaluation of its use in a family medicine clerkship. Med Educ 1987;21:114–8.
23. Webber RH. Structured short answer questions: an alternative examination method. Med Educ1992;26(1):58–62.
24. Pai MRSM, Sanji N, Pai PG, Kotian S. Comparative Assessment in Pharmacology Multiple Choice Questions Versus Essay with focus on Gender difference. J Clin Diagnostic Res 2010;4:2515–20.
25. Epstein RM. Assessment in medical education. N Engl J Med 2007;356:387–96.
26. DFES (Department for education and skills) Towards a Unified e-Learning Strategy: consultation document, Nottingham: DFES Publications; 2003.
27. 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.
28. Moosa MYH, Jeenah FY. The assessment of undergraduate psychiatry training: a paradigm shift. A S Psychiatry Re Rev 2007;10:88–91.
29. Swanson DB. A measurement framework for performance based tests. In: IR Hart and RM Harden, Editors, Further developments in assessing clinical competence. Montreal: Can-Heal; 1987.p.13–45.
30. Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. Med Educ 1996;22:325–34
31. Dauphinee, D. Determining the content of certification examinations. In: D Newble, B Jolly and R Wakeford, Editors, The certification and recertification of doctors: issues in the assessment of clinical competence, Cambridge: Cambridge University Press: 1994p. 92–104.
32. Hettiaratchi ES. A comparison of student performance in two parallel physiology tests in multiple choice and short answer forms. Med Educ 1978;12:290–6.
33. Leamnson R. Thinking about teaching and learning. Sterling, VA: Stylus Publishing; 1999.

Address for Correspondence:

Dr. Junaid Sarfraz Khan, Controller of Examinations, University of Health Sciences, Lahore, Pakistan. **Tel:** +92-42-99231218, Fax: +92-42-99231857

Email: junaidsarfraz@hotmail.com