

A subjective comparative analysis of quality of Short Answer Questions(SAQ) Items between Orthopedics and Other specialties used in internal examinations in MBBS course in Medical College of Lahore

Amina Husnain¹, Asif Khan², Tariq Aziz³, Faisal Nazeer Hussain⁴

¹ Associate Prof.
Medicine SIMS, Lahore

² Senior Registrar,
Avicenna Medical
College DHA9, Lahore.

³ Resident Orthopedics
Avicenna Medical
College

⁴ Professor of Orthopedics
Avicenna Medical
College Lahore

Authorship and contribution

Declaration: Each author of this article fulfilled ALL 4 Criteria of Authorship:

1. Conception and design or acquisition of data, or analysis & interpretation of data. 2) Drafting the manuscript or revising it critically for important intellectual content. 3) Final approval of the version for publication. 4) All authors agree to be responsible for all aspects of their research work.

Correspondence

Author:

Faisal Nazeer Hussain
Email:
fnhussain@hotmail.com

ABSTRACT

Objective: To compare the subjective quality of various items used in short answer questions (SAQ) papers between Orthopaedics and other specialties used in internal examinations in MBBS course in a Medical College of Lahore.

Methods: This retrospective descriptive study was conducted over a period extending from 25th October 2019 to 25th February 2020 at Avicenna Medical College Lahore. Previously used short answer questions (SAQ) papers (in internal assessments) of MBBS examination were collected from various departments of medical college. They were segregated in two groups as those from Orthopedic Surgery and those from Non Orthopedics subjects like general surgery, medicine and other subjects in the Miscellaneous group. SAQ in both the groups were assessed for the level of cognitive domain like C1(Recall),C2(Identify) and C3(Analyze) level and for clinically focused and relevant scenarios. The statements were evaluated whether scientifically correct and leading to an appropriate diagnosis without providing a lead into the question that would follow. All items were checked for mistakes in language or spelling and grammar. A comparison was made and *P* value calculated with Chi-square test. *P* value of < 0.05 was considered significant.

Results: A total of 12 Orthopaedic SAQ papers comprising of 84 items and 14 miscellaneous papers comprising of 101 items were evaluated and compared. The number of C1 questions in Orthopaedic SAQ papers were 24(13%),C2 questions were 32(17.4%) and C3 was 28(15.2%).In the miscellaneous group C1 questions were 33(17.4%),C2 questions were 31(16.8%) and C3 questions were 37(20.1%). The number of SAQ scenarios were focused in 74(40%) SAQ in Orthopaedics versus 48(25.9%) in miscellaneous group(*P* value 0.013).The quality of statement was clear in 56(30.3%) SAQ in Orthopaedics while only 14(7.6%) of the miscellaneous papers had clear statement. Spelling mistakes were detected in 12(21.1%) SAQ of Orthopaedics and 22(38.8%) in miscellaneous group. Grammatical mistakes were noted in 10(17.5%) Orthopaedics SAQ and 5(8.8%) in miscellaneous group.

Conclusion: The number of Orthopaedic SAQ with cognitive levels(C3,C4) were less than other specialty SAQ. However, the quality of scenarios were better focused and statement much clear in Orthopaedic SAQ than in miscellaneous group.

Key Words: Cognition, Quality, Short Answer Questions.

This article may be cited as:

Husnain A, Khan A, Aziz T, Hussain FN. A subjective comparative analysis of quality of Short Answer Questions(SAQ) Items between Orthopedics and Other specialties used in internal examinations in MBBS course in Medical College of Lahore. J Pak Orthop Assoc. 2020;32(2):

INTRODUCTION

Orthopedic surgery is taught as part of surgery in the under graduate curriculum. Most of the current

teaching practice lies midway somewhere between didactic teacher centered old school teaching and problem based, interactive, student centered

learning. As our teaching methodology evolves so should the assessment methods. Commonly written and oral methods of testing are used. Short answer questions(SAQ) are known to test creativity, spontaneity and innovative thinking. Most of the questions to test cognitive domain are generally divided into three types, i.e, C1 or recall based where the knowledge is tested as a fact, C2 or analysis where the student shows the depth of his understanding by inferring results from the given data (like to make a diagnosis) and C3 or application of the knowledge, e.g. to suggest a treatment strategy. The language and philosophical depth determine the quality of testing item so that it does what it is means to do. In formal examination based systems the student is part of the educational environment rather he is the major partner. The students are quick to pick what is expected of them in the examinations hence develop their learning methods which ensure success. This impacts their learning practices and strategy for preparation. This dilemma between proactive learning and teacher based learning still haunts the system as most of the present day teachers have themselves been taught and tested in the past through a system that favored C1 level response. A teacher who has not been trained and taught in teaching methodology and in developing testing items is less likely to produce assessment tools of higher grade especially in our under graduate courses. Moreover most young institutes have not yet developed periodic review and analysis of examination item formally(making large scale scientifically reviewed question pools). Lack of these formal introspective reflection strategies and analytical studies promotes low level testing practices by under graduate teachers and departments.

Resultant learning may be directed to passing the examination rather than learning the art of treating the patients.¹⁻³ Newer learning methods have been developed which may change the learning environment totally, e.g. online lectures through Moodle, Zoom, learning through YouTube videos, hypermedia and simulation devices. The way skill acquisition changes with these methods so does the assessment methodology. The questions asked in the examinations should be intellectually challenging and motivating for the modern day student so that he finds the testing process attractive. His interest can only be increased or captured if the quality of item used in examinations is of high standard and requires application of the knowledge that the student has gained. Items based on recall are easy and least challenging. Here it can be safely cited that the one

who writes these items matters more than generally acknowledged.

Shortcomings of Short Answer Questions(SAQ) are well known. The students may lose marks due to bad spellings and hand writing. They are only suitable for questions that can be answered with short responses necessitating the examiner is very clear on the type of answers expected when setting the item. The tools to assess the items such as discriminatory index, difficulty index and reliability calculations cannot be done for SAQs like multiple choice questions(MCQs).³ Examination items like MCQs and SAQs development is directly related to the experience and training of the faculty.^{3,4} SAQs need more time to solve, the stationary cost and logistics may be more expensive.⁴As SAQ is an open-ended question it lend the students freedom to answer any way they choose. SAQs are famous for difficulties in grading if the question is not worded carefully. It needs a clear headed and experienced teacher to set such papers. The main advantage lies in assessing knowledge but the danger being students may go to rote learning. This can only be prevented if the questions are phrased carefully avoiding recall for easy to answer C1 responses. A clever scenario can prevent this by elevating the item to C2 or C3 level.

The objective of our study was to compare the subjective quality of various items used in short answer questions(SAQ) papers between Orthopaedics and other specialties used in internal examinations in MBBS course in a Medical College of Lahore. The results of our study would be utilized to formulate uniform guidelines for Orthopaedics SAQ paper setting in our institution.

METHODS

This retrospective descriptive study was conducted over a period extending from 25th October 2019 to 25th February 2020 at Avicenna Medical College Lahore. We collected 12 SAQ papers from Orthopedics comprising of 84 items and 101 items from 14 SAQ of Surgery, Medicine and ENT and were analyzed in a uniform pattern.^{5,6} They were segregated in two groups as those from Orthopedic Surgery and those from Non Orthopedics subjects in the Miscellaneous group. All SAQs were reviewed for their cognitive levels and quality of phrasing the question. All papers used in the undergraduate clinical subjects were assessed by a single experienced medical teacher. No blinding was used. The cognitive level of the questions was determined by the modified Bloom's taxonomy⁶ as

C1(Recall),C2(Identify) and C3(Analyze) level. The quality of the question was assessed by seeing whether the data given lead to or matched the subsequent questions asked later were in line with each other. The clinical scenario Items were classified as flawed if they contained one or more than one flaw. The statement was assessed independently for being clear and focused. The questions to follow later were assessed for being relevant or suggestive of a likely diagnosis or containing clues to solution of problem. Scenario was checked to be organized, relevant, complete, clear and logically correct. Language mistakes were checked in all items in both groups by the same assessor. Items were classified as flawed if they contained one or more than one flaw. SPSS version 20 was used to analyze the data. A comparison of important variables between the Orthopaedics SAQ and miscellaneous group was made and *P* value calculated with the help of Chi-square test. *P* value of < 0.05 was considered significant. Data presented in tables where necessary.

RESULTS

In this study we reviewed 12 SAQ papers of Orthopedics comprising of 84 Items and 14 SAQ papers of miscellaneous subjects comprising of 101 items. Both groups showed fairly close proportions of C1, C2 and C3 items as shown in table I. However, it can be seen that the percentage of C2 and C3 items was 60 in Orthopedics versus 68 in the Miscellaneous group showing a slightly higher quality of items in miscellaneous group but statistically non significant (*P* value 0.600).The quality of scenarios were focused in 40%(n=74) Orthopaedics SAQ and 27.9%(n=48) in the miscellaneous group as shown in table II(*P* value 0.013). The quality of statement was clear in 56(30.3%) Orthopaedic SAQ and 14(7.6%) miscellaneous SAQ (*P* value 0.000) as shown in table III. Spelling mistakes were detected in 12(21.1%) SAQ of Orthopaedics and 22(38.8%) in miscellaneous group while grammatical mistakes were noted in 10(17.5%) Orthopaedics SAQ and 5(8.8%) in the miscellaneous group(*P* value 0.060).

Table I: Comparison of cognitive quality of SAQ items of Orthopedics with Miscellaneous papers.

Cognitive level	Orthopaedics SAQ		Miscellaneous SAQ	
	Frequency	Percentage	Frequency	Percentage
C1(Identify)	24	13%	33	17.4%
C2(Interpret)	32	17.4%	31	16.8%
C3(Analyze)	28	15.2%	37	20.1%

Table II: Comparison of quality of scenario in Orthopaedics SAQ and miscellaneous SAQ.

S.No	Quality of SAQ Scenario	Orthopaedic SAQ		Miscellaneous SAQ	
		Frequency	Percentage	Frequency	Percentage
1	Focused	74	40%	48	25.9%
2	Unfocused	07	3.8%	16	8.6%
3	Vague	16	8.6%	10	5.4%
4	Log, Clues	02	1.1%	08	4.3%
5	Data problems	02	1.1%	02	1.1%

Table III: Comparison of quality of statement of Orthopaedic SAQ and miscellaneous SAQ.

S.No	Quality of Statement	Orthopaedic SAQ		Miscellaneous SAQ	
		Frequency	Percentage	Frequency	Percentage
1	Clear	56	30.3%	14	7.6%
2	Diagnosis Suggested	10	5.4%	50	27%
3	Scenario Unclear	06	3.2%	18	9.7%
4	Confusion	12	6.5%	11	5.9%
5	No Scenario	--	--	08	4.3%

DISCUSSION

No examination system can be unanimously declared to be devoid of shortcomings. Every testing item must stand the test of reliability, validity, educational impact, cost effectiveness, and acceptability.¹ Reliability reflects the precision with which a score on a test is obtained. It may be different for multiple choice questions and short answer questions. The user may use them differently when phrasing a set of data or choosing the questions to ask. The practical usage expects us to maintain the validity which actually tests what it is expected to test. Phrasing of the data statement followed by suitable questioning through short answer questions makes the candidate think and apply his knowledge to understand, interpret and analyze the given facts before he takes up answering the question. This can only succeed if the problem is presented in quiz manner for the student to use higher thinking. We are yet to make an assessment method be it objective or subjective where an improvement is not possible. There are many studies which have shown that very few examination items especially MCQs are accepted by all without reservations.^{6,7} A higher level of cognitive reasoning which can discriminate between students who are high and low achievers is possible with high quality MCQs.^{2,4} It is known that good quality SAQs or MCQ are relatively difficult but not impossible to construct for example problem solving questions.⁸ This can easily be assumed for SAQs as well. The student has to be made to think, analyze deeply then only we can expect a response regarding management of the problem. In a quest for an ideal assessment method we can say that the distinction between objective and subjective assessments is neither useful nor accurate. Rather there is no such thing as "objective" assessment. It is said that all assessments are created with some inherent biases built into decisions about relevant subject matter and content, as well as cultural (class, ethnic, and gender) biases.⁹ Advantages of SAQs remain well known e.g. easy to make, used in both formative or summative assessment, familiar for students and are relatively fast to mark and can be marked by different assessors controlling examiner bias. Guessing is less likely to succeed as in MCQs.^{5,9,10}

Every specialty has its own contextual or content variations which make it unique and different from others. Medicine for example has to deal with almost all parts of the body and a multitude of pathophysiological phenomenon like variations in blood sugar can lead to changes in urinary output or

any change in metabolic response of the body may ultimately effect the respiratory system thence changes in acid base balance and blood gas levels may get to be asked in one scenario. Perhaps this was the reason that we noted more C2 and C3 based questions in Miscellaneous group (68% versus 60%) than in Orthopedics (table I) which is a specialty with low occurrence of problems affecting the whole body. Orthopedics scenarios are mostly related to trauma for undergraduates and multi organ involvement is less likely. In our study the cognitive level of Orthopaedic SAQ was C1 in 24(13%) while miscellaneous SAQ had 33(17.4%) items of C1 levels. Tariq and his colleagues⁸ observed that 83.3% of SAQs in their set up was C1 level. Most medical educationists agree that commonly practiced form of formative assessment is diagnostic assessment. Can the student make a diagnosis? While the contrast lies in Performance-based assessment which resembles summative assessment, as it focuses on achievement. World Health Organization⁵ in a 2001 quality assurance endeavor desires that "Student learning is driven by assessment and it is therefore essential that methods of student assessment match and reinforce the goals and objectives of the medical course" However since our data collection did not segregate the quality of items temporally we cannot comment on the quality of work improvement with experience. Tariq,⁸ however noted that the question made by faculty in internal examinations improved in quality with passage of time. We observed that when the data statement was analyzed for quality most of the data statements in Orthopedics were far better than 56/84(66.7%) and 14/101(14%) the comparative group(table III). The probable reason could be due to smaller size of the curriculum in Orthopedics and not much detailed or elaborate questioning was possible.

Language remains a common cause of problems in many examination items especially so in a society like ours where the mother tongue is not English. We have found that the Miscellaneous group showed problems in writing the items like spelling mistakes and grammatical faults. Item writing mistakes are common and well reported in literature. We noted far lesser mistakes of this kind than others i.e.28% than 46% in another study.^{11,12}

Our study was limited by its small scale and without having added the scores the students obtained in these SAQ papers. Moreover only one observer was asked to analyze the items. Further

well designed large scale studies are needed to confirm our results.

CONCLUSION

The number of Orthopaedic SAQ with cognitive levels(C3,C4) were less than other specialty SAQ. However, the quality of scenarios were better focused and statement much clear in Orthopaedic SAQ than miscellaneous group.

Conflict of Interests: None

Fundings/Grants: None

REFERENCES

1. Cantillon P, Hutchinson L, Wood D: Editors: ABC of learning and teaching in medicine: 2003. BMJ Publishing Group Ltd, BMA House, Tavistock Square, London.
2. Gunderman RB. 2nd ed. Achieving Excellence in Medical Education; Springer-Verlag London Limited 2006. P96
3. Zia ul Islam, Usmani A. Pak Psychometric analysis of Anatomy MCQs in Modular examination. J Med Sci. 2017; 33 (5):1143-1150.
4. Raina SK, Singh M, Sood A, Chander V. Assessment of need for capacity building in framing multiple choice questions for undergraduate medical students. J Sci Soc. 2016;43:11-3.
5. WHO Guidelines for Quality Assurance of Basic Medical Education in the Western Pacific Region 2006. Manila: WHO Regional Office for the Western Pacific.
6. Chan C. Assessment: Short Answer Questions, Available: Accessed: DATE April 19, 2020, Assessment Resources@HKU, University of Hong Kong[<http://ar.cetl.hku.hk>]:https://ar.cetl.hku.hk/am_saq.htm
7. Patil R, Vell K, Boratne AV. Evaluation of multiple choice questions by item analysis in a medical college at Pondicherry, India. Int J Community Med Public Health. 2016;3:1612-1616.
8. Tariq S, Tariq S, Maqsood S, Jawed S, Baig M. Evaluation of cognitive levels and item writing flaws in medical pharmacology internal assessment examinations. Pak J Med Sci. 2017;33(4):866-870.
9. Joint Information Systems Committee (JISC). "What Do We Mean by e-Assessment?" JISC Info Net, available at JISCinfonet.ac.uk. Retrieved April 19, 2020.
10. Tarrant M, Ware J. Impact of item-writing flaws in multiple choice questions on student Achievement in high stakes nursing assessments. Med Educ. 2008;42:198-206.
11. Buckwalter JA, Schumacher R, Albright JP, Cooper RR. Use of an educational taxonomy for evaluation of cognitive performance. J Med Educ. 1981;56:115-121.
12. Baig M, Ali SK, Ali S, Huda N. Evaluation of Multiple Choice and Short Essay Question items in Basic Medical Sciences. Pak J Med Sci. 2014;30:3-6.