# Learner Competence Across Cognitive Levels In Undergraduate Medical Education: An Item Response Approach.

Zeeshan A Rehman[1*], Syed Imran Mehmood[2]

ABSTRACT:

**Objective:** The current study endeavors to evaluate competence of learners using item response theory approach in undergraduate medical education.

**Methodology**: The responses of the examinees from first professional year (MBBS) were evaluated for Anatomy discipline to gauge learner competence through knowledge, understanding and application in the subject using item response theory approach (IRT) which is an effective tool to measure the validity (discrimination and difficulty index) of assessment, while KR-20 is used to gauge reliability of the items with statistical software STATA 17. The quality of question and level of examinee responses provide skill level and competence of learners.

**Results:** The current study found that the C1 and C3 level items from the Anatomy discipline are mostly acceptable and effectively gauge learner competence while flawed C2 level items need complete restructuring. The study also identified weak areas of average learner for items in each cognitive level.

**Conclusion**: The study found significant role of IRT in effectively estimating competence of the examinees across cognitive levels for each discipline and in building of an effective test pool through assessment of learner ability and construction of quality SBAs to achieve valid and reliable assessment tool.

**Keywords:** Difficulty index; Discrimination index; Single best answer questions; Cognitive levels; Item characteristic curve.

## Introduction:

Contemporary medical curriculum witnessed a remarkable shift from traditional to a more learner centered approach with a purpose to produce competent medical professionals. Besides teaching strategies assessment and feedback are also major constituents of educational curriculum to corroborate achievement of the learning outcomes, quality reassurance[1] and fulfillment of mandatory competence expected from learner to ensure patient safety.

Competency based Medical education (CBME) demands persistent assessment and feedback mechanisms to enhance the learning process and overcome the learner deficiencies [1,2]. In the context of life long and self directed learning approach, CBME focused on formative assessment to enhance knowledge and skill of the learner throughout the learning process[3] by continuously recognizing learner weak areas and providing chances to improve competencies and skill development[4,5].

According to Bloom Taxonomy[6], assessment should encompass three domains i.e. cognition, psychomotor and affective. The cognitive domain in bloom taxonomy[6] is divided into 6 cognition levels (C1-C6), differentiate into remembers (knowledge), comprehend, apply, analyze, evaluate and create (knowledge being the lowest level and the creation being the highest level). Lately, medical institutes are endeavoring to timely assess the competence of medical learners with reliability and accuracy[7, 2, 8, 9] with prime objective to improve all learners' potential and competence via motivational direction towards learning flaws. The conventional assessment often ignores the existence of presumption and false concepts in learner's understanding and focuses on acquiring memorizing and gaining knowledge(C1), while the cognitive assessment advocates for gauging both understanding (C2) and application (C3) beside knowledge by correlating various concepts for multiple scenarios in accordance with groups of interrelated concepts besides testing knowledge[10, 11.]

Assessment is deemed appropriate if the student could accurately judge one's knowledge, understanding, relevance and analysis with capability to appropriately appraise the learning content related to the subject being examined [12, 13,14], which significantly depends on educator efficiency especially in initial years of undergraduate medical education where educators from basic science departments' greatly influence learner's competence by interconnecting learner's knowledge with understanding and clinical application across different disciplines[15].

Single best answer question (SBAs) is a multiple choice assessment tool to evaluate learner's competence especially in health related education[16] as it could effectively gauge high level cognitive skills besides factual knowledge, such as understanding and problem-solving skills based on the respective learning outcome.[17]

The assessment tool should be valid and reliable to gauge the knowledge of the learner according to the level of cognition[9, 18] and its quality could be ascertained through collection, recapitulation and interpretation of student's abilities to respond[8]. Hence item drafting should be precise and clear as any fault in item writing such as unclear terminology could generate ambiguity and reduce reliability of the data[19 -22] and true evaluation of learner competency[13]

Item response theory (IRT) also known as a latent trait theory is a model based psychometric technique with attribute rank estimates of examinee responses based on abilities and characteristics or parameters of the item administered.

1. Lead "Department of Medical Education, SMBB Medical College, Lyari Karachi, Pakistan"

2. Faculty of Medicine, Northern Border University, Saudi Arabia & Dow Institute of health Professions Education, Dow University of Health Sciences, Karachi. Pakistan

*=corresponding author :
Email: dr.zeeshanarehman@gmail.com.

IRT emphasized on assessing each item individually rather than test score
with a concept that learner performance is a function of both learner ability and item psychometric characteristics[23]. The IRT model is appropriate to determine the reliability of test items with item characteristic curve (ICC) as the main attribute of the item response theory, which also impacts other constructs of the theory as well. S silhouette of the ICCs curve indicate association of ability scale with probability of correct answer {P(θ)} by positioning learner score on scale of ability (theta) ranges from negative to positive perpetuity, zero mean and element unit verses probability of correct response from each learner ranging from 0 to 1 for lowest to highest ability student respectively[23].

**Objective:** Therefore, objective of the current study is to provide insight for the educators to identify targets to be achieved for attaining necessary competency by students. The study evaluates learners' competence at three cognitive levels in terms of ability (θ) for Anatomy discipline using IRT approach to specifically approximate the true student ability in these cognitive disciplines with null hypothesis; Ho: Learners are not equally competent in Anatomy at different cognitive levels.

## Methodology:

The cross-sectional study was conducted in Public sector college of Pakistan for first professional MBBS students (100) with ethical approval. Purposive sampling was done and data was collected through the end of cardiovascular module formative assessment with Anatomy SBAs to gauge competence at three initial cognitive levels. The responses of students were employed to evaluate 20 SBAs (Anatomy C1=7, C2=6 and C3=7) with 1 and 0 score for correct and incorrect respectively without any negative marking) through IRT (2-PL) two parameter (dichotomous) logistic model approach with major parameters of IRT[23] i.e. discrimination and difficulty index (segregate quality items) and ICC (ability illustration). Item analysis was done to identify acceptable or desirable items to determine learner competence.

## Results:

Anatomy Cognitive Level.

The analysis of C1 level Anatomy items (Table 1) found 57 %, 14% very low, 14% low and 14% very highly discriminated items with p-values significant only for 3 items i.e. Ac12, Ac14 and Ac17. The difficulty index for 71.4% items is moderate, 14% easy and 14% is very difficult with p-value significant only for 3 items i.e.Ac12, Ac14 and Ac17. The reliability for the test is 0.45. The C2 level analysis (Table 1) shows that 33.33% items are moderate, 33.33% negative, 16.66% low and 16.66% are very highly discriminated. The difficulty index appears moderate for 66.66% and very difficult for 33.33% items with p-value significant only for 1 item (Ac22) and the overall test reliability 0.42. The analysis of C3 level items (Table 2) shows that discrimination of 57.1% items are moderate, 28.6 % low and

*Discrimination: negative (<0), very low (0.1-0.34), low (0.35-0.64),*

| Anatomy C1 level | Coefficient | Interpret | Std. errs. | Z | P>z | Item Alpha | Test Alpha | Action |
|---|---|---|---|---|---|---|---|---|
| Ac11<br>Discrim<br>Diff | .2418108<br>-2.164961 | very low<br>Easy | .3144031<br>.873195 | 0.77<br>-0.75 | 0.442<br>0.451 | 0.4494 | 0.4509 | Discard |
| Ac12<br>Discrim<br>Diff | 1.137174<br>-1.133064 | moderate<br>moderate | .5053291<br>.4156461 | 2.25<br>-2.73 | 0.024<br>0.006 | 0.4018 | | Acceptable for test pool |
| Ac13<br>Discrim<br>Diff | 2.18097<br>.0837288 | very high<br>moderate | 1.471165<br>.1620419 | 1.48<br>0.52 | 0.138<br>0.605 | 0.3546 | | Highly desirable for test pool |
| Ac14<br>Discrim<br>Diff | .706579<br>1.232824 | moderate<br>moderate | .3900126<br>.6527427 | 1.81<br>1.89 | 0.070<br>0.059 | 0.4292 | | Acceptable for test pool |
| Ac15<br>Discrim<br>Diff | .6849395<br>2.5086 | Moderate<br>very difficult | .4447295<br>1.434389 | 1.54<br>1.75 | 0.124<br>0.080 | 0.4414 | | Revise |
| Ac16<br>Discrim<br>Diff | .3742176<br>.2794272 | low<br>moderate | .3325106<br>.6024407 | 1.13<br>0.46 | 0.260<br>0.643 | 0.4251 | | Revise//Discard |
| Ac17<br>Discrim<br>Diff | .8450766<br>-.883065 | moderate<br>moderate | .4316136<br>.4498231 | 1.96<br>-1.96 | 0.050<br>0.050 | 0.3813 | | Acceptable for test pool |

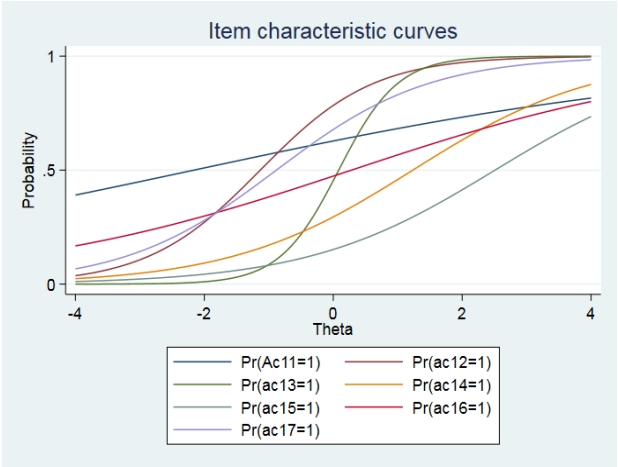
| Anatomy C2 level | Coefficient | Interpret | Std. err. | Z | P>z | Alpha | Reliability | Action |
|---|---|---|---|---|---|---|---|---|
| Ac21 | | | | | | | | |
| Discrim | 1.061652 | moderate | .4395064 | 2.42 | 0.016 | 0.3478 | | Acceptable for test pool |
| Diff | -.0696062 | moderate | .2345715 | -0.30 | 0.767 | | | |
| Ac22 | | | | | | | | |
| Discrim | 1.289375 | moderate | .6670452 | 1.93 | 0.053 | 0.2988 | | Acceptable for test pool |
| Diff | 1.265358 | moderate | .4663272 | 2.71 | 0.007 | | | |
| Ac23 | | | | | | | | |
| Discrim | -.3787838 | negative | .3461498 | -1.09 | 0.274 | 0.4280 | | Discard |
| Diff | 2.956972 | very diff | 2.616469 | 1.13 | 0.258 | | | |
| Ac24 | | | | | | | 0.4175 | |
| Discrim | .5408103 | low | .3620558 | 1.49 | 0.135 | 0.3856 | | Revise/discard |
| Diff | -.9336706 | moderate | .6846781 | -1.36 | 0.173 | | | |
| Ac25 | | | | | | | | |
| Discrim | 2.446097 | very high | 1.746145 | 1.40 | 0.161 | 0.2727 | | Highly desirable for test pool |
| Diff | -.0763543 | moderate | .1567179 | -0.49 | 0.626 | | | |
| Ac26 | | | | | | | | |
| Discrim | -.3380045 | negative | .4025157 | -0.84 | 0.401 | 0.4763 | | Discard |
| Diff | 5.456223 | very diff | 6.27671 | 0.87 | 0.385 | | | |
| Anatomy C3 level | Coefficient | Interpret | Std. err. | Z | P>z | Alpha | Reliability | Action |
| Ac31 | | | | | | | | |
| Discrim | 2.295253 | very high | 1.780258 | 3.02 | 0.003 | 0.3509 | | Highly desirable for test pool |
| Diff | .6875573 | moderate | .2462355 | 2.81 | 0.005 | | | |
| Ac32 | | | | | | | | |
| Discrim | .7247598 | moderate | .3876478 | 2.31 | 0.021 | 0.4322 | | Acceptable for test pool |
| Diff | 1.353859 | moderate | .678972 | 2.22 | 0.027 | | | |
| Ac33 | | | | | | | | |
| Discrim | .7752361 | moderate | .3803037 | 1.34 | 0.180 | 0.4201 | | Acceptable for test pool |
| Diff | .2680531 | moderate | .3121128 | 0.76 | 0.449 | | | |
| Ac34 | | | | | | | | |
| Discrim | .8195885 | moderate | .4351836 | 3.18 | 0.001 | 0.3910 | | Acceptable for test pool |
| Diff | .722926 | moderate | .4135849 | 2.12 | 0.034 | | 0.4527 | |
| ac35 | | | | | | | | |
| Discrim | .3972787 | low | .3123016 | 0.04 | 0.969 | 0.4506 | | Revise/discard |
| Diff | -. 9053695 | moderate | .8506057 | -0.04 | 0.969 | | | |
| Ac36 | | | | | | | | |
| Discrim | .4196424 | low | .3315924 | 0.87 | 0.384 | 0.4365 | | Revise/discard |
| Diff | -1.068684 | moderate | .9324719 | -0.82 | 0.414 | | | |
| Ac37 | | | | | | | | |
| Discrim | .660952 | moderate | .3766508 | 1.96 | 0.050 | 0.4148 | | Acceptable for test pool |
| Diff | -.2344806 | moderate | .3536343 | -0.67 | 0.500 | | | |

*moderate (0.65-1.34) high (1.35-1.69) very high (>1.70) perfect (+infinity)* [33]
*Difficulty: very easy (<-2.00), moderate (-2.00 to 2.00), very difficulty (>2.00)* [46].
*KR 20 (Alpha) value > or equal to 0.7.is considered significant.*
*Z value >1.96 and P value > 0.5*
***AC1: Cognitive level C1, 1= 1st item.***

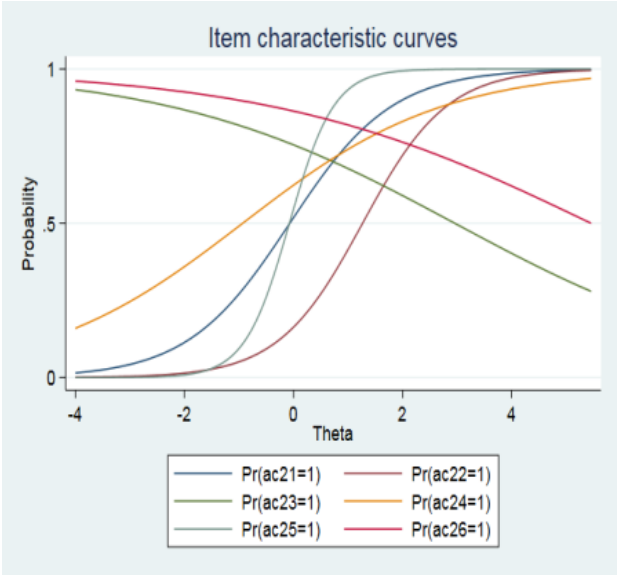
14.3% is very high while difficulty index for all the items is moderate with P value significant only for 4 items (AC31, 32, 34 & 37). The reliability of the test is 0.45. The analyses of responses for C1 Anatomy level impart a pattern of the learner competence for individual items with ICCs (Graph 1) with ability (θ) score at x-axis and probability of correct items at y -axis. Curve Ac16 illustrate very low discrimination and moderate difficulty with small slope and slow change in prospect of the correct response from lower to higher ability level. The curve Ac11 also shows low discrimination as there is only a slight upward projection of the slope as high ability level approaches. The item is also easy as the prospect of correct answer is 0.4 for low and reaches to approximately 0.7 for high-scorer examinees. The curve Ac14 and Ac15 are moderately discriminated with a slight steep slope as the probability of the correct response is approximately stumpy till θ =-1.0 and rise as high ability levels approaches. The curve Ac15 appears more difficult as compared to Ac14 with lower slope of correct responses for low ability which gradually increases for high ability examinees. The curve Ac13 appears highly discriminated and moderately difficult with steep slope in the middle as probability of correct response changes very rapidly with rise in ability. The curves Ac12 and 17 shows much less steep slope as compared to Ac13 and are moderately difficult and discriminated. The probability of correct response is 0.7, 0.5 0.3 and 0.65 at ability θ=0 (average learner) and 1 at ability θ=1.5 for Ac12 and Ac13 and θ= 4 for Ac 14 and 17. ICCs C2 level curve (Graph 2) shows that curve Ac24 is less discriminated and moderately difficult with less steep slope which changes slowly over the ability scale. The curve Ac21 and Ac22 both are moderately discriminated and difficult with a quite steep slope over the increasing range

of ability after score -3.0 **and -1.0 respectively, hence Ac22 displays more discrimination as compared to Ac21.**

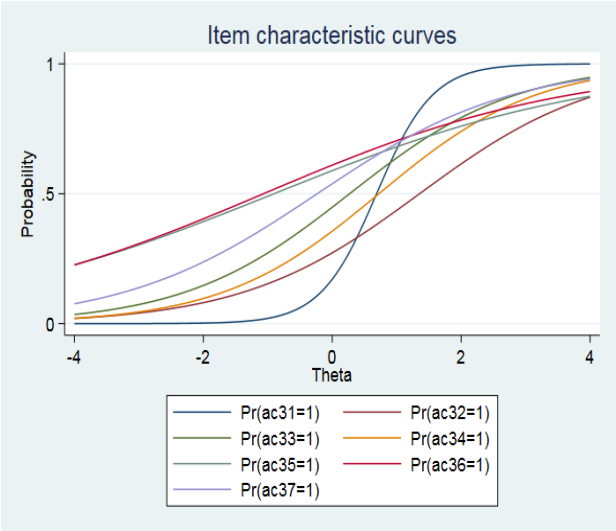
**Graph 1**(Anatomy at C1 level)




**Graph 2**(Anatomy at C2 level)




**Graph 3**(Anatomy at C3 level)




The curve Ac25 exhibits moderate difficulty level and high discrimination as slope shows upstroke at θ =-1.0 onwards with probability of correct response increase as higher ability level achieved. The curve Ac23, Ac26 shows downward slope as probability of correct answer is higher for low ability which is decreasing for high ability examinees; thus the item is either keyed, copied or incorrect having negative discrimination and is very difficult. The probability of correct response is 0.6, 0.25 and 0.7 at θ=0 and 1 at θ=4.3, 4.3 and 1.8 for Ac21, Ac22 and Ac25 respectively. The C3 level ICC curve (Graph 3) shows a small slope for Ac35 and Ac36 thus it has moderate difficulty and low discrimination. The curves Ac32, Ac33, Ac34 and Ac37 show only a slight rise in slope as ability range increases that represent moderate difficulty and discrimination. The curve Ac31 shows a quite steep slope in the middle showing prominent change in correct response with increase in ability of examinee. Thus, the item has moderate difficulty and a high level of discrimination. The probability of correct response is 0.1, 0.3, 0.45, 0.4, 0.6 at θ=0 while 1 at θ= 2.5, 4, 4, 4, 4 for Ac31, Ac32, Ac33, Ac34 and Ac37 respectively.

### Hypothesis Decision

| Hyp | Decision |
|---|---|
| Ho | Retained |

## Discussion:

The current study depicts that anatomy items at C1 level (Knowledge) are mostly having moderate difficulty and discrimination hence regarded as acceptable (Ac12, 14 and 17), whereas only one item is desirable i.e. very high discrimination and moderate difficulty (Ac13), hence could easily differentiate between students of low and high ability level. The very difficult item with moderate discrimination and easy with low discrimination need to be revised as too difficult or easy item could manipulate the result and should be written properly as per requirement for assessment of factual knowledge as differentiation is less between learners at different ability scores. The items that illustrate very low discrimination and easy difficulty should be discarded from the item pool as they fail to differentiate between learners at different ability levels. C2 level anatomy items with very high discrimination and moderate difficulty (Ac25) are considered as desirable while items with moderate DI & DIF are acceptable (Ac 21, Ac22) and could be retained in the test pool. The items with low discrimination and moderate difficulty should be considered for redrafting as these items were not able to effectively assess the ability of the examinees therefore should either be reassessed, edited or corrected to improve quality of item for better assessment. The items with negative discrimination and very high difficulty index depict that low achiever examinees answer more correctly than the high achievers which could be due to unclear questions or wrongly marked answer keys. Hence, such items should be discarded from the Q bank. Lastly, most of the anatomy items at C3 level are acceptable (Ac 32, Ac33, Ac 34 and Ac 37) and one in desirable range (Ac 31) while 2 items show low discrimination and moderate difficulty. The overall test reliability appears satisfactory as number of items is less in each test.

This study specifies that high discrimination and moderate difficulty items are authentic to evaluate learner competence while negative and low discrimination items with low or high difficulty are considered poor level items and should be discarded. The high incidence of difficult or easy

items and moderate to high frequency of poorly discriminating and negative items in present study stipulate continuing corrective measures to improve the quality of SBAs to store in question banks by identifying weak areas of competency for learners through item analysis and to effectively gauge competency of learners.

## Conclusion:

The study employed an item response approach to identify and differentiate between competence (ability) of the learner for correct responses using difficulty and discrimination indices in corroboration with item characteristic curve. The study found that C1 and C3 items are mostly acceptable while 50% of C2 items are of poor quality and required to be reassessed/ discarded. Hence, desirable and acceptable items should be regarded as a benchmark for a high quality item to enhance outcome for gauging competence of the examinee. The current study also signifies that either average learner is weak in understanding concept of Ac14 (C1 level), Ac 22 (C2 level) and Ac31, Ac32 (C3 level) item or was not reflected significantly by educator, hence need attention. The study also suggests mandatory provision of just-in-time evaluation and feedback to the learners for improving weak areas and to educators for writing a quality SBAs through training and experience. The learning outcomes for the SBAs should also be explained briefly to the educators to increase teaching and item writing skills of the educator.

## Contribution:

The current study provides a landmark for improving the quality of items and prepares a viable question bank for subsequent use as it is predicting actual performance and achievement of both educators and learners gauged through real performance scenarios[25]. The study will encourage the learners to learn more effectively with focus on weak areas and provide feedback to the teacher for improvement in item drafting and teaching methods.

## Future Recommendation:

The current study conducted for formative assessment only hence further study should be pursued with summative assessment across disciplines with more items and learners to improve competence estimation.

## Limitations:

The current study conducted on one-time formative assessment and one discipline due to time constraint, hence unable to gauge competence in all disciplines.

## Conflict of interest:

There is no conflict of interest for the study.

## References:


1 Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. Med Teach. 2010;32(8):676-82. [PMID: 20662580.] [DOI: https://doi.org/10.3109/0142159x.2010.500704.]

2. Hattie J., Gan M., Brooks C. (2017). Instruction based on feedback. In Mayer R. E., Alexander P. A. (Eds.), Handbook of research on learning and instruction. 2017 (2nd ed., pp. 263-285).

3. Kotur PF. Introduction of evidence-based medicine in undergraduate medical curriculum for development of professional competencies in medical students. Curr Opin Anaesthesiol. 2012 Dec;25(6):719-23. [DOI: 10.1097/ACO.0b013e32835a1112] [PMID: 23103843.]

4. Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: from Flexner to competencies. Acad Med. 2002 May;77(5):361-7. [PMID: 12010689] [DOI: 10.1097/00001888-200205000-00003]

5. Bing-You RG, Trowbridge RL. Why medical educators may be failing at feedback. JAMA. 2013;04102:22-3. » https://doi.org/10.1001/jama.2009.1393

6. Bloom, BS, Engelhart, MD, & Committee of College and University Examiners. Taxonomy of educational objectives: The classification of educational goals. London: Longmans, 1956.

7. 7. Hirsh DA, Ogur B, Thibault GE, Cox M. "Continuity" as an organizing principle for clinical education reform. N Engl J Med. 2007 Feb 22;356(8):858-66. [PMID: 17314348] [DOI: 10.1056/NEJMsb061660]

8. 8. Mirbahai L, Adie J. Applying the utility index to review single best answer questions in medical education assessment. Archives of Epidemiology and Public Health. 2020 Mar 25;1:1-5. DOI: 10.15761/AEPH.1000113.

9. 9. Miller GE. The assessment of clinical skills/ competence/performance. Acad Med. 1990 Sep;65(9 Suppl):S63-7. PMID: 2400509. DOI: 10.1097/00001888-199009000-00045

10. Lurie SJ, Nofziger AC, Meldrum S, Mooney C, Epstein RM. Temporal and group-related trends in peer assessment amongst medical students. Medical Education. 2006 Sep;40(9):840-7. https://doi.org/10.1111/j.1365-2929.2006.02540.x

11. Patel MS, O'Hare K. Residency training in transition of youth with childhood-onset chronic disease. Pediatrics. 2010 Dec;126 Suppl 3:S190-3. PMID: 21123485. PMID: 21123485. DOI: 10.1542/peds.2010-1466P

12. Lumley JSP, Craven JL, editors. (2004). MCQs in Anatomy. 3rd ed. New York: Churchill Livingstone, 2-4.

13. Crossley J, Humphris G, Jolly B. Assessing health professionals. Med Educ. 2002 Sep;36(9):800-4. PMID: 12354241. DOI: 10.1046/j.1365-2923.2002.01294.x

14. Begum T. A Guideline on Developing Effective Multiple Choice Questions and Construction of Single Best Answer format. J. Bangladesh Coll. Phys. 2012 Nov. 3 ;30 (3):159-66. DOI: http://dx.doi.org/10.3329/jbcps.v30i3.12466

15. Griffith CH 3rd, Wilson JF, Haist SA, Ramsbottom-Lucier M. Do students who work with better housestaff in their medicine clerkships learn more? Acad Med. 1998 Oct;73(10 Suppl):S57-9. PMID: 9795652. DOI: 10.1097/00001888-199810000-00045

16. Belay LM, Sendekie TY, Eyowas FA. Quality of multiple-choice questions in medical internship qualification examination determined by item response theory at Debre Tabor University, Ethiopia. BMC Med Educ. 2022 Aug 22;22(1):635. PMCID: PMC9394015. DOI: 10.1186/s12909-022-03687-y

17. Zafar U, Ali Z, Khalid A, Zaki S, Naeem FJ, Chaudhry

MA. Difficulty Index, Discrimination Index, Sensitivity and Specificity of Single Best Answer Questions to Assess Medical Students' Performance in Send Up Examination at Lahore Medical & Dental College. Pak J Phsyiol [Internet]. 2022 Sep. 30 [cited 2024 Apr. 17];18 (3):44-7. Available from: https://www.pjp.pps.org.pk/index.php/PJP/article/view/1472.

18. Badyal DK, Singh T. Learning Theories: The Basics to Learn in Medical Education. Int J Appl Basic Med Res. 2017 Dec;7(Suppl1):S1-S3. PMCID: PMC5769162. DOI. 10.4103/ijabmr.IJABMR_385_17

19. Holsgrove G, Elzubeir M. Imprecise terms in UK medical multiple-choice questions: what examiners think they mean. Med Educ. 1998 Jul;32(4):343-50. PMID: 9743793 DOI: 10.1046/j.1365-2923.1998.00203.x

20. Oliver CW, (ed.). Developing and maintaining an assessment system: A PMETB guide to good practice. 2007. 50 p. Available from: https://www.gmc-uk.org/-/media/documents/designing-and-maintaining-postgraduate-assessment-programmes-2109_pdf-70434370.pdf.

21. Hift RJ. Should essays and other "open-ended"-type questions retain a place in written summative assessment in clinical medicine? BMC Med Educ. 2014 Nov 28;14:249. PMCID: PMC4275935 DOI: 10.1186/s12909-014-0249-2

22. Schuwirth LW, van der Vleuten CP. ABC of learning and teaching in medicine: Written assessment. BMJ. 2003 Mar 22;326(7390):643-5. PMCID: PMC1125542. PMCID: PMC1125542 DOI: 10.1136/bmj.326.7390.643

23. Montgomery, Julie L. and Wendy Baker. "Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. Journal of Second Language Writing. June 2007 16(2):82-99. DOI:10.1016/j.jslw.2007.04.002.

24. Kusumawati M, Hadi S. An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school. Res Eval Educ.2018;4(1):70-8.

25. Abdul Rahim AF, Simok AA, Abdull Wahab SF. A guide for writing single best answer questions to assess higher-order thinking skills based on learning outcomes. Education in Medicine Journal. 2022;14(2):111-124. DOI. https://doi.org/10.21315/eimj2022.14.2.9.

.