International Journal of Research and Review
[Indexed, Peer-Reviewed Journal]
Volume 12; Issue: 3; March 2025
Website: www.ijrrjournal.com
E-ISSN: 2349-9788; P-ISSN: 2454-2237

Research Paper

Scientific Authenticity of Multiple-Choice Questions (MCQ) for Knowledge Test of Kinesiology and Biomechanics for Undergraduate Students of Physical Education and Sports Sciences

Syed Murtaza Hussain Andrabi[1], Dhananjoy Shaw[2], Rabiya Husain[1]

[1]PhD Scholar, Department of Physical Education and Sports Sciences, University of Delhi, Delhi, India
[2]Professor, Indira Gandhi Institute of Physical Education and Sports Sciences, University of Delhi, Delhi, India

Corresponding Author: Rabiya Husain

DOI: https://doi.org/10.52403/ijrr.20250343

ABSTRACT

The study evaluates the scientific authenticity of multiple-choice questions (MCQ) in kinesiology and biomechanics by analysing their difficulty index, discrimination index, and internal consistency to determine their effectiveness in student assessment. Eighty-one physical education students enrolled in a kinesiology and biomechanics subject in graduation programme at the University of Delhi. Over 20 weeks, students attended lectures followed by multiple-choice question (MCQ) tests, with responses recorded for analysis. Students were divided into high-scoring (n=21) and low-scoring (n=21) groups, and the discrimination index (D) was calculated using the formula $D = p_h - p_l$. The Split-Half and Odd-Even reliability methods were used to assess internal consistency. The item analysis revealed that 60% of the MCQs were too easy ($p > 0.70$), 37% were acceptable ($p = 0.30–0.70$), and 3% were too difficult ($p < 0.30$). The discrimination index showed that 49% of the items were excellent ($D > 0.35$), 23% were good (0.25–0.34), 14% were acceptable (0.15–0.24), and 14% required revision ($D < 0.15$). Cronbach's Alpha (0.809) confirmed strong internal consistency. Split-half reliability analysis showed a moderate correlation (0.605) between Part A and Part B, but when correlated with the total test score, Part A and Part B had high correlations of 0.892 and 0.900, respectively. Odd-even reliability analysis indicated a 0.780 correlation between odd- and even-numbered questions, increasing to 0.944 and 0.942 when correlated with the total score, reinforcing the test's reliability. The results indicate that MCQ for kinesiology and biomechanics are a valid and reliable tool for assessing knowledge in kinesiology and biomechanics, demonstrating strong discrimination and internal consistency. The high split-half and odd-even reliability correlations with the total test score validate the assessment's accuracy. While a more balanced difficulty distribution is needed, the findings highlight the importance of systematic MCQ validation for effective student evaluation.

Keywords: Biomechanics, item analysis, kinesiology, MCQ, reliability.

INTRODUCTION

Education is a continuous process that integrates teaching, learning, and assessment.[1,2] One of the most critical aspects of this process is evaluating student outcomes after instructional delivery.[3] Assessment provides insights into students' comprehension, identifies learning gaps, and informs instructional modifications.[4-6] Traditionally, academicians have relied on subjective and time-consuming evaluation methods, but the introduction of Multiple-Choice Questions (MCQ) revolutionized assessment.[7] Multiple-Choice Questions (MCQ) are widely used due to their efficiency, objectivity, and ability to test a broad range of knowledge within a short time.[8,9] MCQs allow for standardized evaluation, enabling fair comparison across students while minimizing examiner bias.[10] They can assess not only factual recall but also higher-order cognitive skills, such as application, analysis, and evaluation.[11]

Despite their advantages, MCQs also have certain limitations. One major concern is the possibility of students guessing the correct answer, particularly in true/false and single-response formats, which may compromise assessment accuracy.[12] Poorly constructed MCQs can fail to effectively differentiate between high- and low-performing students, reducing their discriminatory power.[13] Additionally, they may encourage rote memorization rather than deep learning, potentially limiting students' ability to apply knowledge in practical contexts.[10] When systematically constructed and thoroughly analyzed, MCQ are considered among the most reliable and valid assessment tools, provided they undergo rigorous validation through item analysis.[14] The effectiveness of MCQ largely relies on their scientific validity, which necessitates comprehensive evaluation prior to administration.[3] Item analysis is a psychometric process used to assess the quality of MCQ based on difficulty index, discrimination index, and distractor effectiveness.[15,16]

The difficulty index (p-value) determines how challenging an item is for students, with an ideal range between 30% and 70%.[19] Questions that are too easy or too difficult fail to effectively differentiate between students with varying levels of competency.[3] The discrimination index (d-value) measures how well an item distinguishes between high and low performers, where values above 0.30 are considered effective in differentiating student abilities.[17] Additionally, distractor analysis ensures that incorrect answer choices function effectively by attracting students with misconceptions rather than being ignored.[13] Poorly designed distractors lead to guessing tendencies and reduce the validity of assessment outcomes.[7]

In the field of kinesiology and biomechanics, MCQ are integral to evaluating students' theoretical understanding of movement, force application, and injury prevention.[10] The demand for competency-based education in sports sciences has led to a greater emphasis on constructing reliable and valid MCQ.[7] However, research suggests that many biomechanics MCQ are poorly structured, ambiguous, and misaligned with learning objectives.[18] Given the critical role of biomechanics in physical education, it is essential to ensure that MCQ accurately assess conceptual and applied knowledge.[3]

The present study aims to evaluate the scientific authenticity of multiple-choice questions (MCQ) in kinesiology and biomechanics through item analysis to determine their effectiveness as an assessment tool.

The specific objectives of this study are to:

1. Analyse the difficulty index to assess the relative ease or challenge of each test item.
2. Evaluate the discrimination index to determine the effectiveness of each MCQ in distinguishing high- and low-performing students.
3. Assess internal consistency using split-half and odd-even reliability methods to measure the overall reliability of the test in relation to total score (student performance).

By conducting a comprehensive item analysis, this study seeks to enhance MCQ assessment quality, ensuring fair, objective, and valid evaluations of students' competencies in kinesiology and biomechanics. The findings will guide educators in refining test construction strategies, ultimately improving the effectiveness of competency-based learning assessments.[9]

MATERIALS & METHODS

The study involved a total of eighty-one (N=81) physical education students from Indira Gandhi Institute of Physical Education and Sports Sciences (University of Delhi), enrolled in a kinesiology and biomechanics course as part of their bachelor's degree program. Over a period of 20 weeks, students attended regular weekly lectures delivered by the professor, followed by a knowledge test in the form of multiple-choice questions (MCQ). The students' responses were recorded and organized into a spreadsheet format for further analysis. To ensure the relevance and appropriateness of the test items, an internal validation process was conducted through item analysis, where variables such as item difficulty and discrimination index were computed for each test item. The data collected from student responses were exported to Excel, where statistical analysis was performed. In order to calculate the level of discrimination (**D**), the subjects were put in a descending order according to the total score earned on the whole test. For the purpose of the study the group was divided, breaking them into two groups i.e., high scoring group (n=21) and low scoring group (n=21). The following formulae was used to compute the level of discrimination. In which P_h represents number of correct responses in the high group divided by 21 and P_l represents the number of correct responses in the low group divided by 21.

$$D = p_h - p_l$$

This analytical approach provided insights into the effectiveness of each test item in differentiating between high- and low-performing students, ensuring the reliability and validity of the assessment.

RESULT

Table 1: Descriptive Statistics of Score and Numbers in High and Low Score Group of Kinesiology and Biomechanics Test

Items	Test takers (n)	Score		High Score Group (n)	Low Score Group (n)
		Mean ± S. D.	Median		
35	81	25.38 ± 5.26	26	21	21

Note: **Rounded to two digits after the decimal; n= number**

The Table 1 presents the descriptive statistics of the test scores, including the mean (25.38 ± 5.26) and median (26) scores of the 81 participants. The high-scoring group (n=21) and low-scoring group (n=21) were determined based on total test scores, ensuring an effective analysis of item discrimination.

Table 2: Categorization of the Items in regard to their Ratings of Difficulty Index (p)

S. No.	Number of Items			Total Number of Items
	Too Difficult	Acceptable	Too Easy	
	(Less than 0.30)	(0.30 to 0.70)	(More than 0.70)	
Total count	1	13	21	35
Item No	14	1,7,8,9,12,16,17,18, 21,23,27,30,35	2,3,4,5,6,10,1113,15,19,20, 22,24,25,26,28,29,31,32,33,34	

Table 2 displays the difficulty index (p) of the administer items. One item was rated as "too difficult", 13 items were rated as "acceptable" and 21 "items" were regarded as "too easy" as per the rating scale proposed by Natekar & De Souza, 2016.

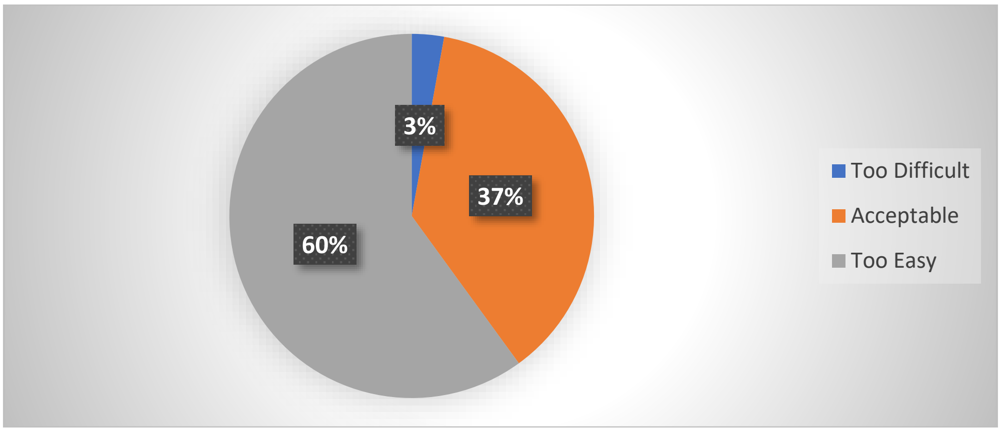

Figure 1: Percentage Distribution of Three Level of Difficulties

The Figure 1 illustrates the proportion of test items classified under different difficulty levels. Based on the item difficulty index, questions are categorized as too easy (>0.70), acceptable (0.30–0.70), and too difficult (<0.30). The figure provides a visual representation of how well the MCQ were distributed across these categories.

Table 3: Categorization of the Items in regard to their Ratings of Index of Discrimination (D)

S. No.	**Number of Items**				**Total Number of Items**
	Discard	**Acceptable**	**Good**	**Excellent**	
	(Less than 0.15)	**(0.15 to 0.24)**	**(0.25 to 0.34)**	**(Greater than 0.34)**	
Total count	5	5	8	17	35
Item No	12, 13, 14, 24, 25	1, 6, 21, 22, 30	2, 5, 15, 19, 20, 31, 32, 34	3, 4, 7, 8, 9, 10, 11, 16, 17, 18, 23, 26, 27, 28, 29, 31, 32, 33, 34	

Table 3 displays the rating on the index of discrimination with four categories of rating (Discard, Acceptable, Good, and Excellent) as proposed by Natekar & De Souza, 2016. Out of the total 35 items, 5 items were discarded, 5 items were rated as "acceptable", 8 items were rated as "good" and 17 items were rated into "excellent" category.

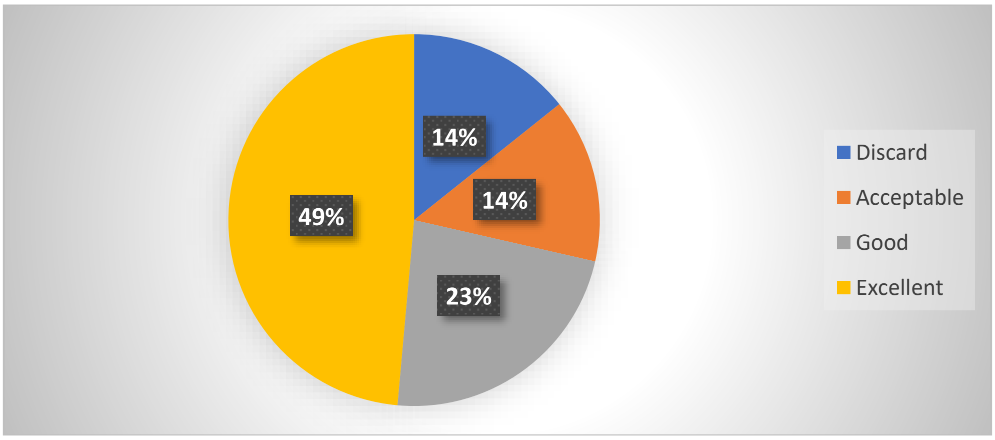

Figure 2: Percentage Distribution of Ratings of Index of Discrimination

The Figure 2 represents the proportion of MCQ categorized under different discrimination levels, showing the effectiveness of each item in distinguishing between high- and low-performing students. The analysis follows established rating criteria, with items classified as excellent (>0.35), good (0.25–0.34), acceptable (0.15–0.24), or poor (<0.15).

Table 4: Common Items

S. No.	Too Easy (More than 0.70)	Index of Discrimination (Less than 0.15)	Common
Total count	21	5	3
Q. No.	2,3,4,5,6,10,11,13,15,19,20, 22, 24, 25, 26, 28, 29, 31, 32, 33, 34	12,13,14,24,25	13,24,25

Note: **Prashant E. Natekar & Fatima De Souza (2016) rating criteria was used to identify "too difficult", "acceptable", and "too easy"**

The Table 4 identifies MCQ that were either too easy ($p > 0.70$), had a low discrimination index ($D < 0.15$), or both. The question numbers corresponding to each category are listed, highlighting specific items that may require revision for improved test quality.

Table 5: Summary of Internal Consistency

Criteria	No of items deleted	Alpha	Rating
Deleting zero items	-	0.809	High
Deleting common* items	3	0.804	High
Deleting items with D less than 0.15	5	0.819	High

Note: ***d less than 0.15 and p value more than 0.70**

The Table 5 presents the Cronbach's Alpha values computed after deleting different sets of items. The initial reliability was 0.809, which remained high (>0.8) even after removing common items or those with poor discrimination indices. This indicates that the test maintained a strong internal consistency across different conditions.

Table 6: Descriptive Statistics for Split-Half and Odd-Even Method

Split-Half		Odd-Even	
	Mean ± S.D.		**Mean ± S.D.**
Part A	12.23 ± 2.89	**Odd**	13.23 ± 2.82
Part B	13.15 ± 2.99	**Even**	12.15 ± 2.76
Total	25.38 ± 5.26	**Total**	25.38 ± 5.26

Note: **N=81; Part A (Q1–Q18); Part B (Q19–Q35); Odd Part (Odd S.No. questions, N=18), Even Part (Even S.No. questions, N=17)**

The Table 6 compares the mean and standard deviation of scores from two methods of testing the reliability (split-half method and odd-even method). In split-half method, the Part A section consists of multiple-choice questions from serial numbers 1 to 18, while Part B includes questions from serial numbers 19 to 35. In the odd-even method, the odd part comprises all questions listed at odd serial numbers (1, 3, 5, etc.), whereas the even part includes questions at even serial numbers (2, 4, 6, etc.).

Table 7: Correlations of Split-Half Reliability

		Part A	Part B	Total
Part A	**Pearson Correlation**	1	.605**	.892**
	Sig. (2-tailed)		.000	.000
Part B	**Pearson Correlation**	.605**	1	.900**
	Sig. (2-tailed)	.000		.000

Note: **Correlation is significant at the 0.01 level (2-tailed).

The Table 7 displays Pearson correlation coefficients between the test's two halves, having acceptable correlations (r = 0.605) indicate that the two halves (Part A and Part B) of the test produce consistent results (reliable). There was excellent validity coefficient between total score and Part A (r= 0.892) as well as total score and Part B (r= 0.900).

Table 8: Correlations of Odd-Even Reliability

		Odd	Even	Total
Odd	**Pearson Correlation**	1	.780**	.944**
	Sig. (2-tailed)		.000	.000
Even	**Pearson Correlation**	.780**	1	.942**
	Sig. (2-tailed)	.000		.000

Note: **Correlation is significant at the 0.01 level (2-tailed).

Table 8 presents the correlation values for odd- and even-numbered items. The high correlation (r = 0.780) between odd and even groups confirm strong internal consistency. Additionally, the excellent validity coefficients between the odd group and total (0.944) and between the even group and total (0.942), both significant at the 0.01 level, further validate the test's effectiveness in student evaluation.

DISCUSSION

The objective of the present study was to evaluate the scientific authenticity of MCQ used in kinesiology and biomechanics by analysing their difficulty index, discrimination index, and internal consistency. The results demonstrated that 60% of the items were categorized as too easy (p > 0.70), 37% fell within the acceptable range (p = 0.30–0.70), and 3% were too difficult (p < 0.30). This skewed distribution suggests that the test may not have sufficiently challenged students at higher competency levels, similar to the observations made by Natekar & De Souza.[3]

A study found a more balanced difficulty distribution, reinforcing that a greater proportion of moderately difficult items enhances the effectiveness of MCQs. The predominance of easy questions in this study may lead to an overestimation of student competency, emphasizing the need for a more balanced range of difficulty levels.[16]

The discrimination index analysis revealed that 49% of the MCQ were excellent at distinguishing high- and low-performing students (D > 0.35), 23% were rated as good (0.25–0.34), 14% as acceptable (0.15–0.24), and 14% required revision (D < 0.15). Despite the high proportion of easy questions, the majority of the test items effectively differentiated student performance, indicating the scientific validity of the MCQ used in this study. These results align with Tarrant et al.[13], who emphasized that well-constructed MCQ maintain strong discriminatory power, even if difficulty levels vary. However, another study observed lower discrimination indices in biomechanics-related MCQ, suggesting that subject complexity and student familiarity with the topic might impact question effectiveness.[11] The presence of items with a low discrimination index in this study suggests that certain questions may require revision to improve their clarity, alignment with learning objectives, and ability to challenge students appropriately.

A significant strength of this study lies in its reliability analysis. Cronbach's Alpha (0.809) indicates high internal consistency, reinforcing that the MCQs measured the intended knowledge domain effectively. The split-half reliability analysis showed that when Part A and Part B were correlated directly, the coefficient was 0.605, suggesting moderate reliability. However, when each part was correlated with the total score, Part A had a much higher correlation of 0.892, and Part B had an even stronger correlation of 0.900, indicating that both sections contributed significantly to overall test reliability. According to Kirkendall's

Reliability Rating Scale, these values confirm moderate reliability for the halves but excellent reliability when correlated with the total test score.[20] This aligns with findings of the study who noted that higher total-score correlations typically indicate well-structured assessments with balanced item difficulty.[7]

Similarly, odd-even reliability analysis revealed a correlation of 0.780 between the odd and even numbered questions, indicating strong internal consistency. However, when the odd-numbered questions were correlated with the total score, the reliability coefficient increased significantly to 0.944, while the even-numbered questions correlated at 0.942. These values fall within the "excellent reliability" category according to Kirkendall's scale. Such findings confirm that while split-half and odd-even correlations within sections showed moderate to strong reliability, their relationships with the total score demonstrated exceptional test consistency. This aligns with previous research that suggests internal consistency improves significantly when all test components are considered together rather than in isolated halves.[7,14]

Despite these strengths, certain limitations were identified. The high proportion of easy questions may lead to inflated student scores, reducing the ability of the test to accurately assess high-level competency.[12] Additionally, while most items had strong discriminatory power, a small subset (14%) required revision, suggesting the need for improved question design and more effective distractor choices. Future assessments should consider integrating more application-based and conceptually challenging MCQs to enhance test effectiveness.

CONCLUSIONS

The present study confirms that MCQs are a scientifically valid and reliable tool for assessing kinesiology and biomechanics knowledge, demonstrating strong effective differentiation of student performance, strong reliability, and consistent assessment outcomes. The split-half and odd-even reliability analyses revealed excellent correlations with the total test score, reinforcing the assessment's reliability. While the majority of items effectively differentiated student performance, the high proportion of easy questions suggests a need for better balance in difficulty levels to enhance assessment effectiveness. Future assessments should incorporate more application-based MCQs to improve discriminatory power.

Declaration by Authors

Acknowledgement: We sincerely appreciate all the participants (students) for taking the time to complete the MCQ.

Source of Funding: None

Conflict of Interest: No conflicts of interest declared.

APPENDIX

List of Questions in Kinesiology and Biomechanics MCQ.

S. No.	Questions
1	The branch of mechanics that deals with the study of causes of angular motion
2	The adduction/abduction of the arm happens in which plane
3	Walking happens predominantly in which plane
4	Which of the following is not a property of skeletal muscles
5	Factor/s that affect a projectile is/are
6	Force is
7	Mass and weight
8	Skiers wax the bottom of the skies in order to
9	Which of the following is not a phase of motor action
10	Work done can be
11	Muscles can be
12	The muscles that assist the prime movers and make movement more refined are called

13	The horizontal plane cuts the body into
14	The term Kinesiology means
15	Speed is
16	The degree of conformity between result and predicted result of motor action is
17	The principle of conservation of angular momentum applies to bodies that are
18	Which among the following is not a cyclic motor action
19	Newton's second law is also called
20	One Newton is an equivalent of
21	The most common type of lever in human body is
22	Which among the following is an example of law of reaction
23	In which of the following can friction prove to be a disadvantage
24	Unit of pressure is
25	Unit of work is
26	A body doesn't move unless a force is applied to it is
27	Force is produced by
28	Newtons third law is also called
29	A swimmer is able to swim when he pushes the water backwards. Which law is directly linked to it
30	As a batsman hits the ball, the greater the force applied the greater will be the acceleration which law is linked to it
31	A car is moving and the break is applied, the passengers in the car tend to fall in forward direction this is due to
32	F=ma, here "a" stands for
33	The equation of potential energy is "mgh", here "h" stand for
34	The path of a projectile is called
35	The distance travelled by a projectile depends upon

REFERENCES

1. Gagné, R. M. (1985). The Conditions of learning and theory of instruction. Holt.
2. Yambi, T. D. A. C., & Yambi, C. (2018). Assessment and evaluation in education. University Federal do Rio de Janeiro, Brazil.
3. Natekar, P. E., & De Souza, F. (2016). Item analysis (post-validation) of MCQ in anatomy for the first MBBS course. Indian Journal of Anatomy, 5(1), 29–32. https://doi.org/10.21088/ija.2320.0022.5116.5
4. Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74. doi:10.1080/0969595980050102
5. Brinkley, A. J., Sherar, L. B., & Kinnafick, F. E. (2022). A sports-based intervention for pupils excluded from mainstream education: A systems approach to intervention acceptability and feasibility. Psychology of Sport and Exercise, 61, 102217. https://doi.org/10.1016/j.psychsport.2022.102217
6. Zulkifli, A. F., & Danis, A. (2022). Technology in physical education: Using movement analysis application to improve feedback on sports skills among undergraduate physical education students. Social Sciences & Humanities Open, 6(1), 100350. https://doi.org/10.1016/j.ssaho.2022.100350
7. Harper, R. (2003). Multiple-choice questions -a reprieve. Bioscience Education, 2(1), 1-6. https://doi.org/10.3108/beej.2003.02000007
8. Namdeo, S. K., & Sahoo, B. (2016). Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar, India. Int J Res Med Sci, 4(5), 1716-19. DOI: http://dx.doi.org/10.18203/2320-6012.ijrms20161256
9. Elgadal, A. H., & Mariod, A. A. (2021). Item analysis of multiple-choice questions (MCQ): Assessment tool for quality assurance measures. Sudan Journal of Medical Sciences, 16(3), 334–346. https://doi.org/10.18502/sjms.v16i3.9695
10. Anderson, J. (1979). For multiple choice questions. Medical Teacher, 1(1), 37-42. https://doi.org/10.3109/01421597909010580
11. El-Gohary, T. M., & Abdelkader, S. M. (2020). Computation is used to denote application level within the cognitive process dimension of the multiple-choice questions derived from biomechanics and kinesiology courses. International Journal of

Contemporary Teaching and Research, 11(3), 173–189.
12. Chandratilake, M., Davis, M., & Ponnamperuma, G. (2011). Assessment of medical knowledge: The pros and cons of using true/false multiple choice questions. The National Medical Journal of India, 24(4), 225-228.
13. Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. BMC Medical Education, 9(40).
14. Klůfa, J. (2016). Multiple choice question tests – Advantages and disadvantages. Recent Advances in Educational Technologies, 1(1), 39-46. ISBN: 978-1-61804-322-1
15. Ferketich, S. (1991). Focus on psychometrics. Aspects of item analysis. Research in nursing & health, 14(2), 165-168. https://doi.org/10.1002/nur.4770140211
16. Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple-choice questions: A quality assurance test for an assessment tool. Medical Journal of Armed Forces India, 77(S85–S89). https://doi.org/10.1016/j.mjafi.2020.11.007
17. Rehman, A., Aslam, A., & Hassan, S. H. (2018). Item analysis of multiple-choice questions. Pakistan Oral & Dental Journal, 38(2), 291–297.
18. Haqqi, M., Rifki, M., Arsil, A., & Sepriadi, S. (2025). Analysis of Multiple-Choice Items in Physical Education Subject: A Systematic Literature Review. Journal of Education, Teaching and Learning, 10(1), 47-53. doi: http://dx.doi.org/10.26737/jetl.v10i1.6701
19. Rao, C., Prasad, K. H. L., Sajitha, K., & Permi, H. S. (2016). Item analysis of multiple-choice questions: Assessing an assessment tool in medical students. International Journal of Educational and Psychological Researches, 1(1), 50–57.
20. Kaur, M., Rawat, S., & Shaw, D. (2023). Relationship among five-point, seven-point and nine-point scale of health assessment questionnaire when administered to female sportspersons: A validation. Global Journal for Research Analysis, 12(3), 227-230. https://doi.org/10.36106/gjra

How to cite this article: Syed Murtaza Hussain Andrabi, Dhananjoy Shaw, Rabiya Husain. Scientific authenticity of multiple-choice questions (MCQ) for knowledge test of kinesiology and biomechanics for undergraduate students of physical education and sports sciences. *International Journal of Research and Review*. 2025; 12(3): 352-360. DOI: *https://doi.org/10.52403/ijrr.20250343*

******