Teaching and Learning Research Journal Volume: 1 Nomor 3. 2025, Hal 80-86
EISSN:3090-1480

Scoring Assessment on Multiple-Choice Questions

Zulpan,[1] Sahrul,[2] Muhammad Rizky Rochmawan[3] Arianto,[4]Ahmad Husein[5], Ahmad Fauzi Pulungan[6], Siti Sarah,[7] Tri Idha Sari[8],
[1] Akbid Madina Husada , [2 3] Universitas Ivet Semarang, [1 4 5 6 7 8] STAIN Mandailing Natal

*Correspondensi: Zulpan
Email: zulpan200990@gmail.com.

Abstrak: *This article aims to explain the scoring method used in processing the results of multiple-choice tests. The focus of the discussion includes how to construct test items, how to assign scores, how to analyze the results, and how to make decisions based on the test outcomes. The article was developed by conducting an in-depth analysis and synthesis of theories from scientific articles and books related to scoring multiple-choice items. The result is a structured concept that outlines the process of item construction, scoring, and decision-making for multiple-choice tests. This article is expected to assist and facilitate teacher education students and classroom teachers in improving their teaching quality and becoming more professional in their field. The scoring method for test results is generally adjusted to the type of test used, whether it is an objective test, essay test, or other forms. For multiple-choice items, which fall under the category of objective tests, each correct answer is usually assigned a score of 1 (one), while each incorrect answer is assigned a score of 0 (zero).*

Keywords: ***Arrangement, Scoring andIinterpretation of Multiple Choice.***

Introduction

The effectiveness of teaching is often measured by the extent to which students achieve mastery of the intended learning objectives. To determine this, teachers are required to conduct systematic assessments of students following the learning process. In many educational contexts, multiple-choice tests remain one of the most widely used assessment formats. This preference is largely due to their clear answer keys, which ensure objectivity in scoring and produce consistent results regardless of the examiner. Generally, multiple-choice items consist of three essential components: the stimulus, the question stem, and the set of answer alternatives—comprising both distractors and the correct answer. Additionally, adherence to established item-writing guidelines plays a crucial role in determining the overall quality of the test and the cognitive level being measured (Kadir, 2015).

While teachers and students often expect test results to reflect positive learning outcomes, the reality can be quite the opposite. Such discrepancies are not always attributable to ineffective teaching methods or poor student study habits. Instead, they may stem from flaws in the evaluation tools themselves—particularly in the quality of the test items used (Widodo, 2010). The construction of multiple-choice questions, therefore, represents a potential source of assessment inaccuracy. As noted by Yusron et al. (2024), many teachers and pre-service teachers continue to overlook established principles of multiple-choice question design when developing assessment items.

This concern is reinforced by the findings of Akbar (2020), which reveal that students preparing multiple-choice questions frequently make errors across several indicators, including insufficient attention to the development of plausible distractors. Similar patterns were observed in other

participants. Questionnaire results further indicate that 28.1% of respondents strongly agreed and 50% agreed that designing effective stimuli and crafting well-structured question stems posed significant challenges. Moreover, lecturers were perceived as offering limited clarity regarding the processes and procedures of item construction. Explanations about question development, as well as guidance on determining item difficulty levels, were reported to be insufficiently conveyed (Purwasih et al., 2020).

In light of these findings, this article seeks to examine the construction, scoring, and decision-making processes involved in multiple-choice assessments. It presents a structured conceptual framework and practical implementation strategies for scoring such items. Ultimately, the article aims to contribute to the professional competencies of pre-service teachers and in-service educators by enhancing their ability to develop high-quality multiple-choice questions and conduct objective, reliable learning assessments.

Methods

This study employs a literature review approach, conducted by gathering a range of relevant and credible references in the form of scholarly articles, books, and other sources related to the construction and scoring of multiple-choice questions. The process began with an in-depth understanding of the theoretical foundations obtained from these sources, followed by analysis, evaluation, and synthesis of the theories. This process ultimately resulted in a structured conceptual framework outlining the procedures for developing and scoring multiple-choice items.

Discussions

Multiple-Choice Questions (MCQs)

Multiple-choice questions (MCQs) are among the most widely used test formats in educational evaluation, from primary to higher education levels. MCQs are a form of objective test consisting of a question or statement accompanied by a set of predetermined answer options. Ardania et al. (2022) define MCQs as tests containing questions and several alternative answers. Similarly, Surapranata (2006) describes MCQs as questions for which the answer must be selected from a range of possible choices provided by the test developer. Widoyoko (2017) further explains that each MCQ contains more than one possible answer, typically ranging between two to five options.

In general, an MCQ consists of two main components: the question stem and the answer options. According to Alwi (2010), the Assessment Center of the National Education Research and Development Agency sets specific rules for the number of answer options in MCQs: four options for primary and junior secondary school levels, and five options for senior secondary school level. MCQs can measure students' abilities across a range of cognitive levels, from basic recall to complex reasoning, with items typically offering between two and five alternatives (Daniel, 2019).

Answer options include one correct answer (the answer key) and several incorrect alternatives (distractors). There are two main types of MCQs: (1) *single correct answer*, where only one answer is correct and the others are incorrect; and (2) *best answer*, which measures learning outcomes that require understanding, application, or interpretation, where the best answer is the one agreed upon by experts (Wartoni, 2020). MCQs have several advantages: (1) they can be scored quickly and accurately, producing highly reliable results; (2) they can assess a wide range of cognitive levels, from basic knowledge to complex reasoning; and (3) they cover a broad range of material in a single test. Slameto (2001) highlights that MCQs are especially suitable for large-scale testing—such as

semester exams, school exams, and final examinations—where results need to be reported quickly. However, MCQs also have limitations: (1) their development requires substantial time, effort, and expertise; (2) they are challenging to design for assessing higher-order cognitive skills; and (3) there is a possibility that students may guess the correct answer. In addition, constructing effective and homogeneous distractors can be difficult, and test-takers may still arrive at correct answers through guessing.

Therefore, the development of MCQs must be carried out with care and in adherence to established principles of good test construction. According to Alwi (2010), the quality of MCQs is highly dependent on the skills of the test developer. Producing high-quality, objective MCQs requires professional competence, dedication, and creativity (Suryabrata, 2004). In MCQ item-writing guidelines, Arif (2014) outlines three key aspects to consider: (a) content, (b) construction, and (c) language. Hazraini (2017) similarly identifies these three aspects, which are further detailed into the following indicators:

Aspect	Indicators
Content	Items must align with the indicators specified in the test blueprint; distractors must function effectively; each item should have only one correct answer.
Construction	The question stem must be clearly and precisely stated; avoid double negatives; answer options must be homogeneous and logically consistent with the content.
Language	Use appropriate and standardized language; ensure language is communicative and easily understood by students; avoid the use of regional dialects.

Hazraini (2017)

Susilo (2014) further emphasizes that MCQ development should: (1) select content that assesses the ability to understand, apply, analyze, synthesize, or evaluate information—while recognizing that recall is an essential foundation for higher-order thinking; (2) routinely design questions that test critical thinking and problem-solving skills; and (3) provide an initial stimulus—such as an illustration, reading passage, case study, or table—before presenting the main question.

Scoring Methods

Scoring is a method or technique for assigning scores. According to the *Kamus Besar Bahasa Indonesia* (KBBI), a score is defined as the total points of achievement or the value obtained in a test, examination, or quiz. To score means to assign a score, while scoring refers to the process or act of assigning scores (Irwan et al., 2022). In essence, scoring is the process of converting responses on an instrument into numerical values representing the quantitative measurement of responses to each item. Scoring involves converting test answers into numbers or numerical values (Asrul et al., 2015). The numerical results are then processed into grades (Elis, 2015). According to Syahputra and Arsyam (2020), scoring methods are generally adapted to the type of test, whether objective tests, essay tests, or other formats. In multiple-choice questions (MCQs), which are a form of objective test, each correct answer is typically assigned a score of one (1), while incorrect answers receive a score of zero (0) (Ibrahim & Muslimah, 2021). Zainal (2016) notes that objective tests are often referred to as

dichotomous tests because they have only two possible outcomes—correct or incorrect—scored as 1 or 0, respectively. The scoring is objective because the answer key is fixed and unambiguous, ensuring that all graders will produce identical results.

1) Scoring by Counting Correct Answers
   This method involves counting the number of correct answers from the total items, assigning one point for each correct answer. The total score is calculated using the formula:

$$S = R$$

   where S is the total score, and R is the number of correct answers (Putri et al., 2022).
   Example: If there are 25 questions and a student answers 20 correctly, the total score is 20.
2) Penalty for Incorrect Answers
   This method deducts points for incorrect answers, using the formula:

$$S = R - \frac{W}{O-1}$$

   where *S* is the total score, *R* is the number of correct answers, *W* is the number of incorrect answers, and *O* is the number of answer options (Rosyidi, 2020).
   Example: With 25 questions and 4 answer choices, if a student answers 20 correctly and 5 incorrectly, the score is:
   20- (5/4-1) = 20 – 1,67 = **18,33**
3) Bonus for Unanswered Items
   This method rewards students for leaving questions unanswered, calculated as:

$$S = R + \frac{T}{O}$$

   where T is the number of unanswered questions (Bhakti, 2015).
   Example: With 25 questions and 4 answer choices, if a student answers 21 correctly and leaves 4 unanswered: 21 + (4/4) = 21 + 1= **22**

4) Weighted Item Scoring
   This method assigns different weights to items based on cognitive levels, calculated as:

$$S = R_i x B_i$$

   where R_i is the number of correct answers in a given category, and B_i is the assigned weight (Rofieq, 2008).
   Example: With 20 questions, weighted by cognitive level as follows:

Cogntive Level	Weight	Total Items
Remembering	1	5
Understanding	2	5
Applying	3	5
Analyzing	4	5

If a student answers all 5 remembering items, all 5 understanding items, 3 applying items, and 2 analyzing items correctly, the total score is: 5x1 + 5x2 + 3x3 +2x4 = **32**

Scoring serves as a tool for teachers in determining grading policies, especially in decision-making regarding student evaluation. It is part of multi-criteria decision-making, which plays a crucial role when assessment involves more than one evaluation criterion (Pulungan, 2017).

Interpretation Of Scores

After scoring MCQs, the next step is to interpret the scores to produce a meaningful grade. Scoring is a part of measurement, while grading provides evaluative meaning. Husein et al. (2025) define assessment as a comprehensive concept that integrates measurement results into values or criteria. The conversion from raw scores to grades is carried out by comparing student responses with the predetermined answer key. Scores may be presented on numerical scales such as 0–10 or 0–100, or letter grades such as A–E (Zulkifli, 2009). Example: If a student's raw score is 20 out of 25, and the grading scale is 0–100, the formula is: Raw Score /Maximum Score x 100 becomes 20/25x 100 = 80, A score of 80 can be classified as "Good" according to the applicable grading standard.

Result

Based on the discussion above, the following conclusions can be drawn:

1) Multiple-choice questions consist of a stem and two to five answer options, one of which is correct (the answer key) and the rest are distractors.
2) Advantages of MCQs include the ability to measure a wide range of cognitive skills and objective, efficient scoring. Limitations include the time required to construct quality items and the difficulty in creating effective distractors.
3) MCQ construction should follow established guidelines in terms of (a) content, (b) construction, and (c) language.
4) The most common scoring method for MCQs is *dichotomous scoring* (1 for correct, 0 for incorrect).
5) Four alternative scoring methods for MCQs include: 1) Counting correct answers 2) Applying penalties for incorrect answers 3) Giving bonuses for unanswered questions 4) Applying weighted scoring based on cognitive level.
6) The output of scoring is a raw score, which must be converted into a grade or classification using either criterion-referenced or norm-referenced approaches.

References

Ahiri, J. (2018). Keefektifan Bimbingan Tes Dalam Mengatasi Ketimpangan Skor Tes Akuntansi Ditinjau Dari Kecemasan Siswa. Jurnal Teknodik, 21(3), 140-163.

Akbar, A. (2020). Kemampuan Mahasiswa Dalam Penyusunan Soal Pilihan Ganda. *Attadib: Journal of Elementary Education*, *4*(1), 44-53.)

Ardania, R., Fitriyah, L. A., & Kuswanti, N. (2022). Pengembangan Instrumen Soal Pilihan Ganda Berbantu Aplikasi Quizizz Materi Sistem Pernapasan Pada Manusia Untuk Peserta Didik Kelas VIII SMP. *Discovery: Jurnal Ilmu Pengetahuan*, *7*(1).

Arif, M. (2014). Penerapan aplikasi anates bentuk soal pilihan ganda. *Jurnal Ilmiah Edutic: Pendidikan dan Informatika*, *1*(1).

Ariyanti, E., & Bhakti, Y. B. (2020). Perbandingan Bentuk Tes Pilihan Ganda dan Teknik Penskoran Terhadap Reliabilitas Tes Mata Pelajaran Kimia. *Titian Ilmu: Jurnal Ilmiah Multi Sciences*, *12*(2), 66-76.

Asrul, Rusydi Ananda, dan Rosnita. (2015) Evaluasi Pembelajaran. Bandung: Cipta Pustaka Media,

Bhakti, Y. B. (2015). Pengaruh jumlah alternatif jawaban dan teknik penskoran terhadap reliabilitas tes. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, *5*(1)

Crocker, L., & Algina, J. (1986). Introduction to Classical and Modern Test Theory. Florida: Holt, Rinehart, and Winston, Inc.

Daniel, M. (2019). Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance. Academic Medicine, https://doi.org/10.1097/ACM.0000000000002618.

Elis Ratna Wulan, E., & Rusdiana, A. (2015). Evaluasi pembelajaran.

Hazraini. (2017) "Upaya Meningkatkan Kompetensi Guru Kelas Dalam Penyusunan Soal Pilihan Ganda Yang Baik dan Benar Melalui Pendampingan Berbasis KKG Semester Satu Tahun Pelajaran 2017/2018 di SD Negeri 40 Cakranegara", Jurnal Pendidikan Mandala (JUPE), Vol. 2 No. 2 Desember 2017. https://doi.org/http://dx.doi.org/10.31332/atdb.v8i2.411

Husein, A., Sahrul, S., & Nugraheni, F. W. (2025). Penilaian Proyek Penguatan Profil Pelajar Pancasila Pada Kurikulum Merdeka. *Teaching and Learning Journal of Mandalika (Teacher) e-ISSN 2721-9666*, *6*(1), 190-195.

Ibrahim, I., & Muslimah, M. (2021). Tekhnik Pemeriksaan Jawaban, Pemberian Skor, Konversi Nilai dan Standar Penilaian. *Jurnal Al-Qiyam*, *2*(1), 1-9.

Kadir, A. (2015). Menyusun dan menganalisis tes hasil belajar. *Al-TA'DIB: Jurnal Kajian Ilmu Kependidikan*, *8*(2), 70-81.

Khaerudin, K. (2016). Teknik Penskoran Tes Obyektif Model Pilihan Ganda. *Madaniyah*, *6*(2), 195078.

Pulungan, I. R. (2017). Konsumen Kendaraan Bermotor Menggunakan Metode Scoring System Pada Pt . Adira Finance. 84–87.

Purwasih, J. H. G. (2020). Kendala calon pendidik dalam membuat soal pilihan ganda higher order thinking (HOT). *Jurnal Sosial Humaniora (JSH)*, *13*(1), 12-22.

Putri, H., Susiani, D., Wandani, N. S., & Putri, F. A. (2022). Instrumen penilaian hasil pembelajaran kognitif pada tes uraian dan tes objektif. *Jurnal Papeda: Jurnal Publikasi Pendidikan Dasar*, *4*(2).

Rosyidi, D. (2020). Teknik dan instrumen asesmen ranah kognitif. *Tasyri': Jurnal Tarbiyah-Syari'ah Islamiyah*, *27*(1), 1-13.

Soulisa Irwan, dkk, (2022), *Evaluasi Pembelajaran*, Bandung: Widina Bhakti Persada Bandung

Susilo, J. (2014). Teknik penyusunan soal pilihan ganda untuk meningkatkan kualitas mata uji kediklatan dan mata uji kompetensi. *Swara Patra: Majalah Ilmiah PPSDM Migas*, *4*(4).

Syahputra, A. T., & Arsyam, M. (2020). Pemberian skor dan sistem penilaian dalam pembelajaran.

Wartoni, W., & Benyamin, P. I. (2020). Strategi pengembangan tes objektif (pilihan ganda). *Diegesis: Jurnal Teologi*, *5*(1).

Widoyoko, E. P. (2017). Evaluasi program pelatihan. *Yogyakarta: Pustaka Pelajar*.

Yusron, A., Hanida, R. S., & Marfu'ah, S. (2024). Valitidas Tes Objektif Bentuk Pilihan Ganda. *Jurnal Pendidikan Matematika*, *1*(3), 10-10.

Zaenal, A. (2016). Evaluasi pembelajaran. Bandung Remaja Rosdakarya.