Volume 77 No 1 (2025)
Wahana : Tridarma Perguruan Tinggi
ISSN : 2654-4954 (online) | 0853-4403 (Print)
http://jurnal.unipasby.ac.id/index.php/whn

Analysis of Chemistry Multiple-Choice Test Items on the Periodic Table and Periodic Properties of Elements at Srijaya Negara Senior High School

Revnika Fethya [a, 1], **Effendi Nawawi** [b, 2]
[1,2] Program Studi Pendidikan Kimia, Universitas Sriwijaya, Palembang, Indonesia
email coresponden author : revnikabastian@gmail.com

Abstrak

Penelitian ini memiliki tujuan untuk mengevaluasi mutu butir soal pilihan ganda yang berkaitan dengan materi Sistem Periodik Unsur dan Sifat Periodik Unsur di SMA Srijaya Negara. Pendekatan yang digunakan yaitu deskriptif kuantitatif dengan teknik dokumentasi, observasi, dan wawancara. Sampel dalam penelitian ini berupa soal-soal yang telah diujikan kepada siswa. Hasil penelitian menunjukkan bahwa sejumlah soal belum memenuhi kriteria kualitas yang diharapkan. Diperlukan revisi dan pengembangan kemampuan guru dalam merancang alat evaluasi pembelajaran. Perbaikan terhadap soal serta pengembangan profesional guru menjadi langkah penting untuk meningkatkan kualitas evaluasi.

Abstract

This study aims to evaluate the quality of multiple-choice test items related to the Periodic Table of Elements and Periodic Properties of Elements material at SMA Srijaya Negara. A descriptive quantitative approach was employed using documentation, observation, and interview techniques. The sample consisted of test items that had been administered to students. The results indicate that several questions did not meet the expected quality criteria. Revisions and improvements in teachers' competence in developing assessment instruments are necessary. Improving test items and enhancing teacher professionalism are essential steps to improve the quality of evaluation.

Article History
Accepted : 14 May 2025
Approved: 25 May 2025

Kata kunci:
Analisis Soal, Sistem Periodik Unsur, Evaluasi Pembelajaran

Keywords:
Item Analysis, Periodic Table, Learning Evaluation

Introduction

Education represents a conscious, earnest, and systematically organized effort to develop students' potential, playing a vital role in preparing future generations to contribute to societal progress and national prosperity (Fanny, Susiloningsih, & Irianto, 2022). In the educational process, learning evaluation serves as a crucial component for measuring students' competency achievement and the effectiveness of instruction. Through periodic evaluations, teachers can assess the success of the learning process and determine the next steps whether to proceed with new material or to revisit previously taught concepts (Purniasari et al., 2021).

Student learning outcomes are generally assessed through various forms of examinations, including essay questions and multiple-choice tests, which provide feedback for both students and teachers regarding the successes and challenges encountered during the learning process (Saleh & Andini, 2021). Tests serve as the primary instruments in evaluating and measuring learning outcomes, particularly in the cognitive domain, as they are capable of reflecting students' understanding of the subject matter and their achievement of learning mastery in accordance with the applicable curriculum standards (Pamungkasih & Nawawi, 2021). In relation to this, Saragih, et al (2024) state that in chemistry learning, questions typically found at the end of sub-chapters in textbooks can be used. Therefore, the test items administered in this study for Grade X students were in the form of multiple-choice questions with five answer options. According to a study conducted by Sukardi (as cited in Zahiroh, 2021), multiple-choice tests meet the criteria of a good test as they possess characteristics such as objectivity, reliability, and the ability to identify students who have succeeded as well as those who have not. The advantages of this test include its ability to effectively measure students' mastery of the material, its high flexibility, and its facilitation of the correction process by making it quicker and easier.

Item evaluation is a critical series of steps to assess the quality of each question. High-quality questions can be reused, while those that are inadequate must be revised, and questions that do not meet the criteria should be replaced (Mawardi, Fuady, & Sunismi, 2023). This process plays a crucial role in ensuring that the questions effectively represent the competencies being assessed in accordance with the characteristics of the subject matter, as well as guaranteeing fairness for the students. Additionally, the analysis aims to improve the quality of the evaluation by refining or eliminating ineffective questions, while also assessing students' understanding of the taught material (Hamid et al., 2018). According to Sitaresmi, et al (2017), chemistry education at the senior high school level, particularly in the topics of the Periodic Table of Elements and the Periodic Properties of Elements, requires evaluation instruments that can accurately reflect students' conceptual understanding. This subject serves as a crucial foundation for mastering subsequent chemistry concepts. In this regard, the purpose of this study is to assess the quality of multiple-choice items applied to specific topics at SMA Srijaya Negara. The study focuses on analyzing the extent to which these questions are suitable for use based on established quality standards. The analysis is conducted based on several indicators, namely the validity, reliability, difficulty level, and

discriminating power of the items. The findings of this study are expected to contribute to the enhancement of the quality of chemistry learning evaluations at SMA Srijaya Negara.

Method

A quantitative descriptive approach was employed in this study to evaluate the quality of multiple-choice items related to the Periodic Table of Elements and Periodic Properties of Elements. The research was conducted at SMA Srijaya Negara, with the subjects consisting of multiple-choice question documents that had been administered to students. Data collection in this study was carried out using three main methods: (1) documentation, to obtain copies of the test items and students' answer sheets; (2) observation, to directly observe the implementation of the evaluation in the classroom; and (3) interviews, to gather information from chemistry teachers regarding the process of question development and implementation. The data obtained were analyzed using a quantitative approach with the assistance of Microsoft Excel. The aspects analyzed included validity, reliability, consistency, difficulty level, and discriminating power of the items. Validity was determined by calculating the correlation between the score of each item and the total test score, while reliability was measured using the Kuder-Richardson Formula 20 (KR-20). The difficulty level was analyzed by comparing the number of students who answered each item correctly, while the discriminating power was assessed by comparing the performance of the upper and lower groups. According to Arikunto (2012), the difficulty level of a test item is determined by comparing the number of students who answered the item correctly to the total number of test-takers. Based on this calculation, items can be classified into difficult, moderate, or easy categories, depending on the results obtained.

$$P = \frac{B}{N}$$

(Arikunto, 2012)

Explanation:

P = Item difficulty index

B = Number of students who answered the item correctly

N = Total number of students who took the test

Table 1

Item Difficulty Category

P Value Range	Description
0,00 – 0,30	Difficult
0,31 – 0,70	Moderate
0,71 – 1,00	Easy

(Arikunto, 2012)

According to Arikunto (2012), the discriminating power of a test item refers to its ability to distinguish between high-achieving students and low-achieving students. The calculation of discriminating power is conducted by comparing the proportion of students in the high-achieving group and the low-achieving group who answered the item correctly.

$$D = \frac{A - B}{N}$$

(Arikunto, 2012)

Explanation:

D = Discriminating power

A = Number of students in the upper group who answered correctly

B = Number of students in the lower group who answered correctly

N = Number of students in each group (upper and lower, typically 27% of the total number of students in each group)

Table 2

Item Discriminating Power Categories

Discriminating Power Range (D)	Category	Description
≥ 0,40	Very Good	The item is highly effective in distinguishing student ability
0,30 – 0,39	Good	The item is fairly effective and suitable for use
0,20 – 0,29	Fair	The item needs improvement or revision
< 0,20	Poor	The item should be discarded or thoroughly revised

(Arikunto, 2012)

Finding and Discussion

1. Observation

Based on the results of the multiple-choice test taken by 25 students at SMA Srijaya Negara, it was observed that most students experienced challenges in understanding the concepts of the Periodic Table of Elements and Periodic Properties of Elements. This was reflected in the low score distribution, where the majority of students scored below 3 out of a total of 10 questions, and six students even scored 0. The test items were predominantly at the cognitive levels of C3 (applying) and C4 (analyzing), which require analytical thinking skills. The students' low performance indicates that higher-order thinking skills (HOTS) have not yet developed optimally. Therefore, it is crucial to reassess both the quality of the test items and the instructional methods used. The learning process should focus more on in-depth conceptual understanding, for example through visual media, simulations of periodic trends, or problem-based discussions, so that students can better connect theoretical concepts with their application in analytical questions.

2. Interview

Through interviews conducted with the chemistry teacher at SMA Srijaya Negara, a comprehensive understanding was gained regarding how multiple-choice items are designed and utilized in the teaching and learning process for the topic of

the Periodic Table of Elements and Periodic Properties of Elements. The teacher stated that the construction of test items refers to a blueprint developed in accordance with the demands of the revised 2013 Curriculum, covering cognitive levels from remembering (C1) to analyzing (C4). Nevertheless, the teacher also acknowledged that not all items effectively differentiate students' levels of understanding.

In practice, the teacher combines questions from a question bank with self-constructed items, based on personal experience and understanding. The development process does not involve a thorough statistical analysis of item quality due to limitations in time and available resources. As a result, item validity and discriminating power among students have not been fully optimized. According to the interview, the teacher also noted that students often struggle with abstract concepts such as trends in ionization energy, atomic radius, and electronegativity, thereby necessitating a more visual and contextualized teaching approach.

The interview findings further revealed that student assessment is more focused on final scores rather than on evaluating the quality of individual test items. Consequently, the teacher is not always aware of which items are truly effective in measuring students' deep conceptual understanding. The teacher expressed a desire to participate in training or receive guidance on conducting quantitative item analysis, particularly in terms of item difficulty and discriminating power.

3. Test Item Analysis

a. Item Validity

The purpose of validity testing is to ensure that each test item accurately measures the intended competency. In this study, item validity was assessed by calculating the correlation between each item's score and the total test score. The resulting correlation values were then compared to the critical r-value from the correlation table at a 5% significance level with a sample size of 25, which is 0.388. Based on this analysis, it was found that 7 out of 10 test items had correlation values exceeding the critical threshold, thus classified as valid. Conversely, the remaining 3 items fell below the threshold and therefore did not meet the criteria for validity.

Table 3

Result of Validity Test

No	Calculated r	Table r	Description
1	0,1612	0,388	Not Valid
2	0,5211	0,388	Valid
3	0,7768	0,388	Valid
4	0,4282	0,388	Valid
5	0,0546	0,388	Not Valid
6	0,4833	0,388	Valid
7	0,4135	0,388	Valid
8	0,4415	0,388	Valid
9	0,2053	0,388	Not Valid
10	0,5276	0,388	Valid

Based on the data, it can be identified that the items are able to measure students' understanding of the material on the periodic table of elements and their periodic properties. This aligns with Arikunto's (2012) statement that a valid test item is one that accurately represents the data of the variable being measured. A similar study by Prastowo, et al (2020) also highlights the importance of test validity in the development of effective instruments, particularly in the context of chemistry. Based on the table, it can be seen that the validity of the instrument is quite good, as the majority of the items are considered valid.

b. Reliability

Table 4

Reliability Test Results

Component	Value
K (Number of Items)	10
Total Variance	2,4691
Total Variance	4,9261
Reliability (r11)	0,4691
Description	Tidak Reliabel

The purpose of the reliability test is to assess the extent to which the instrument produces consistent measurements for the same variable. In this study,

reliability was tested using the KR-20 formula because the test items are in the form of multiple-choice questions with true-false answers. Based on the data analysis, the obtained reliability coefficient is 0.4691. Compared to standard reliability criteria, this value is considered low, thus it can be concluded that the instrument as a whole cannot be considered reliable. According to Sugiyono (2013), low reliability indicates that the items need to be revised or reassessed to ensure high consistency in measuring students' competencies. This is supported by research by Wahyuni, et al (2022) which states that a reliable instrument will yield stable results even when used under different conditions or over time.

c. Test Item Difficulty Index

Table 5

Test Item Difficulty Index Results

Item Number	Correct Answer	Number of Student	Difficulty Index (P)
1	5	25	0,05
2	8	25	0,25
3	3	25	0,25
4	4	25	0,40
5	4	25	0,15
6	6	25	0,20
7	7	25	0,20
8	6	25	0,30
9	7	25	0,35
10	3	25	0,15

Based on the analysis of data obtained from SMA Srijaya Negara, it was found that the majority of multiple-choice items on the topic of the Periodic Table and Periodic Properties of Elements fall into the difficult category. Out of the ten test items analyzed, eight were classified as difficult, two as moderate, and none as easy. This condition indicates that, in general, the difficulty level of the test items is too high for most students. Several factors may contribute to this outcome, including students' lack of mastery of the material, unclear or overly

complex wording of the questions, and cognitive demands of the items that may exceed the average students' capabilities.

In test construction, it is essential to consider a balanced distribution of item difficulty levels to comprehensively assess students' abilities. The implication of these findings is the need for revision of certain test items to improve the overall quality of the assessment. Items that are too difficult should be reviewed in terms of content accuracy, clarity of wording, and the cognitive level required. Furthermore, educators must ensure that all relevant material has been taught and understood by the students prior to assessment. By doing so, the evaluation items will not only be content-valid but also capable of fairly distinguishing students' levels of understanding.

d. Discriminatory Power of Test Items

Item discrimination analysis is a crucial aspect in evaluating the quality of test items, particularly in assessing the ability of an item to distinguish between high-performing and low-performing students. In the learning evaluation process, effective discriminatory power enables test items to identify learners who have truly mastered the material being taught.

Table 6

Results of the Discriminatory Power Test for Each Item

Item Number	Discriminatory	Category
1	0,076923	Poor
2	0,230769	Fair
3	0,384615	Good
4	0,461538	Excellent
5	0,307692	Good
6	0,076923	Poor
7	0,307692	Good
8	0,461538	Excellent
9	0,384615	Good
10	0,230769	Fair

Based on the analysis of the data above, it is evident that the multiple-choice questions on the topic of the Periodic Table and Periodic Properties of Elements exhibit a range of discriminatory quality. Out of the ten items analyzed, two items fall into the “excellent” category, indicating that these questions are highly effective in distinguishing between students with high and low levels of understanding. Meanwhile, three items are categorized as “good,” meaning they are reasonably reliable as evaluation tools. However, three items fall into the “fair” category. Although still usable, these questions require improvement in terms of phrasing and alignment with the relevant basic competencies. Furthermore, two items are classified as “poor,” indicating a need for thorough revision.

These results suggest that the majority of the test items fall within the “good” to “excellent” categories. This is a positive indicator, showing that most of the items are functioning effectively as evaluation instruments capable of differentiating students based on their level of understanding. Nevertheless, the presence of items with low discriminatory power should not be overlooked. As noted by Sudjana (2005), items with low discriminatory power can compromise the overall validity of a test because they fail to provide meaningful information about the differences in student ability. Therefore, conducting item analysis based on discriminatory power is strongly recommended on a regular basis to ensure the quality of the test instruments. Teachers or item developers should pay close attention to the structure and relevance of the items with respect to the learning objectives, and ensure that the items are clear, unambiguous, and do not mislead students regardless of their actual level of understanding.

Conclusion

The results of the study revealed that, out of the ten test items examined, the majority met the validity criteria (seven items were valid). However, the reliability of the instrument was relatively low ($r_{1\ 1}$ = 0.4691), indicating that the internal consistency of the test items needs improvement. The level of item difficulty was generally high, posing a significant challenge for students and reflecting their limited mastery of the topic "Periodic Table and

Periodic Properties of Elements." The discriminatory power of the items varied, with several questions demonstrating good effectiveness in distinguishing between students of different ability levels. However, some items were identified as needing revision or elimination due to their suboptimal performance. Overall, the test items were not developed based on a comprehensive statistical analysis, which has impacted the overall quality of the evaluation instrument. Based on these findings, it is recommended that future research focus on developing and refining evaluation instruments through more rigorous trials and detailed statistical analysis. Establishing a standardized item bank, incorporating more contextually relevant teaching methods, and utilizing visual media are also suggested strategies. These approaches can enhance students' understanding of the subject matter and improve the accuracy of learning outcome assessments.

References

Arikunto, S. (2012). *Dasar-dasar evaluasi pendidikan* (Edisi Revisi). Jakarta: Bumi Aksara.

Fanny, A. M., Susiloningsih, W., & Irianto, A. (2022). Studi Literatur: Model Pembelajaran Kooperatif Tipe STAD Untuk Mengembangkan Karakter Gotong Royong Dalam Pembelajaran IPS. *Wahana: Tirdarma Perguruan Tinggi*, 74(2), 304-313.

Hamid, F. A., Manoppo, Y., & Kayadoe, V. (2018). Analisis Butir Soal Tes Akhir Semester (TAS) Kimia SMA Negeri II Ambon Tahun Pelajaran 2014/2015. *MJoCE*, 8(2), 89-93.

Mawardi, M. S., Fuady, A., & Sunismi. (2023). Analisis Butir Soal Pilihan Ganda Menggunakan Anates pada Penilaian Tengah Semester Kelas VII D SMP Negeri 1 Ngajum Kabupaten Malang. *Wahana: Tirdarma Perguruan Tinggi*, 75(1), 31-41.

Pamungkasih, R. S. N., & Nawawi, E. (2021). Analisis Kualitas Butir Soal Ujian Akhir Semester Ganjil Pada Mata Pelajaran Kimia Kelas X di SMA Negeri 8 Palembang Tahun Ajaran 2020/2021. *Prosiding Seminar Nasional Pendidikan IPA Tahun 2021*, 1(1), 123-130. Palembang: Universitas Sriwijaya.

Prastowo, A., Kusmiyati, K., & Wulandari, S. (2020). Analisis validitas dan reliabilitas instrumen tes pada materi kimia. *Jurnal Evaluasi Pendidikan, 11*(1), 34-42.

Purniasari, L., Masyukuri, M., & Ariani, S. R. T. (2021). Analisis Butir Soal Ujian Sekolah Mata Pelajaran Kimia SMA N 1 Kutowinangun Tahun Pelajaran 2019/2020 Menggunakan Model Iteman dan Rasch. *Jurnal Penelitian Pendidikan Kimia (JPK)*, 10(2), 205-214.

Purwanto, N. (2013). *Prinsip-prinsip dan teknik evaluasi pengajaran*. Bandung: Remaja Rosdakarya.

Saleh, M., & Andhini, L.W. (2021). Efektivitas Penggunaan Media Animasi Terhadap Hasil Belajar Siswa. *Wahana: Tirdarma Perguruan Tinggi*, 73(2), 236-247.

Saragih, A., & Siregar, R. (2024). Pengembangan soal evaluasi pembelajaran kimia berbasis buku teks untuk siswa kelas X. *Jurnal Riset Pembelajaran Kimia*, 9(2), 95–102.

Sitaresmi, K. S., Saputro, S., & Utomo, S. B. (2017). Penerapan Pembelajaran Berbasis Project Based Learning (PJBL) Untuk Meningkatkan Aktivitas dan Prestasi Belajar Siswa Pada Materi Sistem Periodik Unsur (SPU) Kelas X MIA 1 SMA Negeri 1 Teras Boyoali Tahun Pelajaran 2015/2016. *Jurnal Penelitian Pendidikan Kimia (JPK)*, 6(1), 54-61.

Sudjana, N. (2005). *Penilaian hasil proses belajar mengajar*. Bandung: Remaja Rosdakarya.

Sugiyono. (2013). *Metode penelitian pendidikan: Pendekatan kuantitatif, kualitatif, dan R&D*. Bandung: Alfabeta.

Wahyuni, S., Sari, D. M., & Putri, F. N. (2022). Pengembangan instrumen evaluasi pembelajaran kimia berbasis validitas dan reliabilitas. *Jurnal Pendidikan Sains, 10*(2), 121–129.

Zahiroh, F. (2021). Analisis efektivitas tes pilihan ganda dalam evaluasi pembelajaran. *Jurnal Evaluasi Pendidikan*, 12(1), 45–53.