Available online at
https://jurnal.uinsyahada.ac.id/index.php/Lavoisier/index
LAVOISIER: CHEMISTRY EDUCATION JOURNAL
Vol. 4 (2), 2025

DEVELOPMENT OF CHEMISTRY LEARNING OUTCOME ASSESSMENT INSTRUMENTS (MULTIPLE CHOICE) ON HYDROCARBON MATERIAL

Enjelika Manullang[1], Rabiah Afifah Daulay[2], Fathin Nadira[3], Kayra Aprilia Yosephin Sihaloho[4], Nazwa Azhara[5], Nazwa Nazihah[6]

[1-5]Faculty of Mathematics and Natural Sciences, Chemistry Education, Medan State University, Jalan Wiliam Iskandar Pasar V Medan Estate, 20221, North Sumatra, Indonesia

* Correspondence: rabiahdaulay@unimed.ac.id

Article History

Received 07 08th 2025
Revised 07 20th 2025
Accepted 12 12th 2025
Available Online 12 31st 2025

Keywords:
Learning Assessment
Objective Test
Evaluation Instrument,
Hydrocarbons
Validity and Reliability

Abstract

This study aims to develop an assessment instrument in the form of a multiple-choice objective test on hydrocarbon material in senior high school chemistry. The research method used is research and development (R&D), involving steps such as material analysis, test item construction, validation, field trials, and item analysis. The developed instrument consists of 30 items based on the learning outcomes in the "Kurikulum Merdeka". Validity testing using Pearson correlation shows that 29 out of 30 items are valid ($r \geq 0.388$; sig. 0.05). Reliability testing using Cronbach's Alpha resulted in a score of 0.965, indicating very high reliability. These findings suggest that the developed test instrument is valid and reliable for evaluating students' understanding of hydrocarbon concepts.

1. Introduction

Education is a learning activity that aims to develop and actualize the potential of students. The objectives to be achieved in learning activities are a form of student ability that occurs due to the interaction between students, educators, and the learning environment. The achievement of learning objectives can be determined through assessment activities that provide information about the abilities of students. Development is a process/method used to validate, develop existing products, or create new products. Richey and Kelin, as cited by Sugianto, describe development as a systematic study of how to design a product, develop/produce the design, and evaluate the performance of the product, with the aim of obtaining empirical data that can be used as a basis for creating products, tools, and models used in

DOI : 10.24952/lavoisier.v4i2.16256

learning or non-learning contexts. Assessment is the most important thing in education. With assessment, we can determine the level of success of learning. Similarly, Farida states that assessment to determine learning outcomes and the process of progress needs to be carried out. The meaning of assessment focuses on the stage of calculating and measuring results. Before conducting an evaluation, the first stage begins with assessment, therefore assessment has a smaller scope than evaluation. (Musfirah et al., 2025).

Assessment is an integral part of the overall teaching and learning process. Assessment can be used as one of the factors in determining the success of the learning process and outcomes, not just as a way to assess students' success in mastering subject matter. Teachers, as one of the main factors in improving the quality of education, must be skilled in assessing the learning process and outcomes of students, such as creating assessment tools, processing assessment data, diagnosing learning difficulties, and utilizing assessment results. There are four main elements in assessment, namely the object being assessed, the criteria used as a benchmark, data about the object being assessed, and judgment. The assessment process includes determining the object to be assessed, creating or determining the measurement criteria, collecting data, and making a decision. The tools used as a means of assessment can be tests or non-tests. According to Hamzah and Satria (2012: 111), a test is a set of stimuli given to a person with the intention of obtaining answers that form the basis for determining a score. The score, which is based on a representative sample of the test taker's behavior, is an indicator of the extent to which the person being tested has the characteristics being measured. A test is a way to conduct an assessment in the form of a task or series of tasks that must be completed by a child to produce a score on the child's behavior or achievement, which can be compared with the scores achieved by other children or with a predetermined standard score. (Inteni et al., 2013).

THEORETICAL STUDY

The types of tests based on question format are as follows:

a. Subjective Tests

Subjective tests are generally in essay format. Essay tests are a type of learning progress test that requires answers in the form of discussion or written explanation. The questions are characterized by words such as: describe, explain, why, how, compare, conclude, and so on.

b. Objective Tests

Objective tests are tests that can be administered objectively. This is intended to overcome the weaknesses of essay-type tests.

1) Types of Objective Tests
   a) True/False Tests

DOI : 10.24952/lavoisier.v4i2.16256

The questions are in the form of statements. Some of these statements are true and some are false. The person being tested is tasked with marking each question by circling the letter b if they think the question is true and circling the letter s if they think the question is false.

2) Multiple Choice Tests

A multiple-choice test consists of an incomplete statement or description of a concept. To complete it, the test taker must choose one of several possible answers provided, consisting of one correct answer (the key answer) and several distractors.

3) Matching Test

A matching test can be replaced with the terms “comparing,” “matching,” “pairing,” or “matching.” A matching test consists of a series of questions and a series of answers. Each question has an answer listed in the series of answers.

4) Completion Test

A completion test is commonly referred to as a fill-in-the-blank test, a completion test, or a completion exercise. A completion test consists of sentences with parts that have been omitted. The omitted parts, or the parts that must be filled in by the student, are the concepts that we want the student to understand.

A measuring instrument can be said to be valid if it can measure what it is intended to measure accurately. In terms of validity and reliability, it is certainly influenced by (1) the instrument, (2) the subject being measured, and (3) the person conducting the measurement. In terms of measurement, especially in education, the most important thing is accurate measurement results. This is because inaccurate or insufficient measurement results will provide incorrect information, leading to incorrect conclusions. A test is a systematic procedure, meaning that (a) the items in the test are arranged according to certain methods and rules, (b) the test administration and scoring procedures must be clear and specified in detail, and (c) everyone who takes the test must receive the same items under comparable conditions. From the above opinion, a test is a systematic procedure created in the form of standardized tasks and given to individuals or groups to work on, answer, or respond to, either in written, oral, or behavioral form using a numerical scale or category system. Tests are also measuring tools that have objective standards.

Validity comes from the word validity, which means the extent to which a measuring instrument (test) is accurate and precise in performing its measuring function. A test is said to have high validity if the tool performs its measuring function accurately or provides measurement results that are in accordance with the purpose of the measurement. This means that the measurement results accurately reflect the facts or actual conditions of what is being measured. The concept of test validity can be divided into three types, namely content validity, construct validity, and empirical

DOI : 10.24952/lavoisier.v4i2.16256

validity or criterion validity. The content validity of a test concerns the extent to which a test measures the level of mastery of certain content or material that should be mastered in accordance with the teaching objectives. In other words, a test with good content validity is a test that truly measures the mastery of material that should be mastered in accordance with the teaching content listed in the teaching program outline (gbpp). To determine whether a test is valid or not, the test grid must be examined to ensure that the test questions represent or reflect the entire content or material that should be mastered proportionally. Therefore, the content validity of a test does not have a specific value that is calculated statistically, but it is understood that the test is valid based on a review of the test grid. Therefore, content validity is actually based on logical analysis, not a validity coefficient calculated statistically. Meanwhile, reliability comes from the word reliability, which means the extent to which the results

2. Instrument Development Method

This research is a type of research and development (R&D). According to Borg and Gall (Sugiyono, 2013: 4), research and development is a research method used to develop or validate products used in education and learning. The main output of this research is a multiple-choice objective test instrument for high school chemistry. The following are important steps that must be taken in developing objective test questions:

a Analyzing the Characteristics of the Material to be Tested

For example, textbook material will have an impact on objective test instruments that are different from those used for non-textbook material.

b. Developing Distractors

Distractors are a separate point because they are a distinctive feature of objective questions, distinguishing them from other types of questions.

c. Arrange the Answer Format

The position of the correct answer, the position of the distractors, and the position of the incorrect answers are important to arrange so that the objective test can function properly.

d. Review the Test Instrument

Before the test is administered, it is necessary to review or validate the instrument. Validation can be carried out by experts or practitioners using various methods, such as Delphi.

e. Field Test

After going through the validation stage, the instrument needs to be field tested. In this trial, it is necessary to check the possibility of students revising or changing their answers during the test.

DOI : 10.24952/lavoisier.v4i2.16256

f. Item Analysis

Test validity, reliability, discrimination, and difficulty level as with instrument testing in general.

g. Compare Test Results or Trials

With test results from other types of instruments, such as essays, this is something that has not been widely done by teachers or instrument developers. Triangulation of assessment results from one type of test with another is necessary to determine the reliability of the developed instrument.

3. Results and Discussions

The instrument developed consists of 30 multiple-choice questions compiled based on a grid in accordance with the learning outcomes in the Merdeka Curriculum. The questions cover hydrocarbons, including the characteristics of carbon atoms, saturated hydrocarbon bonds, positional structural isomers, the mechanics of addition and substitution reactions, and the nomenclature of hydrocarbon compounds.

The following questions are being tested and can be accessed through the following Google form:https://docs.google.com/forms/d/e/1FAIpQLSfovx_yU9JkD5q96YAl1Q04hzMPnHm5VUTg1Q2dnIxhp FQ-lQ/viewform?usp=dialog

No.	learning outcomes (CP)	Learning Objectives (AT)	Question Indicator	Material	Shape	Number
1	Students understand the properties of organic compounds based on functional groups and macromolecular compounds. Students understand the properties of organic compounds based on functional groups and macromolecular compounds.	Describing the unique characteristics of carbon atoms in forming hydrocarbon compounds	Identify the elements C, H, and O in carbon compounds through experiments.	The uniqueness of carbon atoms	multiple choice	1
			Describe the characteristics of carbon atoms in carbon compounds. .		multiple choice	

DOI : 10.24952/lavoisier.v4i2.16256

			Distinguish between primary, secondary, tertiary, and quaternary carbon atoms		multiple choice	
2	Managing and analyzing the structure and properties of hydrocarbon compounds based on understanding the characteristics of carbon atoms and the classification of compounds	Understanding hydrocarbon compounds unsaturated	Classify hydrocarbon compounds based on bond saturation	Saturation of hydrocarbon bonds	multiple choice	
3.	Analyzing the possibility of isomers in hydrocarbon compounds	Determining the structural and positional isomers of short-chain hydrocarbon compounds	Determining the type and number of isomers and positional isomers	Structural and positional isomers	multiple choice	
			Classify the types of hydrocarbon isomers based on their structure and position		multiple choice	
4.	Analyzing reactions and structural changes in hydrocarbons	Explaining the steps addition reactions and Simple substitution	Explaining the mechanism of addition and substitution reactions	The mechanism of addition and substitution	multiple choice	
5.	Interpreting the names names and of hydrocarbon compounds	Participants learners determine the name of hydrocarbon compounds from their	Compose IUPAC names and the of hydrocarbons	Naming hydrocarbon compounds	multiple choice	

This instrument was tested on 26 students in grade 12 at SMK Negeri 1 Percut Sei Tuan. The test results were processed using SPSS version 26 for validity and reliability testing. Validity was tested using the Pearson product moment correlation between the score for each item and the total score. The validity test results showed that, out of 29 items, 30 items were valid with a calculated r ≥ table r = 0.388 and sig. 0.05.

DOI : 10.24952/lavoisier.v4i2.16256

		Total
A01	Pearson Correlation	.598**
	Sig. (2-tailed)	.001
	N	26
A02	Pearson Correlation	.746**
	Sig. (2-tailed)	.000
	N	26
A03	Pearson Correlation	.695**
	Sig. (2-tailed)	.000
	N	26
A04	Pearson Correlation	.747**
	Sig. (2-tailed)	.000
	N	25
A05	Pearson Correlation	.327
	Sig. (2-tailed)	.103
	N	26
A06	Pearson Correlation	.635**
	Sig. (2-tailed)	.000
	N	26
A07	Pearson Correlation	.733**
	Sig. (2-tailed)	.000
	N	26
A08	Pearson Correlation	.746**

DOI : 10.24952/lavoisier.v4i2.16256

	Sig. (2-tailed)	.000
	N	26
A09	Pearson Correlation	.503**
	Sig. (2-tailed)	.009
	N	26
A10	Pearson Correlation	.490*
	Sig. (2-tailed)	.011
	N	26
A11	Pearson Correlation	.727**
	Sig. (2-tailed)	.000
	N	26
A12	Pearson Correlation	.564**
	Sig. (2-tailed)	.003
	N	26
A13	Pearson Correlation	.849**
	Sig. (2-tailed)	.000
	N	26
A14	Pearson Correlation	.695**
	Sig. (2-tailed)	.000
	N	26
A15	Pearson Correlation	.668**
	Sig. (2-tailed)	.000
	N	26

DOI : 10.24952/lavoisier.v4i2.16256

A16	Pearson Correlation	.506**
	Sig. (2-tailed)	.008
	N	26
A17	Pearson Correlation	.751**
	Sig. (2-tailed)	.000
	N	26
A18	Pearson Correlation	.709**
	Sig. (2-tailed)	.000
	N	26
A19	Pearson Correlation	.758**
	Sig. (2-tailed)	.000
	N	26
A20	Pearson Correlation	.758**
	Sig. (2-tailed)	.000
	N	26
A21	Pearson Correlation	.811**
	Sig. (2-tailed)	.000
	N	26
A22	Pearson Correlation	.849**
	Sig. (2-tailed)	.000
	N	26
A23	Pearson Correlation	.625**
	Sig. (2-tailed)	.001

DOI : 10.24952/lavoisier.v4i2.16256

	N	26
A24	Pearson Correlation	.821**
	Sig. (2-tailed)	.000
	N	26
A25	Pearson Correlation	.700**
	Sig. (2-tailed)	.000
	N	26
A26	Pearson Correlation	.758**
	Sig. (2-tailed)	.000
	N	26
A27	Pearson Correlation	.932**
	Sig. (2-tailed)	.000
	N	26
A28	Pearson Correlation	.795**
	Sig. (2-tailed)	.000
	N	26
A29	Pearson Correlation	.470*
	Sig. (2-tailed)	.015
	N	26
A30	Pearson Correlation	.647**
	Sig. (2-tailed)	.000
	N	26
Total	Pearson Correlation	1

DOI : 10.24952/lavoisier.v4i2.16256

	Sig. (2-tailed)	
	N	26

Reliability was tested using SPSS version 26 with Cronbach's alpha. The analysis results showed a test score of 0.965. This means that the instrument has very high reliability, because the alpha value is ≥ 0.9.

Case Processing Summary

		N	%
Cases	Valid	25	96.2
	Excluded[a]	1	3.8
	Total	26	100.0

Reliability Statistics

Cronbach's Alpha	N of Items
.965	30

The assessment instrument in this study was developed to measure student learning outcomes in hydrocarbon material, specifically through multiple-choice questions. The development process was carried out using a Borg & Gall theory-based research and development (R&D) approach. The steps included material analysis, grid and question preparation, expert validation, field testing, and statistical analysis for validity and reliability testing. Multiple-choice questions were chosen because of their objectivity and efficiency in measuring a broader cognitive domain. The developed instrument included 30 questions based on the learning outcomes in the Merdeka Curriculum. The hydrocarbon material tested includes the characteristics of carbon atoms, bond saturation, structural and positional isomers, addition and substitution reaction mechanisms, and compound nomenclature. Each question is directly related to learning indicators and objectives. This shows that the question development process has taken into account the logical relationship between content, learning objectives, and instrument form.

The validity test was conducted using Pearson's correlation between the score for each question and the total score. The test results showed that 29 of the 30 questions were valid with a calculated r value ≥ 0.388 and significance ≤ 0.05. This indicates that most of the questions were able to measure what should be measured according to the learning indicators. The one invalid question was most likely due to inaccurate wording, answer choices that were too easy to guess, or the question not matching the learning indicators.

DOI : 10.24952/lavoisier.v4i2.16256

The reliability of the instrument was tested using Cronbach's Alpha method in SPSS version 26. The analysis results showed an alpha value of 0.965, which is very high. This means that the internal consistency between the questions is very strong, and this instrument will provide stable results when used on different groups but under the same measurement conditions. This high reliability value reinforces the quality of the instrument in terms of reliability. The field trial was conducted on 26 students in grade XII at SMK Negeri 1 Percut Sei Tuan. Although the number of respondents was relatively small, this trial was sufficient to provide an initial overview of the instrument's performance. However, to improve the generalization of the results, it is recommended that the instrument be further tested on a larger and more heterogeneous sample. In addition, an analysis of the level of difficulty and discriminating power of each question needs to be added to complete the information on the quality of the questions.

The author also proposes that triangulation of assessment results is important, namely comparing the results of multiple-choice tests with other forms of tests such as essays. This step is still rarely done by teachers, even though it is important to know whether the learning outcome information obtained is consistent across assessment forms. This study shows an awareness of the importance of using various forms of tests in an effort to improve the accuracy of learning evaluation.

5. Conclusions

Based on the results of research and data analysis, it can be concluded that the multiple-choice objective test instrument developed for hydrocarbon material has met the criteria for validity and reliability. Of the 30 questions tested, 29 were declared valid based on Pearson's correlation test with a significance of ≤ 0.05. Meanwhile, the reliability value obtained was 0.965 based on Cronbach's Alpha calculation, indicating that the instrument has very high internal consistency, making it suitable for use as a tool for evaluating student learning outcomes.

This instrument was designed based on the learning outcomes in the Merdeka Curriculum, with material covering the characteristics of carbon atoms, bond saturation, isomers, reaction mechanisms, and the nomenclature of hydrocarbon compounds. Therefore, this instrument can be used as a reference for teachers in assessing students' understanding comprehensively and objectively. However, further development is still recommended, such as conducting trials on a wider sample, analyzing discriminatory power, and cross-validating with other test forms to strengthen the reliability and usefulness of the instrument.

DOI : 10.24952/lavoisier.v4i2.16256

References

Damayanti, R. Huda, N. Hermina. D. (2024). Pengolahan Hasil Non-Test Angket,Observasi, Wawancara Dan Dokumenter. *Student Research Journal,* 2 (3), 259 - 273.

Fauziyah, A. Sakinah, Z. A. Mariyanto. Juansah, D. E. (2023). Instrumen Tes Dan Non Tes Pada Penelitian. *Jurnal Ilmiah Pendidikan Dasar,* 8 (3), 6538- 6548.

Inteni, K. A. S. Candiasa,I. M. Suarni, N. K. Pengembangan Instrumen Tes Objektif Pilihan Ganda Yang Diperluas Berbasis *Web* Untuk Mata Pelajaran Tik Kelas XI SMAN Di Kabupaten Karangasem. *Journal Program Pascasarjana Universitas Pendidikan Ganesha,* 3 (1), 01-09.

Khaerudin. (2016). Teknik Penskoran Tes Obyektif Model Pilihan Ganda. *Jurnal Madaniyah*, 2 (11),183-200.

Khaerudin. (2017). Administrasi, Analisis Butir, Dan Kaidah Penulisan Tes. *Jurnal Madaniyah*, 1 (12), 97-128.

Musfirah, N. Nurbaya. Nursalam. Rasyid, A. (2025). Pengembangan Instrumen Hasil Penilaian Belajar Tes Dan Non Tes. *Socius: Jurnal Penelitian Ilmu-Ilmu Sosial,* 2 (11), 120-125.

Ramadhan, M. F.Siroj, R. A.Afgani,M. W. (2024). Validitas and Reliabilitas. *Journal on Education,* 6 (2), 10967-10975.

Saripudin, A dan Durtam. (2024)**.** Pengembangan Instrumen *Non Tes* pada Penilaian Aspek Perkembangan Motorik Anak Usia Dini serta Implementasinya pada Pembelajaran PAUD. *Jurnal Tumbuh kembang: Kajian teori dan Pembelajaran PAUD,* 11 (1), 88-101.

Susanto, s. (2023). Pengembangan Alat dan Teknik Evaluasi Tes Dalam Pendidikan. *Jurnal Tarbiyah Jamiat Kheir, 1(1),* 51-60.

Wartoni dan Benyamin, P. I. (2021). Strategi Pengembangan Tes Objektif (Pilihan Ganda). *Jurnal STTBI,* 1(1), 01-08.

Zainab, S dan Wilujeng, I. (2016). Pengembangan Instrumen Penilaian Tes Objektif Pilihan Ganda Untuk Mengukur Penguasaan Materi Ajar Gerak Lurus Dan Keterampilan Proses Sains Siswa Sma. *Jurnal Pendidikan Fisika,* 5 (2), 106-113

DOI : 10.24952/lavoisier.v4i2.16256