JISE 13 (3) (2024): 124-133

Journal of Innovative Science Education

https://journal.unnes.ac.id/journals/jise/

Development and Validation of a Three-Tier Test for Identifying Misconceptions in Organic Chemistry Course

Asyti Febliza[1], Asep Kadarohman[1]✉, Siti Aisyah[1], Norazilawati Abdullah[2]

[1]Department of Science Education, Faculty of Mathematics and Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[2]Faculty of Human Development, Universiti Pendidikan Sultan Idris, Tanjong Malim, Malaysia

Article Info

Article History :
March 2024
Accepted
June 2024
Published
December 2024

Keywords:
Organic chemistry; Reliability; Three-tier; Validity

Abstract

Misconceptions about organic chemistry topics still exist among university students. Identifying these misconceptions requires appropriate instruments. This study aims to develop and validate a three-tier test instrument for identifying misconceptions in organic chemistry lectures. The topics covered in the test include alkenes, alkynes, aromatic compounds, alkyl halides, alcohols, and ethers. This research follows a design and development approach, specifically adopting a product and tool research type with a focus on tool development & use, where validity is a research design concern. A total of 42 three-tier items were developed, combining multiple-choice questions with open reasoning and Certainty of Response Index (CRI) technique for confidence level assessment using a scale of 0, 1, 2, 3, 4, and 5. The content validity test by three expert judgments showed an average I-CVI of 0.98, with an average item proportion of 0.98. These results showed test item were valid. The test items were piloted with 25 students to examine empirical validity, resulting in 40 valid items and 2 invalid items. The Cronbach's Alpha reliability test yielded a value of 0.953, indicating very high reliability for the three-tier test instrument. Therefore, the three-tier test instrument can be used effectively to collect data and detect student misconceptions in organic chemistry lectures.

✉ correspondence:
Jl. Dr. Setiabudi No.229, Isola, Kec. Sukasari,
Kota Bandung, Jawa Barat 40154, Indonesia
E-mail: kadar@upi.edu

p-ISSN 2252-6412
e-ISSN 2502-4523

INTRODUCTION

Organic chemistry is an essential element in the fields of science, technology, engineering, and mathematics (STEM). Although the importance of organic chemistry is acknowledged by the scientific community, there is a widespread perception among students that organic chemistry is challenging. Organic chemistry presents numerous challenges due to the difficulty of understanding abstract concepts (Ewais & Troyer, 2019). Various studies on learning difficulties in organic chemistry have been conducted by researchers. On the topic of resonance, students struggle with the concept of resonance structures (Duis, 2011; Linenberger et al., 2011). Other research has revealed students' difficulties in classifying bases or nucleophiles and understanding reaction mechanisms in alkyl halides (Arellano, 2014). Omwirhiren et al., (2016) found misunderstandings regarding polymerization, identifying isomers, recognizing aromatic compounds from compressed structural formulas, differentiating between substitution and addition reactions, and applying IUPAC rules for naming organic compounds. Taagepera (2000) identified misconceptions about physical properties, chemical properties, and types of organic reactions.

These research findings demonstrate that misconceptions about organic chemistry topics still exist among university and high school students. Many studies have noted that misconceptions or a lack of understanding of fundamental topics presented in previous chemistry courses are significant barriers to students' success in organic chemistry (Fischer et al., 2019; Lindsay et al., 2017; Stone et al., 2018). hese misconceptions stem from students' learning processes or classroom experiences, not the original concepts they learned in class. Students often make mistakes when applying the terms, concepts, and algorithms they learned in class (Rushton et al., 2008). Additionally, Sani (2010) as cited in (Omwirhiren et al., 2016) states that the main factors identified as causes of these misconceptions include ineffective teaching methods, lack of exposure to laboratory activities, poor organizational skills, and limited familiarity with problem-solving procedures.

To identify misconceptions, an appropriate instrument is required. Instruments can take the form of diagnostic tests with multiple-choice questions. According to (Sadler & Sonnert, 2016) this type of test is most commonly used to diagnose student misconceptions. However, traditional multiple-choice questions generally have limitations in diagnosing student misconceptions. Therefore, researchers have modified multiple-choice questions to include reasoning. These modified diagnostic test instruments can be in the form of two-tier, three-tier, four-tier, or five-tier diagnostic tests (Rokhim, 2023). To diagnose students' misconceptions, a valid and reliable instrument is needed. Therefore, the development and validation of diagnostic test instruments are required before these instruments can be used.

Various studies on the development of instruments to detect misconceptions in chemistry concepts have been conducted by researchers. Mutlu & Sesen (2015) developed a valid and reliable two-tier diagnostic test to assess undergraduate students' understanding of several concepts in the context of general chemistry courses. Siswaningsih et al., (2017) also developed a two-tier diagnostic test to identify students' misconceptions about the concept of moles. The development and application of a three-tier test as a valid and reliable tool for diagnosing students' misconceptions about several fundamental carbohydrate concepts were carried out by Milenković et al (2016). Meanwhile, Jusniar et al (2020) developed a three-tier diagnostic instrument on chemical equilibrium for high school students. Additionally, Habiddin & Page (2019) developed and validated a four-tier diagnostic instrument to explore first-year undergraduate students' understanding of chemical kinetics. A study on the development of a five-tier test instrument was conducted by Rizky Fajar et al (2023) to detect high school students' misconceptions on the topic of hydrocarbons.

The development of four-tier and five-tier instruments is a refinement of three-tier (Anam et al., 2019; Cetin-Dindar & Geban, 2011). However, the three-tier instrument is actually sufficient to measure misconceptions. As long as, this three-tier instrument used in previous research uses three-tier with confidence only using “yes" or "no". This choice of confidence level does not yet show an extreme decision, while it is possible that students have confidence even if only a little. Therefore, it is necessary to develop a three-tier instrument with a CRI level of confidence.

Based on previous research, it is evident that the development of a three-tier test instrument to identify misconceptions in Organic Chemistry 1 lectures has not yet been found. However, three-tier test instrument in previous research rarely used Certainty of Response Index (CRI). Therefore, this study aims to develop and validate a three-tier test instrument to identify misconceptions in organic chemistry lectures with CRI. The topics covered in the test include alkenes, alkynes, aromatic compounds, alkyl halides, alcohols, and ethers. The three-tier test instrument adds reasons and a confidence level for the chosen answers and reasons. If students choose incorrect answers and reasons but have a high level of confidence, this indicates a misconception (Jusniar et al., 2020). The three-tier test instrument is already capable of diagnosing misconceptions, making this test instrument easy to use (Cetin-Dindar & Geban, 2011).

In this study, the tool used to detect misconceptions is a three-tier test, which combines multiple-choice questions with open reasoning and the Certainty of Response Index (CRI) technique. This technique was initially developed by (Hasan, 1999) and later modified by (Potgieter et al., 2005) as well as (Hakim et al., 2012). Data analysis was conducted by evaluating students' answers. If students provide correct answers with appropriate reasoning, even with low confidence, it indicates that they understand the concept but lack confidence (Hakim et al., 2012). The CRI technique allows for distinguishing between students who have a good understanding of concepts but are uncertain and those who experience misconceptions.

METHODS

This study is a design and development research (DDR) (Richey, 2007). Design and development research is divided into two main types: 1) Product and Tool Research: Focuses on studying the design and development process of products and tools used. This can include evaluating the impact of the designed product or tool. 2) Model Research: Involves studying the models or frameworks used in the design and development process. This includes testing the effectiveness of certain models in supporting design and development (Richey, 2007).

This study adopts the product and tool research type, with a focus on tool development and use, where the research design concern is validity. This research develops a three-tier test instrument on the topics of alkenes, alkynes, aromatic compounds, alkyl halides, alcohols, and ethers. The three-tier test instrument includes reasoning and a confidence level for the selected answers and reasons. In this study, the tool used to detect misconceptions is the three-tier test, which combines multiple-choice questions with open reasoning and the confidence level of the given answer using the Certainty of Response Index (CRI) technique. The confidence levels used are on a scale of 0, 1, 2, 3, 4, and 5 (Hakim et al., 2012). The CRI technique allows for distinguishing between students who have a good understanding of concepts but lack confidence and those who experience misconceptions. A total of 42 three-tier test items were developed. The table of specifications for the three-tier test instrument is presented in Table 1.

Table 1. Specifications of the three-tier test instrument

Topic	Concept Mastery Indicator	**Question Numbers**	Cognitive Level
Alkenes and Alkynes	Understanding the naming of alkenes and alkynes	1, 2, 3, 4	C3, C3, C3, C3
	Understanding the physical and chemical properties of alkenes and alkynes	5, 6, 7, 8	C1, C2, C1, C2
	Understanding the concept of addition reactions in alkenes and alkynes	9, 10, 11, 12	C4, C5, C6, C6
Aromatic Compounds	Understanding the naming of benzene	13, 14	C3, C1

	Understanding the physical and chemical properties of benzene	15, 16	C1, C2
	Analyzing various electrophilic substitution reactions in benzene	17, 18, 19, 20, 21, 22	C5, C4, C3, C5, C6, C6
Alkyl Halides	Understanding the classification of alkyl halides	23, 24	C2, C1
	Understanding the naming of alkyl halides	25, 26	C3, C3
	Understanding the physical and chemical properties of alkyl halides	27, 28	C1, C2
	Analyzing the reaction mechanisms of SN1, SN2, E1, and E2 in alkyl halides	29, 30, 31, 32	C4, C5, C4, C6
Alcohols and Ethers	Understanding the naming of alcohols and ethers	33, 34	C3, C3
	Understanding the physical and chemical properties of alcohols and ethers	35, 36, 37, 38	C1, C2, C1, C2
	Identifying alcohols and ethers	39, 40	C6, C5
	Analyzing the reaction mechanisms of alcohols and ethers	41, 42	C4, C4

The research procedure can be seen in Figure 1.

Figure 1. Research Procedure

The procedure starts from preparing the questions and their specifications, followed by validation by 3 expert validators. Once valid, proceed with empirical testing (preliminary testing) to test the validity and reliability of the instrument empirically.

Techniques used to address concern (Richey, 2007):

1. Use experts with differing areas of specialization for tool review

This study involved three experts who evaluated content validity. This type of validity assesses the extent to which the items in the instrument represent the entire domain or aspects intended to be measured. The experts evaluated whether the instrument covered all relevant aspects of the construct or variable being measured. The approach used to evaluate content validity is the Content Validity Index (CVI**)**, as proposed by (Lynn, 1986). Lynn (1986) recommends using a minimum of three experts but notes that involving more than ten may not be necessary. The recommended measurement scale is a 4-point ordinal scale to avoid a neutral or ambiguous midpoint. Common labels include: 1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, and 4 = highly relevant. The Item-Content Validity Index (I-CVI) is then calculated for each item by counting the number of experts who rated the item as good (a score of 3 or 4) and converting the ordinal scale into a dichotomous scale, where relevant = 1 and not relevant = 0. This total is then divided by the number of experts. According to Lynn (1986), the I-CVI should be 1.00 when there are five or fewer raters. When there are six or more raters, the standard can be relaxed, but Lynn recommends the I-CVI not fall below 0.78 (Polit & Beck, 2006).

2. Have participants verify reports of tool use

The three-tier test instrument, which has been validated, underwent a limited trial to obtain empirical validity. The test was administered to 25 students who had studied the topics of alkenes, alkynes, aromatic compounds, alkyl halides, alcohols, and ethers. Empirical validity **r**efers to the validity obtained through direct testing with students or participants. This includes how test results correlate with expected performance or outcomes and how the test functions in real-world contexts. The assessment rubric is presented in Table 2, and an example of a three-tier question is shown in Table 3. The scores obtained from the rubric will be processed using **SPSS version 26**. The test aims to measure the validity and reliability of the three-tier test instrument.

Table 2. Assessment Rubric

Score	Criteria
0	Did not answer the question
1	Incorrect answer and reasoning
2	Correct answer, incorrect reasoning
2	Incorrect answer, correct reasoning
3	Correct answer, partially correct reasoning
4	Correct answer and reasoning

Table 3. Example of a Three-Tier Question

Learning Indicator	Question Indicator	Question	Cognitive Level
Understanding the concept of addition reactions in alkenes and alkynes	Differentiating addition reaction rules in alkenes and alkynes	Consider the following reactions: $(CH_3)_2C{=}CH_2 + HBr \rightarrow (CH_3)_2CH\text{-}CH_2Br$ $(CH_3)_2C{=}CH_2 + HBr \rightarrow (CH_3)_2CBr\text{-}CH_3$ $CH_3CH_2CH_2CH_2C{\equiv}CH + HBr \rightarrow CH_3CH_2CH_2CH_2CBr{=}CH_2$ $CH_3CH_2CH_2CH_2C{\equiv}CH + HBr \rightarrow CH_3CH_2CH_2CH_2CH{=}CHBr$ Which reactions demonstrate addition following Markovnikov's rule? 1 and 2 1 and 3 2 and 3 3 and 4 **Reason:** __________ **Confidence Level:** 0 1 2 3 4 5	C4

The limited trial was scored according to the criteria, and the results were processed using SPSS 26. Validity testing was conducted using the Pearson product-moment correlation, and reliability testing was conducted using Cronbach's Alpha. To determine whether each three-tier test item is valid, a comparison was made to the r-table with a significance level of $\alpha = 0.05$. If r_calculated ≥ r_table, the item is valid, and if r_calculated ≤ r_table, the item is not valid. The value of r_table is 0.3961 with a significance level of 5%. The reliability test criteria are presented in Table 4 (Guilford, 1956).

Table 4. Instrument Reliability Test Criteria

Correlation Coefficient	Criteria
$0.80 \leq r \leq 1.00$	Very High
$0.60 \leq r \leq 0.80$	High
$0.40 \leq r \leq 0.60$	Moderate
$0.20 \leq r \leq 0.40$	Low
$-1.00 \leq r \leq 0.20$	Very Low

RESULTS AND DISCUSSION

The results of the validation of the three-tier test instrument by three expert judgments are presented in Table 5. The instrument was given to the experts, and if any item was deemed incorrect, revisions were made until the items were considered accurate. As a result, values for relevant and highly relevant assessments were obtained.

Table 5. Calculation of I-CVI for Each Three-Tier Test Item

Item	Expert 1	Expert 2	Expert 3	Number of Agreements	I-CVI
1	1	1	1	3	3/3 = 1
2	1	1	1	3	3/3 = 1
3	1	0	1	2	2/3 = 0.67
4	1	1	1	3	3/3 = 1
5	1	1	1	3	3/3 = 1
6	1	1	1	3	3/3 = 1
7	1	1	1	3	3/3 = 1
8	1	1	1	3	3/3 = 1
9	1	1	1	3	3/3 = 1
10	1	1	1	3	3/3 = 1
11	1	1	1	3	3/3 = 1
12	1	1	1	3	3/3 = 1
13	1	1	1	3	3/3 = 1
14	1	1	1	3	3/3 = 1
15	1	1	1	3	3/3 = 1
16	1	1	1	3	3/3 = 1
17	1	1	1	3	3/3 = 1
18	1	1	1	3	3/3 = 1
19	1	1	1	3	3/3 = 1
20	1	1	1	3	3/3 = 1
21	1	1	1	3	3/3 = 1
22	1	1	1	3	3/3 = 1
23	1	1	1	3	3/3 = 1
24	1	1	1	3	3/3 = 1
25	1	1	1	3	3/3 = 1
26	1	1	1	3	3/3 = 1
27	1	1	1	3	3/3 = 1
28	1	1	1	3	3/3 = 1
29	1	1	1	3	3/3 = 1
30	1	1	1	3	3/3 = 1
31	1	0	1	2	2/3 = 0.67
32	1	1	1	3	3/3 = 1
33	1	1	1	3	3/3 = 1
34	1	1	1	3	3/3 = 1
35	1	1	1	3	3/3 = 1
36	1	1	1	3	3/3 = 1
37	1	1	1	3	3/3 = 1
38	1	1	1	3	3/3 = 1
39	1	1	1	3	3/3 = 1
40	1	1	1	3	3/3 = 1
41	1	1	1	3	3/3 = 1
42	1	1	1	3	3/3 = 1
∑	42	40	42	Mean I-CVI	0.98
Proportion Relevant	1	0.95	1		
Average Proportion Relevant		0.98			

Based on Table 5 it can be seen that the average I-CVI is 0.98, while the average proportion of items is 0.98. According to Lynn (1986), the I-CVI value must be 1.00 if there are five or fewer raters. If there are six or more raters, the standard can be more lenient, but Lynn (1986) recommends that the I-CVI should not fall below 0.78. Based on this, it can be said that the items in the three-tier test instrument for identifying misconceptions among students in organic chemistry are valid.

Polit & Beck (2006) in their study, concluded that it is important to clarify content validation in development research. Validity indicates how well a measuring instrument (test) can accurately measure what it is supposed to measure. Validity is not only related to the test itself but also to how the test is used for specific purposes. Therefore, validity reflects the researcher’s efforts to evaluate the usefulness and appropriateness of the test for a particular purpose, requiring various supporting evidence. This is necessary to maintain the test's use for the intended purpose, so that adequate evidence can be provided to support its use. Content validity is evaluated through testing the adequacy or relevance of the test content, usually conducted by a panel of experts through expert judgment. Content validity ensures that the measurement involves a sufficient and representative number of items to describe the intended concept. The more items on the scale that reflect the aspects or the entirety of the measured concept, the higher the content validity. In other words, content validity reflects how well the dimensions and elements of a concept are explained (Sekaran, 2006). Measuring and reporting the content validity of instruments is crucial because this type of validity also helps ensure construct validity and increases the confidence of readers and researchers in the instrument, as this validation involves experts evaluating the appropriateness of the instrument both conceptually and operationally (Hendryadi, 2017). In a valid test, students who can understand a question are also able to evaluate whether the answer is right or wrong (Cetin-Dindar & Geban, 2011; Erlianti et al., 2021; Zahro et al., 2021).

After the three-tier test instrument was declared valid in terms of content, a limited test was conducted on 25 students who had taken organic chemistry courses. The output of the Pearson product-moment validity test using SPSS 26 is presented in Table 6.

Table 6. Output of Pearson Product-Moment Validity Test

Question Number	Pearson Correlation	Description	Question Number	Pearson Correlation	Description
Question 1	0.596	V	Question 22	0.555	V
Question 2	0.183	TV	Question 23	0.596	V
Question 3	0.567	V	Question 24	0.715	V
Question 4	0.233	TV	Question 25	0.567	V
Question 5	0.651	V	Question 26	0.672	V
Question 6	0.715	V	Question 27	0.651	V
Question 7	0.672	V	Question 28	0.715	V
Question 8	0.608	V	Question 29	0.715	V
Question 9	0.501	V	Question 30	0.501	V
Question 10	0.651	V	Question 31	0.651	V
Question 11	0.696	V	Question 32	0.696	V
Question 12	0.608	V	Question 33	0.715	V
Question 13	0.485	V	Question 34	0.485	V
Question 14	0.555	V	Question 35	0.555	V
Question 15	0.608	V	Question 36	0.608	V
Question 16	0.651	V	Question 37	0.715	V
Question 17	0.546	V	Question 38	0.546	V
Question 18	0.672	V	Question 39	0.501	V
Question 19	0.696	V	Question 40	0.696	V
Question 20	0.596	V	Question 41	0.496	V
Question 21	0.596	V	Question 42	0.555	V

Based on Table 6, it is evident that the invalid questions are item numbers 2 and 4. These two question items are invalid because the calculated R value is smaller than the table R value. However, these two question items have positive values. This positive value indicates that the question can still be improved. Validity refers to the degree to which an instrument evaluates what it is supposed to evaluate. Validity measures whether a comprehensive evaluation system can accurately reflect the goals and requirements. It refers to the extent to which the measurement tool can measure the accuracy of the intended features. The higher the validity, the better the measurement results can demonstrate the characteristics intended to be measured (Mo et al., 2023). The output of the reliability test using Cronbach's Alpha in SPSS 26 is presented in Table 7.

Table 7. Output of Cronbach's Alpha Reliability Test

Reliability Statistics	
Cronbach's Alpha	**N of Items**
0.953	42

Based on Table 7, it is clear that the Cronbach's Alpha value is 0.953. This value indicates the reliability criterion of "very high." (Kimberlin & Winterstein, 2008) state that reliability estimation is used to evaluate (1) the stability of a measurement given at different times to the same individuals or using the same standard (test-retest reliability), or (2) the equivalence of sets of items from the same test (internal consistency), or the assessment by different observers of a behavior or event using the same instrument (inter-rater reliability). The reliability coefficient ranges from 0.00 to 1.00, with higher coefficients indicating higher reliability levels. The validity and reliability results obtained are consistent with the validity results of the three-tier test instrument developed by previous researchers (Damayanti et al., 2023).

CONCLUSION

Based on the results and discussion, it can be concluded that the developed three-tier test items were deemed valid according to expert judgment. The limited testing results showed that fourty questions were valid, while two questions were invalid. This indicates that, on average, the developed three-tier questions met the validity criteria. The two invalid questions cannot be used as research instruments. The reliability test results for all the questions indicated a reliability rating with very high criteria. These findings suggest that the questions, overall, have a very high level of reliability. Therefore, the three-tier test instrument can be used as a data collection tool to detect students' misconceptions in organic chemistry lectures.

ACKNOWLEDGEMENT

The author would like to express sincere gratitude to the Agency for the Assessment and Application of Technology (BPPT) for providing full support in the form of facilities and technical guidance throughout the research process. Special thanks are also extended to the Indonesia Endowment Fund for Education (LPDP) for the scholarship and financial support that made this research possible.

REFERENCES

Anam, R. S., Widodo, A., Sopandi, W., & Wu, H. K. (2019). Developing a five-tier diagnostic test to identify students' misconceptions in science: an example of the heat transfer concepts. *Elementary Education Online*, *18*(3), 1014–1029. https://doi.org/10.17051/ilkonline.2019.609690

Arellano, D. C. R., and T. M. H. (2014). Students' understanding of alkyl halide reactions in undergraduate organic chemistry. *Chemistry Education Research and Practice*, *15*(4), 501–515. https://doi.org/10.1039/c3rp00089c

Cetin-Dindar, A., & Geban, O. (2011). Development of a three-tier test to assess high school students' understanding of acids and bases. *Procedia - Social and Behavioral Sciences*, *15*, 600–604. https://doi.org/10.1016/j.sbspro.2011.03.147

Damayanti, Y. D., Priatmoko, S., Susilaningsih, E., & Nurhayati, S. (2023). Chemistry in Education Pengembangan Instrumen Tes Diagnostik Three-Tier Multiple Choice Test (TTMCT) untuk Menganalisis Miskonsepsi Siswa pada Konsep Sifat Keperiodikan Unsur. In *Chemined* (Vol. 12, Issue 2). http://journal.unnes.ac.id/sju/index.php/chemined

Duis, J. M. (2011). Organic chemistry educators' perspectives on fundamental concepts and misconceptions: An exploratory study. *Journal of Chemical Education*, *88*(3), 346–350. https://doi.org/10.1021/ed1007266.

Erlianti, S., Lisdiana, L., & Widiyaningrum, P. (2021). Development of Teaching Materials" The Negative Effects of Electric Cigarette" As a High School Student Learning Source. *Journal of Innovative Science Education*, *10*(1), 102-108.

Ewais, A., & Troyer, O. De. (2019). A Usability and Acceptance Evaluation of the Use of Augmented Reality for Learning Atoms and Molecules Reaction by Primary School Female Students in Palestine. *Journal of Educational Computing Research*, *57*(7), 1643–1670. https://doi.org/10.1177/0735633119855609

Fischer, C., Zhou, N., Rodriguez, F., Warschauer, M., & King, S. (2019). Improving College Student Success in Organic Chemistry: Impact of an Online Preparatory Course. *Journal of Chemical Education*, *96*(5), 857–864. https://doi.org/10.1021/acs.jchemed.8b01008

Guilford, J. P. (1956). *Mcgraw-HILL Pusllcatlons In Psychology Fundamental Statistics In Psychology And Education*.

Habiddin, & Page, E. M. (2019). Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indonesian Journal of Chemistry*, *19*(3), 720–736. https://doi.org/10.22146/ijc.39218

Hakim, A. , Liliasari. , and K. A. (2012). Student Concept Understanding of Natural Products Chemistry in Primary and Secondary Metabolites Using the Data Collecting Technique of Modified CRI. *International Online Journal of Educational Sciences*, *4*(3), 544–553. www.iojes.net

Hakim, A., & Kadarohman, A. (2012). student understanding of natural product concept of primary and secondary metabolites using CRI modified. *International Online Journal of Educational Sciences*, *4*(3), 544–553. www.iojes.net

Hasan, S., B. D., and K. E. L. (1999). Misconceptions and the Certainty of Response Index (CRI). *Physics Education*, *34*(5), 294. https://doi.org/10.1088/0031-9120/34/5/304

Hendryadi. (2017). Validitas isi: Tahap Awal Pengembangan Kuesioner. In *Jurnal Riset Manajemen dan Bisnis (JRMB) Fakultas Ekonomi UNIAT* (Vol. 2, Issue 2).

Jusniar, J., Effendy, E., Budiasih, E., & Sutrisno, S. (2020). Developing a three-tier diagnostic instrument on chemical equilibrium (TT-DICE). *Educacion Quimica*, *31*(3), 84–102. https://doi.org/10.22201/fq.18708404e.2020.3.72133

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. In *American Journal of Health-System Pharmacy* (Vol. 65, Issue 23, pp. 2276–2284). American Society of Health-Systems Pharmacy. https://doi.org/10.2146/ajhp070364

Lindsay, K., Carlsen-Landy, B., Boaz, C., & Marshall, D. (2017). Predictors of Student Success In Supplemental Instruction Courses at A Medium Sized Women's University. *International Journal of Research in Education and Science (IJRES)*, *3*(1), 208–217. www.ijres.net

Linenberger, K. J., Cole, R. S., & Sarkar, S. (2011). Looking beyond Lewis structures: A general chemistry molecular modeling experiment focusing on physical properties and geometry. *Journal of Chemical Education*, *88*(7), 962–965. https://doi.org/10.1021/ed100727r

Lynn, M. R. (1986). Determination and Quantification of Content Validity. *Nursing Research*, *35*(6). https://journals.lww.com/nursingresearcho nline/fulltext/1986/11000/determination_and_quantification_of_content.17.aspx

Milenković, D. D., Hrin, T. N., Segedinac, M. D., & Horvat, S. (2016). Development of a Three-Tier Test as a Valid Diagnostic Tool for Identification of Misconceptions Related to Carbohydrates. *Journal of Chemical Education*, *93*(9), 1514–1520. https://doi.org/10.1021/acs.jchemed.6b00261

Mo, Z., Li, X., Zhai, Y., Men, Y., Tang, Y., Qiao, J., Jia, X., Huang, Y., & Wang, B. (2023). Reliability and validity of a questionnaire measuring knowledge, attitude and practice regarding "oil, salt and sugar" among canteen staff. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-47804-3

Mutlu, A., & Sesen, B. A. (2015). Development of a Two-tier Diagnostic Test to Assess Undergraduates' Understanding of Some Chemistry Concepts. *Procedia - Social and Behavioral Sciences*, *174*, 629–635. https://doi.org/10.1016/j.sbspro.2015.01.593

Omwirhiren, E. M., Ubanwa, A. O., & Uk, U. C. (2016). An Analysis Of Misconceptions In Organic Chemistry Among Selected Senior Secondary School Students In Zaria Local Government Area Of Kaduna State, Nigeria. *International Journal of Education and Research*, *4*(7). www.ijern.com

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing and Health*, *29*(5), 489–497. https://doi.org/10.1002/nur.20147

Potgieter, M., Rogan, J. M., & Howie, S. (2005). Chemical concepts inventory of grade 12 learners and up foundation year students. *African Journal of Research in Mathematics, Science and Technology Education*, *9*(2), 121–134. https://doi.org/10.1080/10288457.2005.10740583

Richey, R. C. , & K. J. D. (2007). *Design and Development Research: Methods, Strategies, and Issues (1st ed.)*. Routledge. https://doi.org/https://doi.org/10.4324/9780203826034

Rizky Fajar, M., Kusasi, M., Brigjend Hasan Basry Banjarmasin, J. H., & Selatan Indonesia, K. (2023). Pengembangan Tes Diagnostik Five-Tier Untuk Mengidentifikasi Miskonsepsi Peserta Didik Dalam Materi Hidrokarbon Kelas XI Mipa Development of a Five-Tier Diagnostic Test To Identify Students Misconceptions In Class XI MIPA Hydrocarbon Materials. *Journal of Chemistry and Education)*, *6*(3), 129–128. http://jtam.ulm.ac.id/index.php/jcae

Rushton, G. T., Hardy, R. C., Gwaltney, K. P., & Lewis, S. E. (2008). Alternative conceptions of organic chemistry topics among fourth year chemistry students. *Chemistry Education Research and Practice*, *9*(2), 122–130. https://doi.org/10.1039/B806228P

Sekaran, U. (2006). *Metode Penelitian Bisnis*. Salemba Empat.

Siswaningsih, W., Firman, H., Zackiyah, & Khoirunnisa, A. (2017). Development of Two-Tier Diagnostic Test Pictorial-Based for Identifying High School Students Misconceptions on the Mole Concept. *Journal of Physics: Conference Series*, *812*(1). https://doi.org/10.1088/1742-6596/812/1/012117

Stone, K. L., Shaner, S. E., & Fendrick, C. M. (2018). Improving the success of first term general chemistry students at a liberal arts institution. *Education Sciences*, *8*(1). https://doi.org/10.3390/educsci8010005

Taagepera, M., and N. S. (2000). Mapping Students' Thinking Patterns in Learning Organic Chemistry by the Use of Knowledge Space Theory. *Journal of Chemical Education*, *11*(9), 1224–1229.

Zahro, U. R., Sumarni, W., & Linuwih, S. (2021). The Development of Test Instruments to Measure the Science Literation Skills of Junior High School Students in Global Warming Themes. Journal of Innovative Science Education, 10(1), 17-27.