ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV429

Creation of a Diagnostic Test Instrument with Four Tier Multiple Choice (FTMC) to Determine Students' Misconceptions About Chemical Bonding Material

Mita Laretasya[1]; Utiya Azizah[2*]
Department of Chemistry Education, Faculty of Mathematics and Science
Universitas Negeri Surabaya
Surabaya City, East Java 60231

Abstract:- As part of chemistry learning, students are faced with the need to master various chemical concepts that have been previously learned and be able to connect these concepts with the material they are learning to achieve a comprehensive understanding. Students who have difficulty in understanding concepts correctly will cause concept errors or misconceptions. Finding these misconceptions is the first step in helping students overcome them. A diagnostic test is one technique for identifying misconceptions. The purpose of this study is to develop a four-tier multiple-choice (FTMC) diagnostic test that may be used to determine students' misconceptions about chemical bonding materials in terms of validity, practicality and efficacy. Without the implementation stage, this kind of research is research and development (R&D) utilizing the ADDIE development model (Analyze, Design, Development, Implementation, Evaluation). The limited trial was conducted on 30 students of SHS 1 Dukun Gresik. Based on content validity and construct validity, the findings demonstrated that the four-tier multiple-choice (FTMC) diagnostic test instrument had a median score of 5, indicating that it was deemed valid. According to the students' questionnaire responses, the diagnostic test instrument's practicality was 86.67%, with a very practical category. Reliability and empirical validity tests of the test items are used to determine effectiveness. Six test items fell into the invalid category and fourteen test items fell into the valid category, according to the test item validity results. The test instrument was deemed reliable after the reliability test yielded a reliability rating of 0.780.

Keywords:- *Diagnostic Test Instruments, Four Tier Multiple Choice (FTMC), Misconceptions, Chemical Bonds.*

I. INTRODUCTION

Chemistry is part of a group of sciences that studies the properties, structure and composition of substances [1]. In the chemistry learning process, students must be able to master various concepts that have been acquired and new concepts that are being studied to achieve a comprehensive understanding. To embrace complex notions, one must first have a solid comprehension of the fundamentals. If students do not understand the concept correctly, they will have difficulty learning and have the potential to cause conceptual errors or misconceptions [2].

Misconceptions are wrong thoughts or ideas about an idea and are contrary to concepts believed to be true by experts. The causes of students' difficulty in understanding concepts can start from several factors, such as the reading books used, the environment, and the learning approach used by the teacher. Another factor that can result in concepts to be difficult for students to understand is the difference in interpretation between concepts that are understood and scientific concepts that are considered correct by experts. These differences in perception can hinder an accurate and comprehensive understanding of the chemical substance being studied. To overcome misconceptions among students is to identify these misconceptions [3]. A diagnostic test is one tool that can be used to find misconceptions while learning [4]. Diagnostic tests are used to identify learning problems faced by students, including errors in understanding concepts. A diagnostic exam can be conducted if it is discovered that most students have not been effective in engaging in learning in a specific subject of study [5]. Diagnostic exams are crucial in the independent curriculum since they help create learning plans that are specific to each student's needs.

Multiple-choice diagnostic test instruments have seen various advancements, including two-, three-, and four-tier multiple-choice versions. The four-tier multiple-choice diagnostic test is the most straightforward and precise of the different test kinds, classifying concepts as understood, concepts not understood, misconceptions, and errors. The four-tier multiple-choice diagnostic test is an evolution of the three-tier multiple-choice test. Four levels of questions serve as the foundation for creating the four-tier multiple-choice diagnostic test instrument. The first level consists of a multiple-choice question. The degree of student confidence in selecting a response is the second stage. At the third level, students respond to the question by providing five pre-selected answers and one open-ended response. The degree of confidence students have in selecting a cause is the fourth stage [2]. When compared to the three-tier multiple-choice diagnostic test, the four-tier multiple-choice test has the following advantages: (a) It can more thoroughly determine how well students comprehend concepts based on their confidence in their responses and the justifications they

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV429

provide, (b) Because students need to believe in their responses and the justifications for their content, it can more accurately identify misconceptions that they may have, (c) Identify the sections of the content that need extra attention while being learned, (d) Plan more effective instruction to assist students in minimizing misunderstandings. [6].

In all kinds of schools, chemical bonding is chemical material that is challenging for students to comprehend (National Standard Schools, potential/independent, and pioneering) because the majority of this material is abstract, difficult to experiment with and only a small part can be tried with concrete examples [7]. The results of this research are also related to findings obtained through observations and interviews with chemistry teachers at SHS Negeri 1 Dukun. Pre-research data shows that 85.29% of students scored learning outcomes below 75 or below the Minimum Completeness Criteria in chemical bonding material. Therefore, it is crucial to have a test tool that can determine whether pupils are in the category of comprehending the concept, not understanding it, or even having misconceptions. Apart from that, information was obtained that in chemistry learning teachers had never carried out tests to determine the misconceptions of the students. Students are given essays and one-tier multiple-choice questions.

Considering the results of the chemical bonding material learning procedure, it is possible that there is a lack of understanding or misconception that occurs in students, Therefore, in order to determine students' misconceptions regarding chemical bonding material, researchers are interested in creating a four-tier multiple-choice diagnostic test instrument. The purpose of this study is to develop a four-tier multiple-choice (FTMC) diagnostic test that may be used to determine students' misconceptions about chemical bonding materials in terms of validity, practicality, and efficacy.

II. LITERATURE REVIEW

➢ *Misconception*

Misconception is an error in using and connecting concepts to solve the right problem. This is because errors always occur because there is an initial concept that is understood by students and logistically constructs the misconception from experience [2]. Misconceptions can also be defined as students' understanding of concepts based on everyday experiences that are contrary to scientific concepts. Students are negatively impacted by misconceptions, which are perceived as barriers. Students that have misconceptions may find learning less effective and be less receptive to new information. If there are misconceptions in the cognitive structure of students, it is impossible for students to master further concepts. If misconceptions continue to develop, students will struggle to understand ideas at a higher level. [5].

Students' limited and inaccurate reasoning abilities, incorrect use of everyday terms, incorrect initial concepts, cognitive development stages that do not correspond with the concepts being studied, and students' interest in learning the concepts presented or taught are some of the factors that contribute to misconceptions in students. Thus, it is significant for educators to be mindful to students' initial conceptions before providing new information so that it can be easily accepted in the students' cognitive structure and misconceptions do not occur [8].

➢ *Four Tier Multiple Choice Diagnostic Test*

Diagnostic exams are designed to identify students' learning strengths and shortcomings so that the results can be utilized to inform follow-up instruction. Diagnostic examinations may consist of a series of inquiries or tasks [9]. One tool used to find misconceptions in a content is the four-tier multiple-choice diagnostic exam approach. The three-tier multiple-choice diagnostic exam approach is the basis for this approach. Students who correctly answered at levels 1 and 3 with confidence at levels 2 and 4 (concept understanding category) received a score of 1, while those who did not receive this combination of answers received a score of 0. This is the outcome of the four-tier multiple-choice diagnostic test assessment.

III. METHODS

The Research and Development (R&D) research approach was the one employed in this investigation. Due to the lack of extensive testing in this study, the ADDIE (Analyze, Design, Development, Implementation, Evaluation) research approach is used without going through the implementation stage. At the development stage, a limited trial was conducted at SHS 1 Dukun Gresik with 30 students who had received chemical bonding material. Three criteria are used to evaluate the produced system's viability: validity, practicality, and efficacy.

➢ *Validity*

Validators, who are professionals in their domains, derive validity from the evaluation's outcomes. Validity in this development is divided into 2, namely content validity and construct validity. In processing the validation sheet data using a Likert scale to obtain research data on the validation sheet filled in by three validators (2 chemistry lecturers and 1 chemistry teacher) with the following criteria.

Table 1 Likert Scale Criteria

Criteria	Score
Very valid	5
Valid	4
Quite valid	3
Invalid	2
Very invalid	1

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV429

Ordinal data that has been examined using the median value of data on each facet or indication makes up the validation result data. The aspect is deemed valid if the validator's assessment yields a median score of at least 4. A validator-assessed aspect is deemed invalid if its median score is less than 4. If there are aspects that do not meet the valid requirements, then improvements (revisions) must be made and re-validated until they reach the specified criteria [11].

➢ *Practicality*

Student answer questionnaires were used to collect practicality data, which was then quantitatively descriptively assessed utilizing the Guttman scale assessment [10].

Table 2 Guttman Scale

Answer	Criteria	Score
Positive	Yes	1
	No	0
Negative	Yes	0
	No	1

The four-tier multiple choice (FTMC) diagnostic test instrument's practicality % will be used to calculate the score using the formula below [12].

- *Percentage of Practicality:*

(∑ score obtained/∑ maximum score) x 100%

The percentage results are categorized into the categories listed in the following table.

Percentage (%)	Category
0-20	Very impractical
21-40	Impractical
41-60	Less practical
61-80	Practical
81-100	Very practical

According to these criteria, a percentage of ≥ 61% indicates that the designed four-tier multiple choice (FTMC) diagnostic test instrument is practical.

➢ *Effectiveness*

After conducting a trial of the diagnostic test, students' answers were analyzed to determine the empirical validity and reliability of the developed test instrument. Students' responses to a four-level multiple-choice diagnostic test are interpreted as follows [13].

Table 3 Interpretation of the Results of the Combination of Student Answers

No.	Category	*Tier* 1	*Tier* 2	*Tier* 3	*Tier* 4
1.	Misconception (M)	False	Sure	False	Sure
2.	Not Understanding Concept	False	Sure	False	Not Sure
3.		False	Not Sure	False	Sure
4.		False	Not Sure	False	Not Sure
5.	Understanding Concept	True	Sure	True	Sure
6.	Partially Understanding	True	Sure	True	Not Sure
7.		True	Not Sure	True	Sure
8.		True	Not Sure	True	Not Sure
9.		True	Sure	False	Sure
10.		True	Sure	False	Not Sure
11.		True	Not Sure	False	Sure
12.		True	Not Sure	False	Not Sure
13.		False	Sure	True	Sure
14.		False	Sure	True	Not Sure
15.		False	Not Sure	True	Not Sure
16.		False	Not Sure	True	Sure
17.	Unable to code/error	If one, two, three, or all of them are not filled in			

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV429

To calculate empirical validity with dichotomous item scores, the biserial correlation coefficient is used.

$$r_{pbis} = \frac{Mp - Mt}{St}\sqrt{\frac{p}{q}}$$

Description:

r pbis = point biserial correlation coefficient

Mp = Mean score of subjects who answered correctly the item being correlated

Mt = Mean total score

St = Standard deviation

p = proportion of subjects who answered correctly

q = 1-p

The diagnostic test instrument is valid if r count > r table [14].

The Cronbach's Alpha formula can be used to determine the test's reliability. Testing to measure how consistent (reliable) an instrument developed can use SPSS software. The diagnostic test instrument is valid if r count > r table [14].

IV. RESULTS AND DISCUSSION

➢ *Analyze Stage*

The analysis stage is carried out through a literature review that is relevant to the product to be developed. Based on the literature review that has been conducted, it is known that one of the chemical materials that has the potential to cause misconceptions is the chemical bond material. In addition, field observations were conducted at the school. Field observations were conducted by interviewing chemistry teachers at SHS 1 Dukun Gresik. Observations were also conducted to determine the condition of students, teachers, and the school environment. Considering the results of the conducted interviews, it is known that teachers have never given diagnostic tests to students and the questions given are only one tier multiple choice and essays. In addition, data was obtained that 85.29% of students scored below the Minimum Criteria Completion on the chemical bond material. The use of diagnostic test instruments that have never been carried out and are known to most students still get low scores on the chemical bond material so that efforts are needed to identify students' understanding whether they experience misconceptions or not. Multiple-tier multiple-choice diagnostic test instruments have undergone many developments, including two-, three-, and four-tier versions. The four-tier multiple-choice diagnostic test is the most straightforward and precise of the different test kinds, classifying concepts that are understood, concepts that are not understood, misconceptions, and errors [15].

➢ *Design Stage*

The design stage is the stage of designing a diagnostic test product to identify student misconceptions. The diagnostic test instrument created for this investigation was a four-tier multiple-choice. The first level consists of a multiple-choice question. The degree of student confidence in selecting a response is the second stage. The third level involves pupils responding to questions with one open-ended response and five pre-provided explanations. The degree of confidence students have in selecting a cause is the fourth stage. The following are the stages of designing a four-tier multiple-choice diagnostic test instrument.

- *Determination of Material*

In order to identify the content that could lead to a lot of misunderstandings, the literature was reviewed and field observations were made at the school. The chemical bond material is the one that could lead to misunderstandings, according to the literature analysis and field observations at the school. The material used in this study is the chemical bond material in the ionic bond and covalent bond sub-materials.

- *Determining the Sequence of Material or Concepts*

Concept maps used as a reference in creating a sequence of questions in diagnostic tests.

- *Designing the Diagnostic Test Question Grid*

Includes title, educational unit, subject, class, question form, topic, learning achievement, question number, question indicator, question items, answers, and cognitive domain.

- *Designing Diagnostic Test Items*

There are four tiers in the diagnostic test's four-tier multiple-choice format. The design of questions at the first level must be adjusted to the question indicators and the design of reasons at the third level must be adjusted to the answer choices at the first level. The diagnostic test instrument consists of 7 ionic bond questions and 13 covalent bond questions. The number of questions on the covalent bond sub-topic is greater than on ionic bonds, considering that covalent bonds consist of more sub-concepts.

At the end of the design stage, an evaluation was carried out by the reviewer, namely the supervising lecturer, to obtain suggestions for improving the initial draft of the diagnostic test instrument.

➢ *Development Stage*

The design of the diagnostic test instrument that had been created and reviewed in the previous stage was then validated by 3 validators and a limited trial was conducted.

- *Validity*

Two criteria are used to evaluate the validity of diagnostic test instruments: construct validity and content validity. Content validity is a validity criterion that assesses the extent to which parts of the instrument represent the components of the overall content of the object to be measured. Content validity is viewed from 3 assessment

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV429

indicators, namely: (a) suitability of the question items with the concept of ionic bonds and covalent bonds, (b) suitability between each question item and the question indicators, and (c) suitability between each question and the expected answers and reasons. Construct validity is validity that measures the extent to which parts of the instrument represent the components of the overall content of the object to be measured. Construct validity is viewed from 3 assessment indicators, namely: (a) test question criteria in accordance with the cognitive domain (Bloom's taxonomy), (b) suitability of tables, graphs, pictures and the like with the questions presented, and (c) questions and statements used does not create a double meaning. The ordinal data used for validation is examined using the median data value for each component or indicator. The median value for each facet or indication was used to assess the collected data. Once the data has been sorted by value, the median value is the midway value. Once the data has been sorted by value, the median value—if the amount of data is even—is calculated by averaging the two middle data. [16]. The four-level multiple-choice diagnostic test instrument's validity results are as follows:

Table 4 Validation Results by Validator

Question item	Median score	
	Content Validity	Construct Validity
1	5	5
2	5	5
3	5	5
4	5	5
5	5	5
6	5	5
7	5	5
8	5	5
9	5	5
10	4	5
11	5	5
12	5	5
13	5	5
14	5	5
15	5	5
16	5	5
17	5	5
18	5	5
19	5	5
20	5	5

According to Table 4, one item had a median of 4 with a valid category for the content and construct validity aspects, while 19 items received a median of 5 with a highly valid category. The diagnostic test instrument examined from the construct validity and content validity aspects is deemed valid because it has a median score of at least 4, according to the validation data analysis.

- *Practicality*

Thirty students completed the answer questionnaire following a four-level multiple-choice diagnostic exam. In the response questionnaire there were 10 statements with details of 8 being positive statements and 2 negative statements. The student response questionnaire has ten statements that need to be completed. The following table displays the findings from the examination of the student response survey.

Table 5 Student Response Questionnaire Results

Statement	Response Score	Practicality (%)
1	29	96,7
2	29	96,7
3	23	76,7
4	22	73,7
5	26	86,7
6	23	76,7
7	24	80
8	30	100
9	28	93,3
10	26	86,7
Average (%)		86,67

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV429

According to Table 5, three assertions had a practical category percentage of ≥ 61%, while seven statements received a very practical category percentage of ≥ 81%. The four-tier multiple-choice diagnostic test instrument's total practicality was 86.67%, with a very practical category. These findings show that students are engaged and that the four-tier multiple-choice diagnostic exam instrument was developed well.

- *Effectivity*

Student results are used to assess the test's empirical validity and reliability following a brief trial. Empirical validity is a validity test conducted after students have completed the test questions being tested for validity. A score of 1 is awarded to each student who answered correctly at levels 1 and 3 and with high confidence at levels 2 and 4 (conceptual knowledge category); a score of 0 is awarded for any other combination of responses.

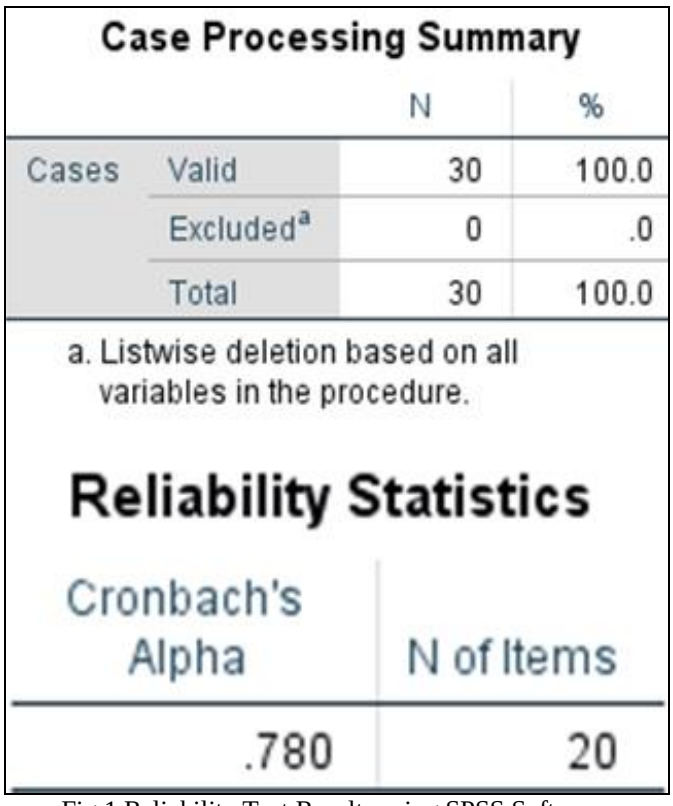
Table 6 Empirical Validity Results

Question	r_{count}	r_{table}	Category
1	0,688	0,361	Valid
2	0,850		Valid
3	0,528		Valid
4	0,148		Invalid
5	0,594		Valid
6	0,495		Valid
7	0,499		Valid
8	0,525		Valid
9	0,436		Valid
10	0,383		Valid
11	0,276		Invalid
12	0,497		Valid
13	0,430		Valid
14	0,215		Invalid
15	0,233		Invalid
16	0,562		Valid
17	0,576		Valid
18	0,310		Invalid
19	0,009		Invalid
20	0,629		Valid

The examination of the small-scale test group's validity test revealed that there were six invalid items and fourteen valid four-tier multiple-choice diagnostic questions.

Whether an evaluation test is valid or not can be influenced by students' answers rather than the interpretation of the items on the test. Inconsistent responses from students who don't grasp the subject matter and the questions' meaning diminish the usefulness of the questions as a means of gauging students' aptitude [17]. Another step towards invalid question items is not to use them or throw them away [18]. However, question items that are invalid in the empirical validity test can still be used based on consideration of the results of the validity test by those declared valid which obtained a median score of 5 with a very valid category.

To ascertain the degree of accuracy of the diagnostic test questions being created, reliability testing was done. The internal consistency approach was used to acquire reliability data; it was tested only once before being analyzed using the Alpha Cronbach coefficient equation with SPSS version 25 assistance. A score of 1 is awarded to each student who answered correctly at levels 1 and 3 and with high confidence at levels 2 and 4 (conceptual knowledge category); a score of 0 is awarded for any other combination of responses.



Case Processing Summary

		N	%
Cases	Valid	30	100.0
	Excluded[a]	0	.0
	Total	30	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.780	20

Fig 1 Reliability Test Results using SPSS Software

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV429

The created diagnostic tool has a Cronbach Alpha value of 0.780 based on the reliability test findings. For N = 30, the r table value is 0.361. If r count > r table, the instrument is considered to have strong reliability, and the produced diagnostic test instrument can be deemed dependable.

V. CONCLUSION

The construction of a four-level multiple-choice diagnostic test instrument to identify students' misconceptions about chemical bonding material is practical for use, according to the analysis and discussion that have been conducted. The following factors are examined while evaluating the viability of diagnostic testing:

- With a median score of 5 (extremely valid), the four-level multiple-choice diagnostic test (FTMC) is considered to be valid in terms of both construct and content validity.
- According to the student response questionnaire, the four-tier multiple-choice (FTMC) diagnostic exam instrument has a practicality percentage of 86.67% (extremely practical).
- Based on an evaluation of the question items' empirical validity, test reliability, and the detection of student misconceptions, the four tier multiple choice (FTMC) diagnostic test instrument is said to be effective. Six questions (30%) fall into the invalid group, whereas 14 questions (70%) pass the empirical validity test. With a r table value of 0.361 and a Cronbach alpha value of 0.780, the reliability test was deemed reliable.

REFERENCES

[1]. Azzahra, F., Milama, B., & Wardani, M. (2023). Identifikasi Kesulitan Guru Menggunakan Laboratorium Virtual Dalam Pembelajaran Kimia. *Dalton : Jurnal Pendidikan Kimia dan Ilmu Kimia, 6*(1), 31-39.

[2]. Harahap, I. P., & Novita, D. (2020). Validitas dan Reliabilitas Instrumen Tes Diagnostik Four-Tier Multiple Choice (4TMC) pada Konsep Laju Reaksi. *Unesa: Journal of Chemical Education, 9*(2), 222-227.

[3]. Haq, I., & Sukarmin. (2023). Pengembangan Instrumen Tes Diagnostik Three Tier Berbasis Web untuk Mengidentifikasi Miskonsepsi Siswa Kelas X pada Materi Reaksi Redoks. *Unesa: Journal of Chemical Education*, *12*(3), 189-199.

[4]. Law, J., & Treagust, D. (2010). *Diagnosis of Student Understanding of Content Specific Science Areas Using On line Two Tier Diagnostic Test.* Australia: Curtin University of Technology.

[5]. Noviani, W., & Istiyadji, M. (2017). Miskonsepsi Ditinjau dari Penguasaan Pengetahuan Prasyarat untuk Materi Ikatan Kimia Pada Kelas X. *Quantum: Jurnal Inovasi Pendidikan Sains, 8*(1), 63-77.

[6]. Fariyani, Q., Rusilowati, A., & Sugianto. (2015). Pengembangan Four Tier Diagnostic Test untuk Mengungkap Miskonsepsi Fisika Siswa SMA Kelas X. *Journal of Innovative Science Education, 4*(2), 41-49.

[7]. Sunyono, Wirya, I., & Sujadi, G. (2009). Identifikasi Masalah Kesulitan dalam Pembelajaran Kimia SMA Kelas X di Propinsi Lampung. *Jurnal Pendidikan MIPA, 10*(2), 9-18.

[8]. Mukhlisa, N. (2021). Miskonsepsi pada Peserta Didik. *SPEED: Journal of Special Education, 4*(2), 66-76.

[9]. Rusilowati, A. (2015). Pengembangan Tes Diagnostik Sebagai Alat Evaluasi Kesulitan Belajar Fisika. *Prosiding Seminar Nasional Fisika dan Pendidikan Fisika, 06*(1), 1-10.

[10]. Sugiyono. (2021). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D.* Bandung: Alfabeta.

[11]. Lutfi, A. (2021). *Research and Development (R&D): Implikasi dalam Pendidikan Kimia.* Surabaya: Universitas Negeri Surabaya.

[12]. Riduwan. (2016). *Dasar-dasar Statistika.* Bandung: Penerbit Alfabeta.

[13]. Amin, N., Wiendartun, & Samsudin, A. (2016). Analisis Intrumen Tes Diagnostik Dynamic-Fluid Conceptual Change Inventory (DFCCI) Bentuk Four-Tier Test pada Beberapa SMA di Bandung Raya. *Prosiding SNIPS, 1*(1), 570-574.

[14]. Arikunto. (2005). *Dasar-dasar Evaluasi Pendidikan.* Jakarta: Bumi Aksara.

[15]. Amelia, T., Elvia, R., & Handayani, D. (2022). Identifikasi Miskonsepsi Siswa Pada Pembelajaran Kimia Menggunakan Metode Four Tier Diagnostik Test Di SMA Negeri 03 Kota Bengkulu. *Alotrop: Jurnal Pendidikan dan Ilmu Kimia, 6*(2), 110-117.

[16]. Sudjana. (2005). *Metoda Statistika.* Bandung: PT. Tarsito Bandung.

[17]. Sukardi. (2010). *Evaluasi Pendidikan: Prinsip dan Operasionalnya.* Jakarta: PT. Bumi Aksara.

[18]. Slamet, R., & Wahyuningsih, S. (2022). Validitas dan Reliabilitas Terhadap Instrumen Kepuasan Kerja. *Aliansi: Jurnal Manajemen & Bisnis, 17*(2), 51-57.