



Development of Four Tier Multiple Choice Diagnostic Test Instrument to Detect Misconception on Chemical Bonding

Hilmalia Febriana Sari¹, Rudiana Agustini²

^{1,2}Universitas Negeri Surabaya, Indonesia

E-mail: hilmaliasari@gmail.com

Article Info	Abstract
Article History Received: 2024-07-22 Revised: 2024-08-19 Published: 2024-09-04 Keywords: <i>Miskonsepsi;</i> <i>Tes Diagnostik;</i> <i>Four Tier Multiple Choice;</i> <i>Validitas;</i> <i>Reliabilitas.</i>	Misconception is a concept that is not in accordance with scientific understanding which contradicts the concept that has been believed to be true by experts. Misconceptions in students will occur continuously if not given the right treatment. The treatment given to students can be done well if misconception detection is carried out on students. Misconceptions that are not detected by the teacher can hinder students in receiving further learning. Diagnostic tests can be used to detect misconceptions in students. This study aims to determine the feasibility of a four-tier multiple choice diagnostic instrument to detect students' misconceptions on chemical bonding material based on validity and reliability aspects. The validity of the four-tier multiple choice diagnostic instrument was assessed by three validators based on content, construct, and language criteria using a 1-5 Likert scale calculation. The reliability of the four-tier multiple choice diagnostic instrument is calculated using the Cronbach Alpha equation. Based on the results of the study, it can be concluded that (1) Content validity, construct validity, and language validity have consecutive mode values of 5, 5, and 4 with an overall mode value of 5 so that it is included in the very valid criteria. (2) The instrument developed is declared reliable with the value of the results of $r_{count} > r_{table}$ with details of the reliability coefficient on the instrument (r_{count}) of 0.381 and the r_{table} value used is 0.373. The diagnostic test instrument that has been developed is declared suitable for use.
Artikel Info	Abstrak
Sejarah Artikel Diterima: 2024-07-22 Direvisi: 2024-08-19 Dipublikasi: 2024-09-04 Kata kunci: <i>Miskonsepsi;</i> <i>Tes Diagnostik;</i> <i>Four Tier Multiple Choice;</i> <i>Validitas;</i> <i>Reliabilitas.</i>	Miskonsepsi merupakan konsep yang tidak sesuai dengan pengertian ilmiah yang bertolak belakang dengan konsep yang telah diyakini kebenarannya oleh para ahli. Miskonsepsi pada peserta didik akan terjadi secara terus menerus apabila tidak diberikan perlakuan dengan tepat. Perlakuan yang diberikan kepada peserta didik dapat dilakukan dengan baik apabila dilakukan deteksi miskonsepsi terhadap peserta didik. Miskonsepsi yang tidak dideteksi oleh guru dapat menghambat peserta didik dalam menerima pembelajaran selanjutnya. Tes diagnostik dapat digunakan untuk mendeteksi miskonsepsi pada peserta didik. Penelitian ini bertujuan untuk menentukan kelayakan instrumen diagnostik <i>four tier multiple choice</i> untuk mendeteksi miskonsepsi peserta didik pada materi ikatan kimia berdasarkan aspek validitas dan reliabilitas. Validitas instrumen diagnostik <i>four tier multiple choice</i> dinilai oleh tiga validator berdasarkan kriteria isi, konstruk, dan bahasa menggunakan perhitungan skala Likert 1-5. Reliabilitas instrumen diagnostik <i>four tier multiple choice</i> dihitung menggunakan persamaan <i>Alpha Cronbach</i> . Berdasarkan hasil penelitian dapat disimpulkan bahwa (1) Validitas isi, validitas konstruk, dan validitas bahasa memiliki nilai modus berturut-turut 5, 5, dan 4 dengan nilai modus keseluruhan adalah 5 sehingga termasuk kedalam kriteria sangat valid. (2) Instrumen yang dikembangkan dinyatakan reliabel dengan nilai hasil $r_{hitung} > r_{tabel}$ dengan rincian koefisien reliabilitas pada instrumen (r_{hitung}) sebesar 0,381 dan nilai r_{tabel} yang digunakan 0,373. Instrumen tes diagnostik yang telah dikembangkan dinyatakan layak untuk digunakan.
I. INTRODUCTION Government Regulation No. 4 of 2022 states that the applicable curriculum can be used in line with changes in national education standards which are a reference in curriculum development. The national standards include content standards, management standards, graduate competency standards, educator and	education personnel standards, process standards, facilities and infrastructure standards, financing standards, and assessment standards. To achieve these national standards, it is necessary to have the role of students, one of which is the aspect of knowledge that students have. Permendikbud Number 5 of 2022 concerning Graduate Competency Standards

states that, one of the knowledge dimensions that must be possessed by SMA / MA / SMALB / Package C students is being able to use mathematical concepts, procedures, facts, and tools to solve practical problems relevant to their vocational field.

The Merdeka Curriculum is the curriculum currently used in Indonesia. The Merdeka Curriculum provides flexibility to teachers in planning and conducting the learning process, by first conducting diagnostic assessments of students (Aziz, 2023). Diagnostic assessment is one of the assessments in the Merdeka Curriculum. Diagnostic assessment or often called assessment for learning can be used as a benchmark for teachers in designing learning according to the characteristics of students. Diagnostic assessment can be used to detect students' strengths and weaknesses in the learning process, especially in materials that have many concepts. One of the branches of science related to conceptual knowledge is chemistry. Chemistry is the study of specific events that occur in a substance and everything related to the structure and properties, composition substances, dynamics and energy of substances and transformations. Chemistry studies the knowledge of facts, rules, principles, concepts, theories, descriptions, chemical intelligence, and also the discovery process (Sa'adah, 2022). Chemistry requires an understanding of the concept in learning it because there is a connection between each concept. Therefore, students must get the concept of chemistry as a foundation to be able to carry out further applicative learning.

One of the subjects in chemistry with many concepts that are interconnected with each other is chemical bonding material. Chemical bonds consist of metal bonds, covalent bonds, and ion bonds. All of which involve a number of concepts including the concept of opposite charge attraction, similar charge repulsion, molecules, atoms, protons, electrons, and neutrons (Tsaparlis, Pappa, and Byers, 2018). Therefore, in learning chemical bonding material, students are required to understand each concept so that it does not interfere with further learning.

The relationship between concepts in learning a material can lead to various forms of understanding in understanding a concept that can be wrong. Learners can develop their own understanding with the assimilation process, which is applying existing concepts to solve a new problem by making small changes as a form of adaptation (Yulianova, 2022). Learners often

experience difficulties and even failures in the assimilation process because they do not accommodate the knowledge to be learned so that there are changes in conception and potentially misconceptions (Yulianova, 2022). Misconceptions found in students can be grouped as; 1) Prejudice, derived from individual experience; 2) Non-scientific beliefs, learners' perceptions that are not based on scientific sources; 3) Conceptual misconceptions, occur when learners apply knowledge that is contrary to scientific concepts; 4) Vernacular misconceptions, word choice errors by learners that give meanings opposite to scientific concepts; 5) Factual misconceptions, misconceptions that arise by learners since childhood and remain until adulthood (Soeharto, 2019).

Misconceptions in students will occur continuously if not given treatment. Misconceptions in students are resistant and presistent. Misconceptions that cannot be detected by the teacher will hinder students in receiving new learning and result in delays in mastering the next material (Kusumawati, et al 2022). Therefore, there is a need for early detection of misconceptions encountered in students so as not to make it difficult for students to understand the concepts to be learned. Misconceptions can be detected using mind maps, interviews, portfolios, and multiple choice tests (Suparno, 2013). Misconceptions can be identified using clinical interviews, two-tier diagnostic tests, pictures, mind maps, and Certainty of Response Index (CRI) (Mukhlisa, 2021). Diagnostic tests with multiple choice forms can measure concept understanding in many students in a shorter time than interviews (Widyatmoko and Shimizu, 2018).

There are two main objectives in diagnostic tests, namely: 1) detect participants who have constraints, 2) seek solutions that are in accordance with the constraints that have been identified (Ebiati, 2021). There are various kinds of diagnostic tests, one of which is a multiple choice test. The development of multiple choice tests can help science teachers detect misconceptions in students (Treagust, 1986 and Gurel, 2015). There are various kinds of multiple-choice diagnostic tests used to detect students' misconceptions, namely: 1) One-level multiple-choice diagnostic tests; 2) Two-level multiple-choice diagnostic test; 3) Three-level multiple-choice diagnostic test; 4) Four-level multiple-choice diagnostic test. A single-level multiple-choice diagnostic test is only used to obtain answers. Two-level multiple-choice diagnostic

tests are used to obtain answers and answer confidence. Three-level multiple-choice diagnostic tests are used to obtain answers, answer confidence, and reasoning. Four-level multiple choice diagnostic tests are used to obtain answers and reasons as well as answer confidence and reasons. The weakness of Two-tier multiple choice is that it is unable to categorize errors caused by misconceptions. The weakness of three-tier multiple choice is that it cannot reveal beliefs at the third level where the choice of beliefs in the first level and second level answers can be different (Gurel, 2015).

The weaknesses of the two-tier and three-tier multiple choice tests are then addressed with the four-tier multiple choice test (FTMC). FTMC is a multiple choice diagnostic test with four tiers of questions. The first tier contains questions with various answer options, the second tier contains the level of confidence of students in answering questions in the first tier, the third tier contains reasons that refer to the answers contained in the first tier, the fourth tier contains the level of confidence of students in answering the reasons in the third tier (Agustina, 2022). Four-tier diagnostic tests have advantages over other diagnostic tests, namely: 1) Allows comparison between the level of confidence in the answers and reasons, so as to know the level of understanding of the concept of learners with more depth; 2) Detect misconceptions experienced by students better; 3) Identify material that requires more focus; 4) Design lessons that can help reduce students' misconceptions better (Fariyani, 2015).

II. METHOD

This study aims to determine the feasibility of four tier multiple choice diagnostic test instruments based on aspects of validity and reliability. This type of research is Research and Development which adapts the 4D development model by Thiagarajan. The stages in this study were carried out up to the 3D stage including define, design, and develop. Activities carried out at the define stage include front end analysis and literature review. At the design stage, writing a grid of test questions, writing test questions, and preliminary designs were carried out. At the develop stage, the test questions were reviewed, the test questions were validated, the questions were revised, and the test questions were limitedly tested. The research subjects used during the limited trial were 26 students of class XI BIKIFIMAE 2 SMAN LABSCHOOL UNESA 1

Surabaya which is a heterogeneous class and has received chemical bonding material.

The diagnostic test instrument that has been prepared is then analyzed for its validity value. The validity data of the diagnostic test instrument was obtained from three expert validators consisting of two lecturers of Chemistry Education Unesa and a teacher of SMAN LABSCHOOL UNESA 1 Surabaya. The validation results were then analyzed quantitatively using Likert scale calculations which can be observed in the following table.

Table 1. Likert Scale

Assessment	Scale Value
Very valid	5
Valid	4
Moderate	3
Bad	2
Very bad	1

Sumber: Riduwan, 2015

Likert scale scores were then analyzed using the mode technique. The mode is a data analysis technique based on the value that appears most often, the value is the value given by the validator. The reliability of the instrument is obtained through the results of students' answers at the limited trial stage. Each learner who answers correctly at levels one and three and is sure at levels two and four (concept understanding category) will be given a score of 1. While other than the combination of answers will be given a score of 0. The reliability of the test will be calculated using the Cornbach Alpha equation which in (Siregar, 2013) is as follows.

$$r = \frac{k}{k-1} \left(1 - \frac{\sum Si^2}{St^2} \right)$$

$$Si^2 = \frac{\sum Xi^2 - \frac{(\sum Xi)^2}{n}}{n}$$

$$St^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n}$$

The instrument is said to be reliable if it has good reliability with $r_{\text{count}} > r_{\text{table}}$ (Arikunto, 2005).

III. RESULTS AND DISCUSSION

A. Results

1. Test Validity Data

Table 2. Calculation Of The Results Of The Validity Assessment Of The Four Tier Multiple Choice Diagnostic Test Instrument

No	Aspect Validated	Mode	Validity Criteria
Content Criteris			
1.	The accuracy of each test item with chemical bonding material	5	Very Valid
2.	The accuracy of each test item with the question indicator.	4	Valid
3.	The accuracy of each test item with the material sequence.	5	Very Valid
4.	The accuracy of question boundaries, answers, and explanation of reasons.	5	Very Valid
	Content Criteria Mode	5	Very Valid
Construct Criteria			
1	Clarity of instructions for using the diagnostic test instrument	5	Very Valid
2	Accuracy of the use of indicators on each item	4	Valid
3	Each item is able to detect misconceptions in students	4	Valid
4	The reason options given are able to state the cause of misconceptions originating from students	5	Very Valid
5	The excuse options are homogeneous and objective with the first tier answer options	5	Very Valid
6	Accuracy of pictures, graphs, tables and the like with the problem presented	5	Very Valid
	Construct Criteria Mode	5	Very Valid
Language Criteria			
1	Question sentences use Indonesian that is in accordance with applicable rules and regulations	4	Valid
2	Various sentences or questions do not cause multiple interpretations	4	Valid
3	Each test question is clearly stated and easy to understand	5	Very Valid
	Language Criteria Mode	4	Valid
	Totally Mode	5	Very Valid

2. Test Reliability Data

Table 3. Test Reliability Calculation Results Using SPSS

Item-Total Statistics				
Test Number	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	2.62	3.126	.227	.330
2	2.73	3.245	.260	.331
3	2.69	3.502	.013	.395
4	2.58	3.374	.046	.391
5	2.65	3.515	-.014	.405
6	2.62	3.366	.066	.383
7	2.54	3.618	-.103	.442
9	2.69	3.742	-.156	.438
8	2.69	3.262	.194	.344
10	2.65	3.195	.208	.338
11	2.73	3.085	.406	.292
12	2.65	3.195	.208	.338
13	2.62	3.206	.172	.348
14	2.69	3.182	.258	.326
15	2.69	3.262	.194	.344

B. Discussion

1. Test Validity

Theoretical validity data is obtained from the results of filling out the validation sheet using a rating scale calculation with a score of one to five. The rating scale contains various aspects that are observed and then described into a specific criteria scale. The aspects assessed consist of content, construct, and language validity. Furthermore, the validator provides a general assessment of the instrument that has been developed and gives a decision whether the instrument is ready to use, needs revision, or has not met the requirements for use. Based on the table, it can be seen that the results of the validity assessment on the content criteria get a mode value of 5, the validity on the construct criteria gets a mode of 4, and the validity on the language criteria gets a mode value of 4. So, overall the four tier multiple choice diagnostic instrument has a mode value of 5. This value is interpreted according to the guidelines in Table 1 by producing very valid criteria.

Based on the data generated, it was found that the four tier multiple choice diagnostic instrument on chemical bonding material that had been developed and tested on a limited basis was included in the very valid criteria.

2. Test Reliability

The reliability of the test shows how consistent the test scores obtained from the measurement results at different times (Kusaeri & Suprananto, 2012). This study aims to determine the reliability of the test instrument can be seen from the combination of students who answer correctly on tier one and tier three and are sure on tier two and tier four (concept understanding category), it will be given a score of 1, while the answer does not meet the combination, will get a score of 0. Reliability data obtained from the diagnostic test results of students. Test reliability is calculated using Cornbach's Alpha consistency which is analyzed using the IBM SPSS Statistic 26 application. Based on table ..., the instrument reliability coefficient value (r_{count}) is 0.381. The r_{table} value used for $n = 26$ with a significance level of 5% is 0.373. Thus, the four tier multiple choice diagnostic test instrument on chemical bonding material that has been developed can be said to be reliable because $r_{count} > r_{table}$ (Arikunto, 2005).

Overall, the four tier multiple choice diagnostic test instrument developed to detect students' misconceptions on chemical bonding material has met the requirements of theoretical validity, empirical validity, and reliability so that it is suitable for use. This is in accordance with the statement of Kimberlin & Winterstein (2008) because it has fulfilled the validity and reliability aspects which are parameters to represent the quality of the instrument as a whole.

IV. CONCLUSION AND SUGGESTION

A. Conclusion

The four tier multiple choice diagnostic instrument on chemical bonding material developed can be declared suitable for use with the following details.

1. Content validity, construct validity, and language validity have consecutive mode values of 5, 5, and 4 with an overall mode value of 5 so that it is included in the very valid criteria.
2. The instrument developed is declared reliable with the value of the results of $r_{count} > r_{table}$ with details of the reliability coefficient on the instrument (r_{count}) of

0.381 and the value of the r_{table} used is 0.373.

B. Suggestion

The four tier multiple choice diagnostic instrument that has been developed can be further researched using IT developments to make it easier to use by teachers and students.

REFERENCES

- Agustina, L. and Indana, S., (2022). Profil Miskonsepsi Siswa pada Materi Protista Menggunakan Four Tier Test. *Berkala Ilmiah Pendidikan Biologi (BioEdu)*, 11(1), pp.60-67.
<https://ejournal.unesa.ac.id/index.php/bioedu/article/download/39945/36324>
- Aziz, A.C.K., (2023). Asesmen Diagnostik Sebagai Penilaian Pembelajaran dalam Kurikulum Merdeka di Sekolah Dasar. *Jurnal Pendidikan Sekolah Dasar*, 1(2), 20-29.
<https://doi.org/10.33830/penaanda.v1i2.6202>
- Ebiati, E., (2021). Pengembangan Instrumen Tes Diagnostik Dengan Certainty of Response Index Untuk Mengidentifikasi Miskonsepsi Siswa Pada Materi Ikatan Kimia. *Journal Evaluation in Education (JEE)*, 2(3), pp.112-116.
<https://doi.org/10.37251/jee.v2i3.220>
- Fariyani, Q., Rusilowati, A., & Sugianto (2015). Pengembangan Four-Tier Diagnostic Test untuk Mengungkap Miskonsepsi Fisika Peserta Didik SMA Kelas X. *Journal of Innovative Science Education*, 4(2), 41-49.
- Gurel, DK, Eryilmaz, A, & McDermott, LC (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science., open.metu.edu.tr.
- Kusumawati, AE, Anggraini, W, & ... (2022). Analysis of Prospective Physics Teacher's Misconceptions on Interference Material using Certainty of Response Index (CRI). *Jurnal Pendidikan Fisika ...*, jurnalkip.unram.ac.id,
<http://jurnalkip.unram.ac.id/index.php/IPFT/article/view/3679>
- Mukhlisa, N., (2021). Miskonsepsi pada peserta didik. *SPEED Journal: Journal of Special Education*, 4(2), pp.66-76.
<https://doi.org/10.31537/speed.v4i2.403>

- Sa'adah, N., Haryani, S. and Mahatmanti, W., 2022. Analisis of Chemical Misconceptions of Stoichiometry Materials Using Certainty of Response Index. *International Journal of Active Learning*, 7(1), pp.94-102.
- Soeharto, S., CsapÃ, B., Sarimanah, E., Dewi, F.I. and Sabri, T., 2019. A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, 8(2), pp.247-266.
- Suparno, P. (2013). Miskonsepsi & Perubahan Konsep dalam Pendidikan Fisika. Semarang: PT. Gramedia Widiasarana Indonesia
- Tsaparlis, G., Pappa, E.T. and Byers, B., 2018. Teaching and learning chemical bonding: research-based evidence for misconceptions and conceptual difficulties experienced by students in upper secondary schools and the effect of an enriched text. *Chemistry Education research and practice*, 19(4), pp.1253-1269. <https://doi.org/10.1039/C8RP00035B>
- Yulianova, T. and Maulana, R.H., 2022, August. Perubahan Konseptual Siswa Pada Materi Kelarutan Dan Hasil Kelarutan Melalui Model Pembelajaran Perubahan Konseptual Dengan Bantuan Media Animasi. In *Prosiding Seminar Nasional Pendidikan Kimia* (Vol. 1, No. 1, pp. 44-51).
- Widiyatmoko, A. & Shimizu, K (2018). The Development of Two-Tier Multiple Choice Test to Assess Students' Conceptual Understanding about Light and Optical Instruments. *Jurnal Pendidikan IPA Indonesia*, [journal.unnes.ac.id, https://journal.unnes.ac.id/nju/jpii/article/view/16591](https://journal.unnes.ac.id/nju/jpii/article/view/16591)