Validity and Reliability of the 4TMC Diagnostic Test Instrument Based on Chemical Representation on Chemical Bonding Material to Detect Student Misconceptions

Sabrina Angeli Almira [*1], Alya Aqilah Zahra [2], Dian Novita [3]

[*1,2,3]Chemistry Education Study Program, Faculty of Mathematics and Natural Science, State University of Surabaya, Indonesia

**Corresponding author: sabrinaangeli.21009@mhs.unesa.ac.id*

Abstract. *This study aims to produce a valid Four-Tier Multiple Choice diagnostic test instrument based on validity and reliability, as well as identify students misconceptions on chemical bonding material. This type of research is research and development (R&D). The research design used is 4D modified by Thiagarjan which consists of four stages, namely define, design, develop, disseminate. The research method used is descriptive quantitative. Based on the results of the study, it can be concluded that (1) the validity of the instrument in terms of content and construct aspects has a mode score of 4, (2) the Cronbach Alpha value is 0.825 which means reliable and (3) the percentage of chemical bond misconceptions held by students for the ion bond sub-concept of 33,8%, metal bonds of 40%, and covalent bonds of 6,20%. Based on the validity and reliability results, the diagnostic test instrument developed can be declared valid for use.*

Keywords: *four-tier multiple choice, misconception, chemical bonding*

INTRODUCTION

Chemistry is a branch of science that explores aspect such as the composition, structure, properties, a substance, theories, concepts, facts, and so on [1]. The concept of chemistry must be obtained by students as a strong foundation to carry out subsequent learning. If students misunderstand the concept, it will result in misconceptions. Student misconceptions should not be ignored. It is important to identify them early to prevent interference with the understanding of related concepts learned later. Student misconceptions have been found in many chemistry topics, especially in abstract concepts at the atomic and molecular level, such as chemical bonding. Chemical bonding refers to the interactions that connect two or more atoms to create compounds or molecules [2] . Chemical bonds are forces that bind two or more atoms to form chemical compounds or molecules.

Misconceptions are caused by students who do not understand chemical concepts at 3 levels of chemical representation [3]. Multiple representation is a chemical representation of macroscopic, submicroscopic, and symbolic dimensions [4]. Identifying misconceptions in students is a very important initial step to take.

Various types of instruments have been used to identify misconceptions. Diagnostic tests can be used to identify misconceptions experienced by students [5]. Diagnostic tests can be categorized into several types, namely one-level, two-level, three-level, and four-level formats. One-tier diagnostic tests consist of simple multiple choices and cannot distinguish between students who answer correctly with correct reasons or correctly with incorrect reasons. Two-tier diagnostic tests consist of multiple choices and reasons that must be selected, but these two-tier diagnostic tests cannot determine how deeply students understand the concept given. Three-tier diagnostic tests consist of multiple choices, reasons, and levels of confidence in the reasons. These diagnostic tests cannot detect students' levels of confidence if they have different beliefs in choosing between answers and reasons. The diagnostic test consists of 4 tiers. The first tier consists of multiple choice questions, the second tier is the level of confidence when answering the first tier

questions, the third tier consists of several choices of reasons for answering the first tier and the fourth tier is the level of confidence when answering the third tier questions [6]. The addition of levels of confidence in each answer and reason can measure differences in student's levels of knowledge so that it can help identify misconceptions experienced by these students.

One of the instruments that can be used is 4TMC diagnostic test instrument. This 4TMC instrument is a two-tier test with the addition of two tiers of confidence also called confidence rating (CR). This confidence rating has six options, namely just guessing, very unsure, unsure, confident, very confident and extremely confident [7]. The confidence rating in this 4TMC test instrument allows students to determine the level of confidence separately for their choices of answers and reasons [8]. The development of this 4TMC instrument was carried out by compiling a two-tier test and adding a confidence rating.

Diagnostic test instruments are evaluated for their feasibility based on validity and reliability. Validity encompasses both content and construct validity. Experts conduct validation to ensure the instrument is suitable for measuring the targeted constructs [9] . In this study, it is student misconceptions. Reliability is the consistency and stability of data. Reliability is related to trust. A test can be said to have a high level of trust if the test can provide consistent results even though several tests are carried out at different times [9]. Based on the description above, the researchers will develop a 4TMC diagnostic test instrument that utulizes chemical representations to detect misconceptions about the concept of chemical bonds.

METHOD

This type of research is research and development (R&D). The research design used is 4D modification Thiagarajan [10] which can be seen in picture 1 below.

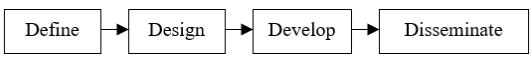

Figure 1. Thiagarajan Modified 4D Research Design Diagram

Based on the modified 4D research design diagaram Thiagarajan, it can be explained in more detail through the following flow:

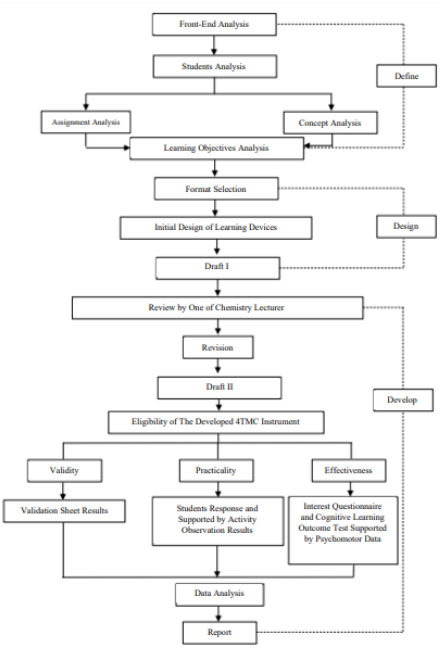

Figure 2. Flow of 4D Research Design Adapted by Ibrahim

Ibrahim's adaptation research design is carried out up to the development stage [11]. The following is an explanation of each stage:

A. Define

Define stage involves initial analysis, student analysis, and material analysis. At this stage, observations were made in class XI of State Senior High School 18 Surabaya. This observation is needed to identify the actual situation, especially those related to the development of diagnostic tests on chemical bonding material, sub-chapter ionic bonds, covalent bonds and metal bonds. The define stage includes five key steps, namely front-end analysis, learner analysis, task analysis, concept analysis, and formulating learning objectives.

B. Design Stage

The next stage of the 4D model is the design stage. Design stage aims to design the product to be developed. The following are the steps in the design stage:

1. Determination of material

Researchers determine what materials are included in the diagnostic test. In this study, the materials used in developing the 4TMC diagnostic test instruments are chemical bonds, sub-chapters of ionic bonds, covalent bonds and metal bonds.

2. Determining the sequence of materials/concepts

In this study, the sequence of concept used as a reference in making the sequence of questions on the diagnostic test namely ion bonds, physical properties of ion compounds, covalent bonds (single covalent bonds, double covalent bonds, coordination covalent bonds, polar and non-polar covalent bonds), physical properties of covalent compounds and metal bonds and physical properties of metal compounds.

3. Designing the diagnostic test question grid

In this study there are several components in the design of diagnostic test question grids, namely title, education unit, subject, class/semester, submaterial, question number, question indicator, question item, answer, cognitive domain and representation level.

4. Designing diagnostic test items

The diagnostic test developed is 4TMC with macroscopic, submicroscopic, and symbolic representations. This 4TMC diagnostic test consists of four levels of questions. The design of the first tier questions is adjusted to the question indicators, the design of the second tier assesses students' confidence in the answers they choose in the first tier, the design of the reasons in the third tier adjusts each answer choice contained in the first tier and the design of the fourth tier assesses students' confidence in the answers they choose in the third level.

In the last stage of design, the first draft 4TMC diagnostic test instrument was produced.

C. Development Stages

The development stage aims to produce a good final draft. The following are the steps in the development stage:

1. Review of Diagnostic Test Instruments

At the initial stage, the review was carried out by one chemistry lecturer to get advice and input on the initial design that had been made. The review was carried out by the supervisor. After getting suggestions and input, revisions were made in accordance with the suggestions and input from the reviewer.

2. Validation of diagnostic test instruments

After review and improvement, the diagnostic test instrument was validated. This diagnostic test instrument was validated by three expert validators, namely two Unesa Chemistry Education lecturers and a chemistry teacher at State Senior High School 18 Surabaya. There are two aspects that will be assessed, namely content validity and construct validity. After receiving assessments and suggestions from three validators, a revision was made to produce a valid 4TMC diagnostic test instrument for use.

3. Limited trial

After analyzing the validation results, the instrument can be tested on students to obtain student answers which will later be identified as students' understanding based on the answers. The diagnostic instrument consisted of 25 multiple choice questions and was tested limited to 29 students at State Senior High School 18 Surabaya. The following is an interpretation of students' answers when working on the 4TMC diagnostic test.

Answer	Answer Confidence Level	Reason	Confidence Level of reason	Criteria
Correct	Tall	Correct	Tall	Understand the Concept
Correct	Low	Wrong	Tall	Misconception
Correct	Tall	Wrong	Tall	
Wrong	Tall	Correct	Low	
Wrong	Tall	Correct	Tall	
Wrong	Tall	Wrong	Low	
Wrong	Low	Wrong	Tall	
Wrong	Tall	Wrong	Tall	
Correct	Low	Correct	Low	Don't Understand the Concept
Correct	Tall	Correct	Low	
Correct	Low	Correct	Tall	
Correct	Low	Wrong	Low	
Wrong	Low	Correct	Low	
Wrong	Low	Wrong	Low	
Correct	Tall	Wrong	Low	
Correct	Low	Correct	Tall	

Figure 3 Interpretation of Student Answers

Based picture 3, students can be categorized into 3 levels of misconceptions as follows:

Table 1 Misconception Criteria

Percentage of Misconceptions	Misconception Criteria
0-30	Low
31-70	Moderate
71-100	High

[12]

The validation results of the diagnostic test instrument were quantitatively analyzed using Likert scale-based percentage calculations, as shown in Table 2:

Table 2 Likert Scale Validation Sheet

Evaluation	Scale Value
Very good	4

Good	3
Not good	2
Very bad	1

[9]

The data obtained from the validation are categorized as ordinal and were analyzed by identifying the mode for each aspect with the following conditions:

a) If the aspect assessed by the validator has a score mode ≥ 3, then the aspect is declared valid.
b) If the aspect assessed by the validator has a score mode < 3, then the aspect is declared invalid.

If any aspect does not meet the applicable requirements, it must be revised and revalidated until it reaches the specified criteria.

After a limited trial was conducted, students' answers were analyzed to determine the reliability of the developed test instrument. Test reliability shows how consistent the test scores are obtained from measurement results at different times [11]. The way to determine the reliability of the instrument is, each student who answers correctly in tiers 1 and 3 and is confident in tiers 2 and 4 (concept understanding category) will be given a score of 1, while other than the combination of answers is given a score of 0. The instrument is regarded as reliably consistent if it yields a Cronbach's alpha value > 0.6 [14]. Testing to measure how consistent (reliable) an instrument that is developed can use SPSS.

Table 3 Level of Instrument Reliability

Cronbach's Alpha	Reliability level
0.00-0.20	Very Low
0.201-0.40	Low
0.401-0.60	Enough
0.601-0.80	Tall
0.801-1.00	Very high

[15]

RESULTS AND DISCUSSION

This study uses research and development (R&D) method using Thiagarajan's modified 4D research design which is limited to develop stage and limited trials. The initial stage is define. At this stage, observations were conducted in class XI of State Senior High School 18 Surabaya to find problems. Students have dificulty in chemistry lesson, especially chemical bonding. At the design stage, the researcher determined the subject matter used in this study, namely chemical bonding which consist of several sub consept namely ionic bonds, covalent bonds and metal bonds. Next, the researcher designed the test blueprint and developed the items for the diagnostic test.

The next stage is develop stage which produces a product in the form of a 4TMC diagnostic test instrument. The diagnostic test questions that were developed are validated by three expert validators, namely two UNESA chemistry education lecturers and one chemistry teacher at State Senior High School 18 Surabaya. After validation, the instrument was tested on students and later the students' understanding was identified based on the answers. The diagnostic test consisted of in the form of 25 multiple choice questions and was tested on limited sample 29 students at State Senior High School 18 Surabaya. The feasibility of the diagnostic test instrument was reviewed based on two aspects, namely validity and reliability.

The feasibility of the diagnostic test instrument was reviewed based on two aspects: validity and reliability. Based on the trial results, the percentage of misconceptions in the sub-concepts was 33.8% for ionic bonds, 6.2% for covalent bonds, and 40% for metallic bonds.

Test Validity

Validity is a measure that shows the level of truth of a test and can measure what should be measured. The validity assessed in this test instrument includes content and construct validity. Content validity includes aspects of the suitability of the question items with the concept and indicators of the question. Construct validity includes aspects of the suitability of the first tier questions with the answer choices, distractors in the first tier answer choices are also rational and homogeneous, distractors in the third tier reason choices are rational and homogeneous with the first-tier answers and the suitability of the images to the problems presented. Validation was conducted to determine whether the 4TMC instrument is valid and can assess students' misconceptions. The instrument is considered valid if it receives a mode score ≥ 3.

Validity of Content

Content validity includes: (1) Suitability of the question items with question concepts and indicators and (2) Suitability of the questions with the answers and justifications. Table 4 below shows the result of the content validity assessment of 4TMC diagnostic test instrument.

Table 4 Calculation Result of Content Validity Assessment of 4TMC Diagnostic Test Instrument

No	Validated Aspects	Mode	Criteria
1	Suitability of question items with question concepts and indicators	4	Very good
2	Compliance of questions with answers and reasons	4	Very good

From the results of the validity assessment in Table 4 can be seen that the content validity of this diagnostic test instrument shows a mode of 4 because the question items are in accordance with the concept and indicators of the questions and the suitability of the questions with the answers and reasons. Construct validity is classified as very good criteria which is as shown in Table 2.

Construct Validity

Construct validity includes: (1) the suitability of the first-tier questions with the answer choices, (2) the distractors in the first tier answer choices are rational and homogeneous and (3) the distractors in the third tier reason choices are rational and homogeneous with the first tier answers. Table 6 below is the calculation of the results of the construct validity assessment of the four-tier multiple choice (4TMC) diagnostic test instrument.

Table 5 Calculation Results of Construct Validity Assessment of 4TMC Diagnostic Test Instrument

No	Validated Aspects	Mode	Criteria
1.	Compliance of first tier questions with answer choices	4	Very good
2.	The distractors in the first tier answer choices are rational and homogeneous.	4	Very good
3.	The distractors in the third tier of reason choices are rational and homogeneous with the first tier answers.	4	Very good

Based on Table 5, the construct validity of the diagnostic test instrument has a mode of 4 because the first tier questions with the appropriate answer choices, the distractors in the first tier answers are rational and homogeneous and the distractors in the third tier reason choices are rational and homogeneous with the first level answers. Content validity is classified as very good criteria in accordance with table 2.

Test Reliability

Reliability is a consistency and stability of data. Reliability is related to trust. A test can be said to have a high level of trust if the test can provide consistent results even though several tests are carried out at different times. Presented below are the outcomes of the diagnostic test reliability calculation performed with SPSS.

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
X01	2.8276	9.862	.460	.816
X02	3.0690	10.138	.430	.817
X03	3.3448	11.591	.000	.826
X04	3.3448	11.591	.000	.826
X05	3.3448	11.591	.000	.826
X06	3.3448	11.591	.000	.826
X07	3.2759	11.064	.268	.822
X08	3.3448	11.591	.000	.826
X09	3.3103	11.079	.387	.820
X10	3.3448	11.591	.000	.826
X11	3.2414	10.761	.361	.819
X12	2.9310	9.567	.572	.808
X13	3.0345	10.463	.298	.825
X14	3.1379	10.195	.466	.814
X15	3.1034	9.382	.757	.796
X16	3.1379	9.623	.703	.800
X17	2.8966	9.453	.605	.806
X18	3.2414	10.618	.434	.816
X19	3.2759	10.850	.397	.818
X20	3.3103	11.293	.211	.824
X21	3.2414	10.690	.397	.817
X22	3.2414	10.547	.471	.814
X23	3.3448	11.591	.000	.826
X24	3.3103	11.293	.211	.824
X25	3.2759	10.993	.311	.821

Cronbach's Alpha	N of Items
.825	25

Figure 6 Result Reliability of 4TMC Diagnostic Test Using SPSS

Based on table 6 above, it can be seen that the developed diagnostic test instrument has a Cronbach's Alpha value of 0.825. This shows that the developed 4TMC diagnostic test instrument is said to be reliable because the Cronbach's alpha value is > 0.6 [13] and is in the very high category according to table 3. Thus, the four-tier multiple choice (4TMC) diagnostic test instrument in the concept of chemical bonds that has been tested can be declared reliable [10]. Based on the results of the validity and reliability aspects, it can be seen that overal the four-tier multiple choice (4TMC) diagnostic test instrument that has been developed can be said to be valid and can be used to identify student misconceptions in the sub-materials of ionic bonds, metal bonds and covalent bonds.

CONCLUSION AND SUGGESTIONS

The 4TMC diagnostic test instrument on the concept of chemical bonding consisting of 25 items along with the scoring rubric is declared feasible to use with validity which obtained mode = 4 with a very good category and reliable with the results of Cronbach's alpha of 0.825.

In this study, only the identification of misconceptions based on 3 levels of chemical representatin was carried out. It is expected that future researchers can identify conceptual understanding, do not understand and percentage of multiple representation the concept on the concept of chemical bonds or other subject concepts.

ACKNOWLEDGMENTS

I would like to express my gratitude to all parties who have provided support and contribution in the preparation of the article on the 4TMC diagnostic test on chemical bonding materials. My special thanks to Dr. Dian Novita S.T, M.Pd for her invaluable assistance, guidance, and input during the research and writing process of this article.

I would also like to express my gratitude to Mrs. Wheny and the students of grade XI-3 of State Senior High School 18 Surabaya who were involved in this diagnostic test trial, who have given their time and dedication so that this research can run smoothly. Thank you to my friends, Alya Aqilah Zahra for your cooperation and support during the implementation of this research. Hopefully this article is useful for readers and can be a meaningful contribution in the field of chemistry.

REFERENCES

[1] A. Indah, “Pengaruh Metode Home Experiment Terhadap Hasil Belajar Kimia Siswa Pada,” UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA, 2023.

[2] K. S. Taber, “Challenging Misconceptions in the Chemistry Classroom: Resources to Support Teachers,” *Educ. Quim.*, vol. 4, pp. 13–20, 2009, [Online]. Available: http://scq.iec.cat/scq/index.html

[3] C. Rahmi, W. Wiji, and S. Mulyani,

"Model Mental Miskonsepsi Pada Konsep Kesetimbangan Kelarutan," *Lantanida J.*, vol. 8, no. 1, p. 64, 2020, doi: 10.22373/lj.v8i1.7108.

[4] H. Retno Widarti, "Desain program pembelajaran berbasis multiple representation melalui cognitive dissonance untuk mereduksi miskonsepsi kimia," 2022.

[5] J. F. Law, "Diagnosis Of Student Understanding Of Content Specific Science Areas Using On-Line Two-Tier Diagnostic Tests," *Sci. Math. Educ. Cent.*, no. April, pp. 1–136, 2008.

[6] U. Agustin, "Pengembangan Instrumen Tes Diagnostik Four-Tier Multiple Choice Untuk Identifikasi Miskonsepsi Siswa Pada Materi Kesetimbangan Kimia," pp. 1–132, 2020.

[7] Q. Fariyani, "Journal of Innovative Science Education PENGEMBANGAN FOUR-TIER Journal of Innovative Science Education," vol. 4, no. November, pp. 66–75, 2015.

[8] I. S. Caleon and R. Subramaniam, "Do students know what they know and what they don't know? using a four-tier diagnostic test to assess the nature of students' alternative conceptions," *Res. Sci. Educ.*, vol. 40, no. 3, pp. 313–337, 2010.

[9] Sugiyono, *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung: Alfabeta, 2019.

[10] S. . Thiagarajan, *Instructional Development for Training Teacher of Exceptional Children*. Indiana: Indiana University, 1974.

[11] M. Ibrahim, *Seri Pembelajaran Inovatif : Konsep, Miskonsepsi dan Cara Pembelajarannya*. Surabaya: Unesa University Press, 2012.

[12] H. Okmarisa *et al.*, "LARUTAN ELEKTROLIT DAN NONELEKTROLIT MENGGUNAKAN FOUR," vol. 5, pp. 23–31, 2021.

[13] K. Suprananto, *Pengukuran dan penilaian pendidikan*, no. 10. 2012.

[14] W. Sujarweni, *SPSS untuk Penelitian*. Yogyakarta: Pustaka Baru Press, 2015.

[15] N. R. Reta, "Chemistry education practice," *Chem. Educ. Pract.*, vol. 1, no. 2, pp. 20–25, 2018.

[16] S. Wiratna, *Metodologi Penelitian*. Yogyakarta: Pustaka Baru, 2014.