Jurnal Pembelajaran Sains
VOLUME 5 Number 2, December 2021

http://journal2.um.ac.id/index.php/
ISSN: 2527-9157

DEVELOPMENT AND VALIDATION OF FOUR-TIER DIAGNOSTICS INSTRUMENT FOR CHEMICAL EQUILIBRIUM

Nuchrurita Rosida[1], Hayuni Retno Widarti[2], Rizki Amelia[3]
[1] Universitas Negeri Malang
[2] Universitas Negeri Malang
[3] Universitas Islam Negeri Maulana Malik Ibrahim Malang

Email: nuchrurita@gmail.com[1]; hayuni.retno.fmipa@um.ac.id[2], rizkiamelia@uin-malang.ac.id[3]

Abstract

Misconceptions can affect students' understanding. Identifying misconceptions is an important first step in gaining an understanding of student learning. One of the instruments that can be used to diagnose student misconceptions is a four-tier diagnostic instrument. This study aims to develop a four-tier diagnostic instrument on the material of chemical equilibrium. This research is a research and development (RnD) research. The research stages consisted of concept mapping, pre-test, identification of unscientific student concepts, developing prototypes, validating prototypes, and completing the final prototype. The product development stages consist of an analysis of learning tools, preparation of test question grids, writing of items, as well as reviewing and revising questions. The data analysis carried out included validity, reliability, difficulty level, and distractor effectiveness. The final product that is produced consists of 20 items on the sub-topic of dynamic equilibrium, equilibrium constant, and equilibrium shift. The results of content validation on two expert lecturers found that the four-tier diagnostic instrument on chemical equilibrium material was valid and suitable for use. In empirical validation, it was found that 3 items were invalid. The resulting test questions are reliable with a reliability score of 0.77. In general, this four-tier diagnostic instrument is at a moderate level and is an effective distractor. This shows that a four-tier diagnostic instrument on chemical equilibrium material can be used to diagnose student misconceptions.

Keywords*: Diagnostic Test, Four-Tier, Misconceptions, Chemical Equilibrium*

Accepted: December 2020, Revised: June 2021, Published: August 2021

INTRODUCTION

One of the important things that must be brought up when students attend class is conception. Conceptions that are contrary to scientific views are often referred to as misconceptions. Misconceptions can occur when students build their understanding based on inadequate initial knowledge so that student constructs are different from those of experts (Barke et al., 2008; Erman, 2017; Qian & Lehman, 2017). These misconceptions will affect students' understanding of the material as a whole because misconceptions develop along with the learning process (Tompo et al., 2016). Therefore, identifying correct misconceptions is an important first step in gaining an understanding of student understanding and learning. One of the chemistry materials studied by high school students is chemical equilibrium. This material has a tiered character, meaning that this material can be a basic material that must be understood by students before they understand other more complex materials, such as acid-base material, salt hydrolysis, solubility product, reduction-oxidation, and buffer solution. There have been quite a several studies that have developed instruments to analyze the misconceptions experienced by students on chemical equilibrium material, but there are still few developments that use four-tier diagnostic tests.

Various measuring tools to identify misconceptions have been developed and implemented by various researchers. Interview method (Chen, 2009), concept map (Novak, 1996), open questionnaire, multiple-choice test (Sadhu et al., 2017), a two-tier test (Siswaningsih et al., 2017), three-tiers (Milenković et al., 2016; Mubarokah et al., 2018), and four tiers (Fariyani et al., 2017) often used to diagnose students' conceptions in science education. Interview methods, open tests, and multiple-choice are often used in science education. However, each diagnostic tool certainly has advantages and disadvantages.

In the interview method, it is crucial because the interview method will get an in-depth diagnosis related to student conceptions. However, it will take more time. The open question method allows respondents to write their answers in the form of their sentences, but this takes time to score students' answers. Furthermore, multiple-choice methods can be used in large numbers of individuals and the analysis is also easy, but the results of the diagnosis are not in-depth (Caleon & Subramaniam, 2010).

The two-tier test is a diagnostic instrument with multiple-choice questions at the first level and a choice of reasons at the second level. Two-tier tests have been used to diagnose misconceptions. However, this two-

tier test is considered to underestimate students' scientific concepts (Chang et al., 2007) or exaggerate the proportion of misconceptions due to lack of knowledge of the two-tier test (Caleon & Subramaniam, 2010). The drawback of this two-tier test is attempted to be solved by the three-tier test. The three-tier test is a reasoned multiple choice with confidence in that choice. Although the three-tier test seems to eliminate many of the shortcomings mentioned for the two-tier test, the three-tier test is still unable to completely differentiate the belief choice for the main (first-order) answer from the belief choice for reasoning (the second level) and may therefore overestimate over-score students and underestimate their lack of knowledge (Gurel et al., 2015).

The four-tier diagnostic test is used to cover the shortcomings that exist in the three-tier test. The four-tier diagnostic test has four levels. The first level is the main question and answers choices, the second level is the scale of confidence in choosing answers, the third level is the choice of reasons, and the fourth level is the scale of confidence in choosing reasons (Habiddin & Page, 2019). This four-tier diagnostic test is considered effective in diagnosing student misconceptions, due to its complex analysis (Gurel et al., 2015).

Chemistry is a science that studies several abstract things such as structure, properties, changes in matter, and energy that accompany the changes. One of the materials taught in chemistry is chemical equilibrium. This material demands conceptual as well as mathematical or algorithmic understanding. To understand abstract chemistry, it also requires an understanding of multiple representations. However, many studies reveal that there are still many students who do not understand the equilibrium material, especially the submicroscopic aspects. This can be the reason for learning difficulties experienced by students. If left without any effort, it can lead to misconceptions. Thus, a diagnostic instrument for misconceptions in this chemical equilibrium material is needed.

RESEARCH METHODS

This research is a research and development (RnD) research. The learning device developed is a four-tier diagnostic test to diagnose student misconceptions on chemical equilibrium material. This research procedure was carried out using the research and development procedure of a four-tier diagnostic test according to Habiddin & Page, (2019). The research stages consisted of concept mapping, pre-test, identification of unscientific student concepts, developing prototypes, validating prototypes, and completing the final prototype. The product development stages consist of an analysis of learning devices, preparation of test question grids, writing of items, and reviewing and revising questions.

Methods of data collection using tests, to determine the validity and reliability of the test. The trial involved 41 students of class XI SMAN 3 Malang and 59 students of class XI SMAN 8 Malang. The data analysis carried out included validity, reliability, difficulty level, and distractor effectiveness. Validiyass testing uses content validity by two expert lecturers. Testing the empirical validity using the product-moment. Reliability testing uses the Alpha Cronbach formula. The level of difficulty and effectiveness of the distractor was analyzed using the difficulty level formula (p).

RESULTS AND DISCUSSION

Development of a Four-Tier Diagnostics Test Instrument

The initial stage of the preparation of the four-tier diagnostic test instrument is to identify the concept of the material and compile a grid of open reasons multiple-choice questions. based on the results of the concept mapping that has been done, it is found that the concept to be studied in this development research is chemical equilibrium taken from the 2013 curriculum syllabus on basic competence (KD) 3.8, which explains the equilibrium reaction in the relationship between reactants and reaction products. Topics of material contained in KD 3.8 include dynamic equilibrium, equilibrium constants, calculations, and application of chemical equilibrium. The next step is to compile a grid of open reasons multiple-choice questions. This grid is used as a guide for constructing multiple-choice questions with open reasons which are then validated by experts. After being revised based on the results of expert validation, the open-ended multiple-choice questions were used for the initial test on students. This test is conducted to obtain a variety of reasons from students which will later be used as a choice of reasons in the third tier for the four-tier diagnostic test. After compiling a four-tier diagnostic test, expert and empirical validation are then carried out to determine the final prototype of the test.

This study developed a four-tier diagnostic test instrument to analyze the misconceptions experienced by students in the chemical equilibrium material, especially in the sub-concepts of dynamic equilibrium, equilibrium constants, calculations, and application of chemical equilibrium. The results of the products developed consist of four-tier diagnostic test lattice, question work instructions, four-tier diagnostic test questions, answer keys, answer sheets, scoring guidelines, results in interpretation guidelines. The four-tier

diagnostics test question grid consists of sub-topics, question indicators, category level questions, and several questions. Instructions for working on questions are instructions for students in working on questions. The instrument for four-tier diagnostics test questions consists of the title, subject, class, subject, processing time, test questions, answer choices, the level of confidence in choosing the answer, the choice of reason, the level of confidence in choosing the reason. The answer key consists of the question number, the choice of answer, and the choice of the correct reason. Scoring guidelines are guidelines for scoring and determining test results.

Each item developed consists of four levels. The first level is in the form of multiple-choice questions with three distractors and one answer key that students must choose. The second level is the level of student confidence in choosing answers. The third level is the student's reason for answering the question, in the form of three choices of reasons that have been provided and one open reason. The fourth level is the level of student confidence in choosing reasons. The level of confidence in choosing answers and reasons is divided on a scale of one to six. A scale of one is chosen if the student guesses, a scale of two if very unsure, a scale of three if not sure, five if very sure, and six if very sure.

The final test questions used consisted of 20 items on the sub-topic of dynamic equilibrium, equilibrium constant, and equilibrium shift. Scoring is given by giving a score of 1 for the choice of answer or choice of the correct reason and a score of 0 is given for the choice of answer or choice of wrong reasons. The level of confidence is high if a scale of 4 or 5 is chosen and the level of confidence is low if it is selected with a scale of 1 or 2 or 3.

Validity of the Four-Tier Diagnostics Test Instrument

Validation is carried out to ensure whether the instrument used is appropriate and can measure what will be measured, in this study, it is a student's misconception. The validity test of the four-tier diagnostics test instrument consists of content validation and empirical validation. Content validation was carried out by two expert lecturers. Content validation relates to 1) suitability of questions with indicators, 2) suitability and correctness of questions with the measured concept, 3) formulation of communicative sentences, and using good and correct language, 4) involving representations and showing the relationship. The results of content validation by two expert lecturers are presented in Table 1.

Table 1. Results of Content Validation by Two Expert Lecturers

Aspect	Validator 1	Validator 2	Average
Suitability of Problems with Indicators	4	4.08	4.04
Conformity and correctness of questions with the concept being measured	4	4.08	4.04
Formulation of communicative sentences and use good and correct language	3.84	4.08	3.96
Involves representations (at least 2) in multiple representations, and shows the relationship	4	4.08	4.04

In general, the results of content validation based on two expert lecturers found that on the aspect of the suitability of the questions with the indicators, the suitability and truth of the questions with the measured concept, and involving the representation an average score of 4.04, and the aspects of the formulation of communicative sentences got a score of 3, 96. Thus, according to two expert lecturers the four-tier diagnostic test instrument is valid.

Empirical validity is the validation result obtained from the trial results. This four-tier diagnostic test instrument was tried out on 100 high school class XI students. Then, the student answers are given a score of 0 for wrong answers and 1 for correct answers. Furthermore, the students' answers were analyzed using the Pearson product-moment. Based on the Pearson product-moment analysis, empirical validation results are obtained as shown in Table 2. Of the 25 four-tier diagnostic test questions on chemical equilibrium, 22 valid and 3 invalid items were obtained, namely the number 3, 8, and 25 questions. The description of the results of the empirical validation is presented in Table 2.

Table 2. Results of Empirical Validation

NO. QUESTION	INFORMATION	NO. QUESTION	INFORMATION
1	Valid	14	Valid
2	Valid	15	Valid
3	Invalid	16	Valid
4	Valid	17	Valid
5	Valid	18	Valid
6	Valid	19	Valid
7	Valid	20	Valid
8	Invalid	21	Valid
9	Valid	22	Valid
10	Valid	23	Valid
11	Valid	24	Valid
12	Valid	25	Invalid
13	Valid		

Reliability of the Four-Tier Diagnostics Test Instrument

Reliability is the level of consistency of questions in assessing what you want to assess. According to Matondang (2009), reliability states the extent to which the results of a measurement can be trusted. The reliability analysis of the four-tier diagnostics test instrument used Cronbach's alpha with the help of SPSS. The results of the reliability analysis using SPSS are shown in Figure 1. The results of Cronbach's Alpha show that Cronbach's alpha value is 0.769, so it can be stated that the four-tier diagnostic test instrument is reliable.

Reliability Statistics

Cronbach's Alpha	Part 1	Value	.723
		N of Items	11[a]
	Part 2	Value	.849
		N of Items	11[b]
	Total N of Items		22
Correlation Between Forms			.653
Spearman-Brown Coefficient	Equal Length		.790
	Unequal Length		.790
Guttman Split-Half Coefficient			.769

Figure 1. Reliability Results with Alpha Cronbach

Difficulty and Distractor Effectiveness of Four-Tier Diagnostics Test Instruments

Apart from validity and reliability, the four-tier diagnostic test instrument also needs to be tested for the difficulty and effectiveness of the distractor. The level of difficulty has been developed for 22 questions, in the first tier, four questions are classified as easy, 13 questions are classified as medium and five questions are classified as difficult. While in the third tier, 14 questions are classified as medium and eight questions are classified as difficult. The description of the difficulty level of the four-tier diagnostics test instrument can be seen in Table 3.

Table 3. Results of Difficulty Level Analysis

No. Question	First-tier		Third-tier	
	P	**Level of Difficulty**	***P***	**Level of Difficulty**
1	0.59	Medium	0.24	Hard
4	0.30	Hard	0.31	Medium
5	0.87	Easy	0.27	Hard
6	0.73	Easy	0.66	Medium
7	0.71	Easy	0.28	Hard
9	0.43	Medium	0.59	Medium
10	0.53	Medium	0.47	Medium
11	0.72	Easy	0.30	Hard
12	0.61	Medium	0.23	Hard
13	0.25	Hard	0.46	Medium
14	0.57	Medium	0.52	Medium
15	0.47	Medium	0.46	Medium
16	0.51	Medium	0.47	Medium
17	0.37	Medium	0.41	Medium
18	0.51	Medium	0.34	Medium
19	0.22	Hard	0.28	Hard
20	0.21	Hard	0.20	Hard
21	0.56	Medium	0.51	Medium
22	0.60	Medium	0.55	Medium
23	0.52	Medium	0.42	Medium
24	0.54	Medium	0.34	Medium

Distractor effectiveness is analyzed on the respective options on the first tier and the third tier on the four-tier diagnostics test instrument. An option is declared ineffective as a distractor if the effectiveness value is less than 0.05. Of the 22 questions, there were 2 questions whose distractors were ineffective, namely items number 5 and 10. Thus, only 20 questions could be used. A description of the distractor's effectiveness can be seen in Table 4.

Table 4. Results of Distractor Effectiveness Analysis

No. Question	DISTRACTOR INDEX									
	FIRST-TIER					THIRD-TIER				
	A	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
1	0.11	0.11	0.12	0.06	0.59	0.22	0.22	0.10	0.24	0.15
4	0.05	0.06	0.55	0.30	0.04	0.31	0.26	0.13	0.11	0.18
5	0.04	0.01	0.06	0.01	0.87	0.07	0.04	0.27	0.35	0.25
6	0.07	0.06	0.73	0.09	0.05	0.05	0.19	0.03	0.06	0.66
7	0.07	0.05	0.06	0.71	0.11	0.28	0.06	0.07	0.17	0.39
9	0.12	0.15	0.43	0.18	0.12	0.59	0.09	0.14	0.09	0.09
10	0.04	0.03	0.01	0.39	0.53	0.47	0.07	0.13	0.14	0.14
11	0.09	0.09	0.06	0.72	0.02	0.12	0.11	0.33	0.09	0.30
12	0.10	0.61	0.13	0.06	0.10	0.29	0.14	0.23	0.29	0.02
13	0.04	0.08	0.10	0.51	0.25	0.07	0.46	0.05	0.19	0.20
14	0.17	0.05	0.57	0.10	0.11	0.12	0.52	0.05	0.09	0.19
15	0.13	0.10	0.47	0.09	0.20	0.08	0.14	0.12	0.46	0.16
16	0.06	0.13	0.08	0.51	0.21	0.47	0.05	0.11	0.20	0.15
17	0.05	0.37	0.07	0.04	0.47	0.13	0.07	0.28	0.08	0.41
18	0.02	0.51	0.10	0.09	0.26	0.10	0.10	0.10	0.34	0.30
19	0.05	0.52	0.02	0.17	0.22	0.21	0.08	0.28	0.05	0.31
20	0.25	0.06	0.21	0.24	0.21	0.25	0.20	0.14	0.08	0.30
21	0.12	0.56	0.10	0.06	0.15	0.14	0.04	0.06	0.21	0.51
22	0.13	0.12	0.60	0.07	0.08	0.04	0.19	0.55	0.05	0.13
23	0.09	0.12	0.11	0.52	0.15	0.42	0.09	0.27	0.10	0.07
24	0.07	0.11	0.54	0.14	0.12	0.30	0.07	0.34	0.10	0.15

CONCLUSIONS AND SUGGESTIONS

A. Conclusion

The four-tier diagnostic test instrument that was produced consisted of a grid of test questions, instructions for handling the questions, test questions, answer keys, answer sheets, scoring guidelines, and results in interpretation guidelines. The test questions consist of four levels, namely questions with one answer key and three cheaters, the level of confidence in the answer, the choice of reason, and the level of confidence in the reason. The final product consists of 20 items on the sub-topic of dynamic equilibrium, equilibrium constant, and equilibrium shift. The results of content validation on two expert lecturers found that the four-tier diagnostic instrument on chemical equilibrium material was valid and suitable for use. In empirical validation, it was found that 3 items were invalid. The resulting test questions are reliable with a reliability score of 0.77. Generally, This four-tier diagnostic instrument is at a moderate level and is an effective distractor. This shows that a four-tier diagnostic instrument on chemical equilibrium material can be used to diagnose student misconceptions.

B. Suggestion

This research is only limited to the process of developing a four-tier diagnostic instrument. For this reason, further research is needed to reveal student misconceptions on chemical equilibrium material using this four-tier diagnostic instrument.

REFERENCES

Barke, H.-D., Hazari, A., & Yitbarek, S. (2008). *Misconceptions in Chemistry: Addressing perceptions in chemical education*. Springer Science & Business Media.

Caleon, I. S., & Subramaniam, R. (2010). 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. *Research in Science Education*, *40*(3), 313–337.

Chang, H.-P., Chen, J.-Y., Guo, C.-J., Chen, C.-C., Chang, C.-Y., Lin, S.-H., Su, W.-J., Lain, K.-D., Hsu, S.-Y., & Lin, J.-L. (2007). Investigating primary and secondary students' learning of physics concepts in Taiwan. *International Journal of Science Education*, *29*(4), 465–482.

Erman, E. (2017). Factors contributing to students' misconceptions in learning covalent bonds. *Journal of Research in Science Teaching*, *54*(4), 520–537.

Fariyani, Q., Rusilowati, A., & Sugianto, S. (2017). A four-tier diagnostic test to identify misconceptions in geometrical optics. *Unnes Science Education Journal*, *6*(3).

Gurel, D. K., Eryılmaz, A., & McDermott, L. C. (2015). *A review and comparison of diagnostic instruments to identify students' misconceptions in science*.

Habiddin, H., & 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.

Matondang, Z. (2009). Validitas dan reliabilitas suatu instrumen penelitian. *Jurnal Tabularasa*, *6*(1), 87–97.

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.

Mubarokah, F. D., Mulyani, S., & Indriyanti, N. Y. (2018). Identifying students' misconceptions of acid-base concepts using a three-tier diagnostic test: A case of Indonesia and Thailand. *Journal of Turkish Science Education*, *15*(Special), 51–58.

Qian, Y., & Lehman, J. (2017). Students' misconceptions and other difficulties in introductory programming: A literature review. *ACM Transactions on Computing Education (TOCE)*, *18*(1), 1–24.

Sadhu, S., Tima, M. T., Cahyani, V. P., Laka, A. F., Annisa, D., & Fahriyah, A. R. (2017). Analysis of acid-base misconceptions using modified certainty of response index (CRI) and diagnostic interview for different student levels cognitive. *International Journal of Science and Applied Science: Conference Series*, *1*(2), 91–100.

Siswaningsih, W., Firman, H., & Khoirunnisa, A. (2017). Development of two-tier diagnostic test pictorial-based for identifying high school students' misconceptions on the mole concept. *JPhCS*, *812*(1), 012117.

Tompo, B., Ahmad, A., & Muris, M. (2016). The Development of Discovery-Inquiry Learning Model to Reduce the Science Misconceptions of Junior High School Students. *International Journal of Environmental and Science Education*, *11*(12), 5676–5686.