



THE EFFECT OF THE ONE-TIER, TWO-TIER, AND THREE-TIER DIAGNOSTIC TEST TOWARD THE STUDENTS' CONFIDENCE AND UNDERSTANDING TOWARD THE CONCEPTS OF ATOMIC NUCLEAR

Yusrizal and A. Halim[✉]

Department of Physics Education, Training Teacher and Education Faculty,
Syiah Kuala University, Aceh, Indonesia.

Article Info

Received April 2017
Accepted June 2017
Published July 2017

Keywords:

*Diagnostic test, misconception,
concept understanding, core
Physics, confidence*

Abstract

The identification of misconception in the Physics or Science learning has used the two-tier or three-tier diagnostic test. The both tests have used as to make sure whether the students only guess the answers or understand the concepts. The students' confidence in choosing the answers also can be identified. However, there has been a study to identify the effectiveness those misconceptions. The purpose of this study is to determine the effect of the one-tier, two-tier and three-tier diagnostics test of the students' answer pattern, confidence level, and understanding on the concept of Atomic Nuclear. The research subject was 8 students attending Atomic Nuclear course in Physics Education Department, Faculty of Teacher Training and Education, Syiah Kuala University. Some tests were consecutively given to the 88 respondents, the test participants could continue to the next test if they finished all questions in the previous test. The data collection used the instrument of basic Physics preconception test (*Tes Pemahaman Konsep Fisika Inti/TPKFI*) in the form of one-tier, two-tier and three-tier multiple choice with eight multiple choice items. The test was developed by researchers with difficulty index = 0.35, discrimination index = 0.31, validity = 0.35, and reliability = 0.90. The analysis result showed that although the students' answer patterns were the same for all types of diagnostic tests, there was a convincing change in the students' concept understanding and confidence between the two-tier and three-tier. Therefore, it could be inferred that the identification of the students' misconception in the Science learning would be better (had high confidence), if the diagnostic test used was higher, that was three-tier, four-tier, and thus on.

©2017 Universitas Negeri Semarang
p-ISSN 2252-6617
e-ISSN 2502-6232

[✉]Corresponding author:

A. Halim

Department of Physics Education, Training Teacher and Education Faculty,
Syiah Kuala University, Aceh, Indonesia.
E-mail: bdlhalim@yahoo.com

INTRODUCTION

A science misconception research involves three main steps, they are (a) developing diagnostic test instruments; (b) analyzing the causes of misconception; and (c) remediation of misconception (Allen, 2010). Related to the test development, many kinds of diagnostic tests had been developed by experts to detect misconceptions, such as (i) closed-ended multiple-choice which was used to detect misconceptions in the force concept inventory (FCI) (Halloun et. al., 1985a), in the Mechanics Baseline Test (MBT) (Viennot, 1979; Clement, 1982; Hestenes et.al 1992), in the Astronomy Diagnostic Test (ADT) (Beth, 2002), in the Fluid Mechanics Concept Inventory (FMCI) (Martin J.K., et.al 2004), (ii) closed-ended multiple-choice and one of the item choices was open-ended had been used to detect misconceptions in the Test of Image Formation by Optical Reflection (TIFOR) (Chung, et.al 2003), in the Materials Concept Inventory (MCI) (Wouldian J, et.al 2005), and in the Kinematics concept inventory (Shaffer, et.al 2005), (iii) multiple choice test accompanied with reasons or two-tier test had been used to detect misconceptions in optic concept (Chung, et. al., 2003), in photosynthesis concept (Griffard and Wandersee, 2010), and in photosynthesis concept (Treagust and Hadlam, 1987), (iv) multiple choice test accompanied with Certain Response of Index (CRI) used to detect misconceptions in mechanics concept (Hadan et. al.; 1999, Masril and Nur Asma, 2002), in Quantum Physics concept (Halim et. al., 2009; Karunia and Rinaningsih, 2013), in Wave concept (Caleon and Subramaniam, 2009), in solid substance concept (Kirbulut et. al., 2014), and Quantum Mechanics Conceptual Survey (QMCS) used to detect misconceptions in Quantum Physics (Wuttifrom et.al., 2006) and understanding in acid and base concept in Chemistry (Dindar et. al., 2011).

The development of diagnostic test forms, from one-tier multiple choice form became two-tier or three-tier diagnostic test based on several considerations, such as the certainty of the students' choices (Fischhoff, et al., 1977; Hadan, et al., 1999; Englehardt and Beischner, 2004), long time to investigate the misconceptions (Chung, et. al., 2003), unable to encompass the students' ability to find reasons (Sahin and Cepni, 2011), unable to differ lack of knowledge from misconception (Arslan, et. al., 2012; Wijaya, et. al., 2016). On the

other hand, the development of diagnostic test from three-tier became four-tier based on the consideration of the students' confidence in choosing the answers and explaining the reasons (Fariyani, et. al., 2015; Hakim, et. al., 2012).

Besides that, there was various new phenomenon related to the effect of test form toward the conceptual understanding of tested concept. Bangert-Drowns (1991) had studied 35 research articles related to the test effect to the conceptual understanding and found that 25 articles showed the positive effects of diagnostic test toward learning and understanding of Science concepts. Phelps (2012) had conducted a meta-analysis study for 10 years from various countries and he found that the tests held had positive effects toward the concepts learning and understanding. Some other researchers found that one-tier type of multiple choice test had positive effects on concept learning and understanding (Butler & Roediger, 2007; Marsh, et.al, 2010; Roediger & Marsh, 2005).

On the other hand, the level addition of diagnostic test would need longer time and there would be difficulties in analyzing the test results and taking decisions related to the concept understanding. According to several results of researches and thoughts, it was necessary to know, for sure, which type of the diagnostic test could encompass misconceptions, the students' confidence in answering the problems, was easier in analyzing and making decisions from the test results. Therefore, this study wanted to find out the students' answer pattern, level of misconceptions, and level of confidence and also the effect of Science concept understanding toward the organization of one-tier, two-tier, and three-tier multiple choice tests.

METHOD

This research used survey technique to 88 the students of Physics Education who were taking Atomic Nuclear course in the odd academic year of 2016/2017. The data collection used Physics preconception test, developed by the researchers in the form of 8 multiple choice items and five answer choices (one choice with the correct concept and four choices as distracters or with the incorrect concept). The Physics preconception test was developed in this research was in the form of -

tier, two-tier, and three-tier and based on Gurcaya and Gulbasa (2005). The one-tier diagnostic test had 8 items with 5 multiple choice answers, while the two-tier test consisted of 8 items with 5 answer choices added with Continuity Response of Indexes (CRI) with 0.5 scales, and the three-tier consisted of 8 items, 5 answer choices, CRI index, and the students were asked to write the reason of the answer or the open-ended reasons. The third instrument had the same content, only the format was different. The result of analysis instrument showed that the discrimination index of the test items in the range of 0.080 – 0.320 with adequate category, the difficulty index of the test items in the range of 0.040 – 0.900 with middle category, the test validity was 0.35 with middle category and valid (2.108), and the test reliability was 0.90 with high category. Based on the result of analysis instrument, it showed that the developed instrument statistically could be used as the instrument of data collection.

This was a survey study (Tamir, 1989; Carpenter, et al., 2008) to 88 the students who were attending Atomic Nuclear course. The study was done for once face to face course with the same rules which were enacted in the mid-term test or the final test. Each respondent got three types of tests, one-tier, two-tier, and three-tier. The same one-tier diagnostic test was given to every respondent. The respondents who finished the one-tier could ask for the second type of diagnostic test that was a two-tier test when handing the answer for the one-tier, as well as for the three-tier diagnostic test. The duration to answer each test was based on the ability of each respondent, the one which was limited only

the total amount of time to answer all tests, which was for one hour of face to face lecturers. For the one-tier test, the respondents only required to choose the correct answer (A, B, C, D, and E), while for the two-tier test respondents needed to fill the index of the answer choice certainty or the confidence index or CRI index with the scale of 0.5, and the three-tier test respondents required to write the reasons for choosing the correct answers in the first step. This kind of research method was based on the method which had been used by several researchers before (Emine, 2015; Widarti, et.al, 2016).

The data analysis used AnaTes2016 program based Microsoft office excel 2007 which had been equipped the instrument analysis and test items, descriptive analysis, and the calculation of *r*. The instrument analysis included the discrimination test of the test items, difficulty index of the test items, test validity, and test reliability with the KR-21 formula. The answer category for the three-tier diagnostic test used the decision matrix which was shown in Table 1 with the category of the conceptual understanding. There were two conditions, which were if the answer and the reason was correct. Nonetheless the CRI index < 2,5 included understanding the concept not the confidence (NK). Secondly, if the answer and the reason was correct, nonetheless the CRI index > 2.5 included understanding the concept correctly (KCC).

Table 1. Student responsibility materials for three-tier diagnostic tests

Answer	Reason	CRI Index	Description
Wrong	Wrong	< 2.5	Do not know the concept
Wrong	Correct	< 2.5	Do not know the concept
Wrong	Wrong	> 2.5	Misconceptions
Wrong	Correct	> 2.5	Misconceptions
Correct	Wrong	< 2.5	Do not know the concept
Correct	Correct	< 2.5	Understand the concept, but not confident with the selected answer
Correct	Wrong	> 2.5	Misconceptions
Correct	Correct	> 2.5	Has the right concept

(Hakim et al., 2012)

RESULT AND DISCUSSION

There were three main purposes which wanted to be obtained after the survey was done with the one-tier, two-tier and three-tier, that was the students' answer patterns (A, B, C, and

D answer choices) from the three types of tests; (ii) the students' confidence level in the two-tier and three-tier diagnostic tests; and (iii) the students' understanding level on the concept of Atomic Nuclear in the two-tier and three-tier diagnostic tests. Therefore, the result and discussion were explained with the appropriate sequence for each purpose.

The students' Answer Patterns

The three types of diagnostic tests had the same answer format that was A, B, C, and D with the same problem materials. The result of the survey for the first purpose was presented in the figure 1.

Based on the profile in Figure 1, it showed that the students' answer patterns entirely the same between one-tier, two-tier, and three-tier tests. In another word, the students had high confidence with the choice in the first level (A, B, C, and D), either

toward the correct answers or the distracters. In other words, the students were not influenced by choosing the answer for the CRI index and writing the reason for the chosen answer. However, there were some the students who were influenced by that request, especially when they answered the questions for number 3 and 6, where the correct answer changed to the wrong answer after they were asked to write the CRI index and reason. According to the same answer pattern as shown in Figure 1, this showed or supported the statement which said that the multiple choice test was not suitable to detect misconceptions, this was also appropriate with the statements from Caleon and Subramaniam (2010) and Treagust (1988).

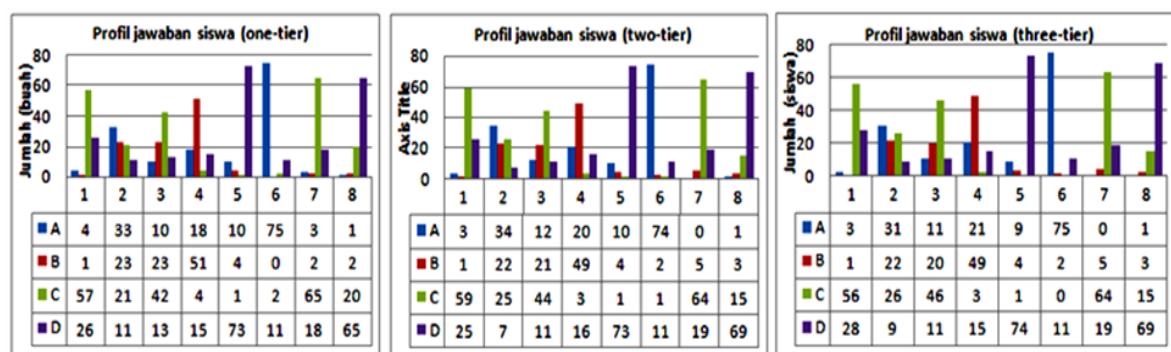


Figure 1. Student response patterns for one-tier, two-tier and three-tier diagnostic tests

The students' Confidence Level

Both two-tier and three-tier types of diagnostic tests had confidence level from the students' answers (CRI index), nonetheless for the three-tier test added with the reason written toward the chosen answer or the open-ended reason. The category of the answer choice for the three-tier test was added by one condition, that was NC category (Not Confidence) with the condition of correct answer and reason, nonetheless the CRI index was low or < 2.5 . It turned out that the category addition or the demand to write the reason for the chosen answer had significantly influenced, as shown in the students' answers on items number 5 and 8. In item number 5, from 62.5% the students who correctly understood the concept for two-tier test decreased to 50% for the three-tier test, as presented in the Figure 2.

Besides that, it could also be inferred that even though the students' answer choices (A, B, C, and

C) stayed the same in the three types of diagnostic tests (Figure 1), there were some convincing changes in the students' concept and confidence categories

(Figure 2). 10% of the students on the average guessed the two-tier diagnostic test, nonetheless in the three-tier one it decreased into 5%, hence the three-tier type of diagnostic test was very effective to reduce the students' guessing habits. Evidently, the addition of the confidence index in the three-tier diagnostic test helped the researcher to understand whether the students chose the answer based on their hunch or their concept understanding, this was corresponding to the statement stated by the previous researchers (Glenberg et al. 1987; Stankov and Dolph, 2000; Yang and Lin, 2015). The confidence of correct concept understanding (KCC) would reduce from 25% in the two-tier test into 18% in the three-tier test, this was why

the level of correct concept understanding indentified through the two-tier diagnostic test could also contain the guessing matter, this was revealed

when the students were asked to write their reasons in the three-tier test.

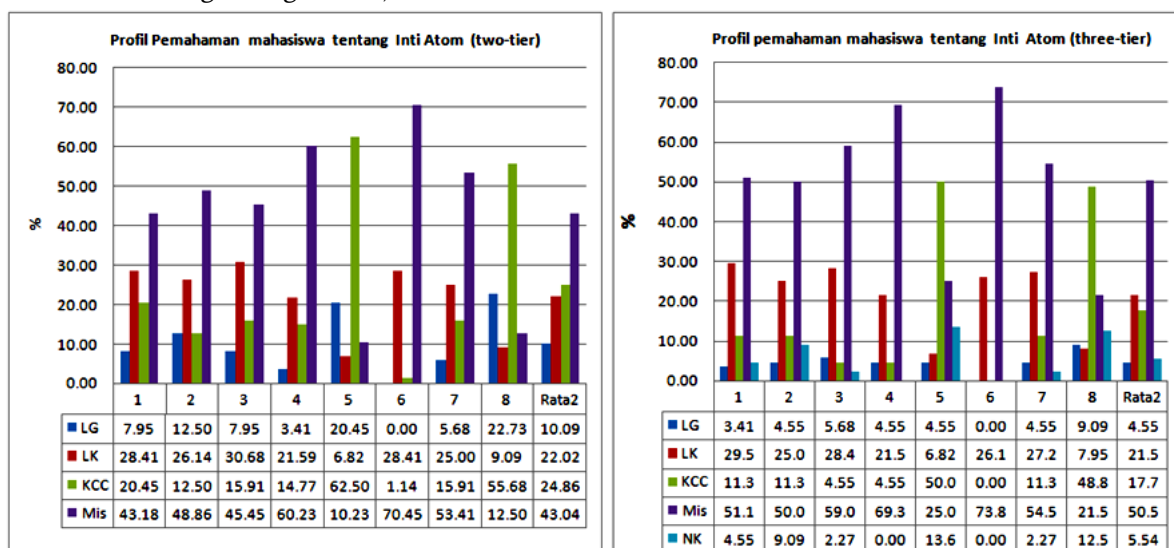


Figure 2. The Effect of two-tier and three-tier tests on confidence and understanding of the Atom Nuclear concept

The students' Understanding Level

The identification of Atomic Nuclear concept understanding or misconceptions in Atomic Nuclear material could only be detected by using two-tier and

three-tier diagnostic tests (because of the confidence scale), while the one-tier multiple choice test could only detect the right and wrong answers.

Table 2. Percentage of Understanding of Atomic Nuclear Concepts from test results with two-tier

Items	1	2	3	4	5	6	7	8	Avr	t23*	r23*
Answer Key	D	B	B	D	D	C	D	D			
LG	7.95	12.50	7.95	3.41	20.45	0.00	5.68	22.73	10.09	0.105	0.737
LK	28.41	26.14	30.68	21.59	6.82	28.41	25.00	9.09	22.02	0.001	0.985
KCC	20.45	12.50	15.91	14.77	62.50	1.14	15.91	55.68	24.86	0.015	0.982
Mis	43.18	48.86	45.45	60.23	10.23	70.45	53.41	12.50	43.04	-0.005	0.974

*) T23 = different identification results between two-tier and 3-tier tests.

R23 = the correlation between the identifiable results between the two-tier and three-tier tests.

The data of the test results using both instruments was shown in Table 2 and Table 3. According to Table 2 and Table 3, it could be known that generally there was no different category in the concept understanding (LG, LK, KCC, and Mis) between the two-tier and three-tier diagnostic tests. This was shown by the t-test with the range value 0.001 – 0.105 and the strong relation between the result of correlation test analysis and the value of correlation coefficient with range 0.737 - .985. However, if each categories of the concept understanding (for example the LG category) was

studied deeper, for every item there would be significant differences. In the item number 5, about 20.45% the students guessed the answers in the two-

tier items, nonetheless the percentage decreased into 4.55% in the three-tier test. This showed that the demand to write the reasons in every answer chosen forced the students to be more careful when choosing the answers, because there were risks in writing the reasons. It was in line with the students' responses after the test. They said that if the items were chosen without

reasons, they would usually guess, another student stated that the responsibility should be set first and foremost, the other one explained that the first question was ordinary, the second one had higher tension and burden, let alone the third question." It meant that the students had higher mental burden and responsibility for the items with more level that was why the guessing matter decreased, this was in line with the statements of the previous researchers (Glenberg et al. 1987; Stankov and Dolph, 2000; Yang and Lin, 2015).

Furthermore, for problem no.8 also shows the same phenomenon, where the LG level (the guess element) in answering the two-tier test is higher (22.73%), while in the three-tier test only about 9.09%. Therefore, it can be said that the student is not confident (13.64%) in understanding the concept (no.5) about: the actual atomic form "and also no confidence (12.5%) in understanding the concept (no.8) about: Nucleons in the atomic nucleus. Based

on the proof of the answer from both questions, it can be said that the guess element (LG) happens in a matter of one-tier choice, not only because it does not understand the concept, but already there is understanding in the form of vague or unclear / unsure.

Either the certainty index of writing the answer or the confidence index of writing the reason for every items had strong correlation between the two-tier and the three-tier tests, as shown by the coefficient value of r_{24} in Table 2 with the range of coefficient value 0.84 – 0.99. This showed that during the diagnostic test the students could also understand the concept of Atomic Nuclear from the test materials, this viewpoint was in accordance with Emine, et al. (2015).

Table 3. Percentage of Understanding of Atomic Nuclear Concepts with three-tiered diagnostic tests

Items	1	2	3	4	5	6	7	8	Average
Answer Key	D	B	B	D	D	C	D	D	
LG	3.41	4.55	5.68	4.55	4.55	0.00	4.55	9.09	4.55
LK	29.55	25.00	28.41	21.59	6.82	26.14	27.27	7.95	21.59
KCC	11.36	11.36	4.55	4.55	50.00	0.00	11.36	48.86	17.76
Mis	51.14	50.00	59.09	69.32	25.00	73.86	54.55	21.59	50.57
NK	4.55	9.09	2.27	0.00	13.64	0.00	2.27	12.50	5.54
r_{23}	0.98	0.99	0.97	0.98	0.84	1.00	0.99	0.89	

CONCLUSION

Based on the research results and the data analysis results as well as the interview result with the students, there were some important notes, as follow; (i) the answer patterns given by the students in the diagnostic tests on the Atomic Nuclear concept were similar with the one-tier, two-tier and three-tier test types; (ii) there were two concepts of Atomic Nuclear triggering the students to guess the answers, that were the concept of the real form of atom and the concept of force between nucleons in the atomic nuclear; (iii) the level addition of the diagnostic test from one-tier test to two-tier, three-tier and thus on, apparently would force the students to be more careful (confidence) in choosing the answers.

REFERENCES

- Allen.M., (2010). *Misconceptions in Primary Science*.M.Graw-Hill, Open University Press, England
- Arslan, H.O., Cigdemoglu, C., & Moseley, C., (2012). A Three-Tier Diagnostic Test to Assess Pre-Service Teachers' Misconceptions about Global Warming, Greenhouse Effect, Ozone Layer Depletion, and Acid Rain. *International Journal of Science Education*. 34(11), 1667–1686
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research*, 85(2), 89–99.

- Beth, H (2002) Development of the Astronomy Diagnostic Test. *The Astronomy Education Review*, 1(1), 47-51
- Butler, A. C., & Roediger, H. L. (2007). Testing improves long term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19(4-5), 514-527.
- Caleon, I., & Subramaniam, R., (2010). Development and Application of a Three-Tier Diagnostic Test to Assess Secondary Students' Understanding of Waves, *International Journal of Science Education*, 32(7), 939-961
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36(2), 438-448
- Chung-Chih Chen, Huann-Shyang, and Ming-Liang Lin (2003), Developing a Two-Tier Diagnostic Instrument to Assess High School Students' Understanding – The Formation of Images by a Plane Mirror, *Procedia National Science Council, ROC(D)*, 12(3), 106-121
- Clement, J., (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, 50(2), 66-71.
- Dindar, Cetin.A., and Geban.O., (2011). Development of a three-tier test to assess high school students' understanding of acids and bases. *Procedia Social and Behavioral Science*, 15(1), 600-604.
- Emine, C., (2015). Effect of Two-tier Diagnostic Tests on Promoting Learners' Conceptual Understanding of Variables in Conducting Scientific Experiments. *Applied Measurement in Education*, 28(2), 253-273
- Engelhardt.P.V and Beichner.R.J., (2004). Students' Understanding of Direct Current Resistive Electrical Circuits. *American Journal of Physics*. 72(1), 99-115
- Fariyani, Q., Rusilawati, A., & Sugianto (2015). Pengembangan four-tier diagnostik test untuk mengungkapkan msikonsepsi fisika siswa SMA kelas X. *Journal of Innovative Science Education*, 4(2), 41-49.
- Fischhoff B, Slovic.P, and Lichtenstein.S., (1977). Knowing with Certainty: The Appropriateness of Extreme Confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4), 552-564.
- Griffard, B. P., & Wandersee, J.H., (2010). The two-tier instrument on photosynthesis: What does it diagnose?. *International Journal of Science Education, International Journal Science of Science Education*, 23(10), 1039- 1052
- Gurcaya, D. & Gulbasa, E., (2015) Development of three-tier heat, temperature and internal energy diagnostic test. *Research in Science & Technological Education*, 5(2), 1-21.
- Hakim.A, Liliasari and Kadarohman.A., (2012). Student Concept Understanding of Natural Products Chemistry in Primary and Secondary Metabolites Using the Data Collecting Technique of Modified CRI. *International Online Journal of Educational Sciences*, 4(3), 544-553.
- Halim, A., (2008). *Pembinaan Ujian Diagnostik dan Modul Pembelajaran Mandiri Fizik Kuantum Kearah Mengatasi Miskonsepsi Siswa*, Thesis Ph.D. Fakultas Pendidikan, UKM.
- Halim.A, T.Subahan.B.M, and Lilia.H., (2009). Development and Implementation of Diagnostic Test to Identification of Misconception in Quantum Physics, *Sains Malaysiana*, 38(2), 126-130.
- Halim.A., dan Melvina (2014g). Misconception on Force Concept; action-reaction pair is same with cause-effect sequence. *Proceeding of THE 4th INTERNATIONAL CONFERENCE ON THEORETICAL AND APPLIED PHYSICS (ICTAP-2014)* 16-17 October 2014, Denpasar-Bali, Indonesia.
- Halloun.I.A., and Hestenes.D., (1985a). Common Sense Concepts about Motion. *American Journal of Physics*, 53(11), 1056-1065.
- Hassan.S, Bagayoko.D, and Kelley.E.L. (1999). Misconceptions and the Certainty of Response Index (CRI). *Physic Education*, 34(5), 294 – 299.
- Hestenes.D and Wells.M., (1992), A Mechanics Baseline Test, Published in: *The Physics Teacher*, 30(2), 159-166.

- Karunia.P.P dan Rinaningsih., (2013). Pengembangan Tes Diagnostik Materi Teori Mekanika Kuantum dan Ikatan Kimia. *Unesa Journal of Chemical Education*, 2(2), 159-172.
- Kirbulut, Z.D., (2014). Using Three-Tier Diagnostic Test to Assess Students' Misconceptions of States of Matter *Eurasia Journal of Mathematics, Science & Technology Education*, 10(5), 509-521
- Marsh, E. J., Roediger, H. L. Bjork, R. A., & Bjork, E. L. (2010). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, 14(3), 194–199
- Martin J.K., Mitchell J, and Newell T., (2004), A Fluid Mechanics Concept Inventory (FMCI), 34th ASEE/IEEE Frontiers in Education Conference, October 20 – 23, 2004, Savannah, GA
- Masril dan Nur Asma.(2002). Pengungkapan Miskonsepsi Pelajar Menggunakan Force Concept Inventory dan Certainty of Response Index. *Jurnal Fisika Indonesia*, HFI, B(5), 1–7.
- Phelps, R. P. (2012). The effect of testing on student achievement, 1910–2010. *International Journal of Testing*, 12(2), 21–43.
- Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 1155–1159.
- Sahin, C., & Cepni, S., (2011). Development of a Two Tiered Test for Determining Differentiation in Conceptual Structure related to “Floating-Sinking, Buoyancy and Pressure” Concepts. *Journal of Turkish Science Education*, 8(1), 111-118.
- Shaffer, P.S., & McDermott, L.C., (2005). A research-based approach to improving student understanding of the vector nature of kinematical concepts. *American Journal Physics* 73(10), 921-931
- Tamir, P. (1989). Some issues related to the use of justifications to multiple-choice answers. *Journal of Biological Education*, 23(3), 285–292
- Treagust, D.F.and Haslam, F. (1987). Diagnosing secondary student' misconceptions of photosynthesis and respiration in plants using a two-tier multiple choice instrument. *Journal of Biological Education*, 21(3), 203-211.
- Viennot,L., (1979). Spontaneous Reasoning in Elementary Dynamics, *European Journal of Science Education*, 2(1), 205-221.
- Widarti, H.R., Permanasari A., & Mulyani, S., (2016). Student misconception on redox titration (A challenge on the course implementation through cognitive dissonance based on the multiple representation). *Jurnal Pendidikan IPA Indonesia*, 5(1), 56-62.
- Wijaya, C.P., Koes, H.S., & Muhardjito (2016). The diagnosis of senior high school class X MIAB Students Misconception about Hydrostatic pressure concept using three-tier. *Jurnal Pendidikan IPA Indonesia*, 5(1), 14-21.
- Wuttiptom,S, M.D. Sharma ,I.D. Johnston, R. Chitaree, and C. Soankwan (2006). Preliminary Results from a New Quantum Mechanics Conceptual Survey, *River Physics*, Australian Institute of Physics 17th National Congress 2006 – Brisbane, 3-8 December 2006.