OPEN ACCESS

DINAMIKA Jurnal Ilmiah Pendidikan Dasar, 2019, 11(2)
ISSN:2655-870X (online)
https://doi.org/10.30595/Dinamika/v11i2.5053

Developing and Validating of The Three Tier Diagnostic Test Based 'Higher Order Thinking Skills' Instrument

Ni Wayan Novita Sari[1*], Sunyono[1], Abdurrahman[1]

[1] Universitas Lampung, INDONESIA

Received 5 August 2019 ▪ Revised 2 September 2019 ▪ Accepted 30 September 2019

ABSTRACT

This study developed a knowledge assessment instrument of three tier diagnostic tests based on higher order thinking skills (HOTS) which aims to diagnose students' understanding in higher-order thinking and reduce misconceptions. The research method uses ADDIE (Analysis-Design-Develop-Implement-Evaluate) and the subject is class VII junior high school students with 2017 odd semester curriculum materials. This study used HOTS-based knowledge assessment instruments of three tier diagnostic tests of 20 items. The results of the instrument expert validation test based on aspects of content, construct, and language assessment have an average score of 90, 88, and 89 with a proper assessment. The data were analyzed using rasch model and the item reliability index was 0.89, the reliability index of the respondents was 0.91, and Cronbach's alpha value was 0.93, which meant the reliability of the questions obtained was good. The results of the students 'answer patterns are six criteria: understanding concepts, misconceptions (+), misconceptions (-), misconceptions, guessing / lack of confidence / luck, and not understanding the concepts used to diagnose learners' understanding in measuring high-level thinking skills (HOTS) which is integrated with the dimensions of factual, conceptual and procedural knowledge. Misconceptions that occur in students have three criteria, namely misconception (+), misconception (-), and misconception, and the most experienced by students is a misconception (-) in the C4 cognitive level which is integrated with the conceptual knowledge dimension on the substance characteristic material . This misconception (-) can be reduced through the correct concept with the right reasons.

Keywords: three tier diagnostic, higher order thinking skills

INTRODUCTION

The thinking ability of Indonesian students can be illustrated from the results of the PISA 2015 test and evaluation [1]. The average score of achievement of Indonesian students for science is 403 which is ranked 62 out of 69 countries evaluated [2]. This shows that the performance of Indonesian students' thinking is still relatively low, as well as evidence that Indonesian students are still dominant in low levels, or more in their ability to memorize in science learning [3]. The questions contained in PISA very demanding reasoning and problem solving skills. One of the factors is the lack of training of students in working on questions that suggest reasoning and problem solving abilities [4]. Besides that, the characteristics of PISA questions also require students' analytical, evaluation and creative thinking skills [5].

The ability to think of analysis, evaluation, and creative is an indicator of high-level thinking skills or higher order thinking skills (HOTS) [6-8]. Higher order thinking skills (HOTS) is the widespread use of the mind to find new challenges [9]. HOTS 'characteristics according to Zohar include critical thinking and creative thinking [4]. One of HOTS students' abilities is low because in evaluating learning there are still those who use questions about low level or memorizing ability [10].

niwayannovitasari04@gmail.com **(*Correspondence)**

The learning mastery of student competence in working on high-level questions is certainly inseparable from assessment activities [12]. The main principle in assessing, not only used for something known to students but also used to assess something that can be done by students in learning [13]. Assessment is the most important part of the overall learning process, one of the main functions of the assessment is to open up the misconceptions that students have before learning [13].

Misconception is a mental representation of a concept that is incompatible with the scientific theory currently held [14]. Incorrect understanding of concepts results in students experiencing misconceptions. If the misconception that occurs in students is left, then it will result in increasing material that is not able to be fully understood and will ultimately affect student learning outcomes [15]. One form of assessment that can be used to assess students' misconceptions is diagnostic multiple choice tests.

One of the multiple choice tests that can assess students' understanding and level of thinking, namely the diagnotic three tier test. There are several studies and developments regarding the three tier diagnotic tests for various purposes including Pesman & Eryilmaz [16] and Bunawan [17]. This three tier diagnostic test is a diagnostic test composed of three levels of questions [19]. The first level of the problem (one-tier) is in the form of ordinary multiple choice, the second level of the question (two-tier) that is in the form of choice of reason, and the level of the third question (three-tier) which is a question of confirmation of the answers has been chosen at the level of questions one and two [20]. Through the above explanation, researchers are interested in conducting development research entitled "Development of Knowledge Assessment Instruments Three-Tier Diagnotic Test Based on higher order thinking skills".

MATERIAL AND METHODS

Methods

This research and development uses the optional ADDIE instructional design model (Analysis-Design-Develop-Implement-Evaluate). The flow of research and development activities is carried out in 5 stages, namely (1) analyzing needs, (2) designing products, (3) developing products, (4) implementing products and (5) evaluating products. Data on research and development was obtained through questionnaires and validation sheets.

Instrument

The conception test instrument that was developed to find out the conception of preservice elementary school teacher students was 7 items. Test items used the four-tier test format. The conception test validation was judged by two physics education experts and one theoretical physicist. The results of the validation of the conception test items indicated that the validator generally stated that the items to be used are valid both in content and construction. Validators provided some notes that are suggested to be revised, especially those that are considered inappropriate, that is, related to accepted scientific theory, image clarity, sentence editorial, and written order. The research instrument was in the form of several multiple-choice items on the concept of force and motion that require explanations about the selected answers. To find out the participants' *conceptual change*, they wrote their responses towards several written statements (Yes/No/Not always) and with confidence levels (Sure/Not really sure/Not sure). The participants answered the questions individually and do not include their names to ensure their anonymity.

Procedures

In the preliminary study, Direct Instruction and discussion methods were used.

Data Analysis

The technical data analysis uses the Rasch model and is assisted by Winsteps 3.73 software developed by Linacre [28]. Parameters used to determine the accuracy or suitability of respondents and items according to Boone in Sumintono & Widhiarso [27] include means-square outfit value, z-standard outfit, and measurre correlation points. The level of difficulty of the item (item measure) can be known from the logit value of each item. Reliability testing uses cronbach alpha formula in **Table 1**

Table 1. Interpretation of Measures of Stability of Alpha Values

Nilai *Alpha Cronbach's*	Keterangan
< 0,5	Poor
0.5-0.6	Ugly
0.6-0.7	Enough
0.7-.0.8	Nice
> 0.8	Very good

Grouping questions and items can be seen from the separation value . then the quality of the question instruments in terms of the overall respondents and items is getting better because it can identify groups of respondents and items in various ways.

RESULTS AND DISCUSSION

Developing and Validating

This development research produces a valid and reliable question instrument for measuring students' high-level thinking skills and student misconceptions. The instrument of the developed question is a question consisting of three levels. The first level is multiple choice questions, the second level is the reason, and the third level is the belief in the answers at the first and second levels. The material used is measurement, classification of living things, energy, and heat. The results of validation data obtained from three validators in the form of quantitative data using Likert scale scores are presented in **Table 2**

Table 2. Results of Overall Data on Expert Validation

No	Assessment Aspect	expert validation			average value	Category
		1	2	3		
1	Isi	88	92	92	90	Very valid (small revision)
2	construct	85	90	90	88	Very valid (small revision)
3	Language	86	91	91	89	Very valid (small revision)

The results of the validation data in Table 2, show that the assessment by the validator regarding the assessment of the instrument as a whole is feasible to use with small revisions. Readability tests are carried out after the test instrument is feasible to use. The instrument readability test results can be seen in Table 3.

Table 3. Results of the question instrument readability test

No	Respondent	Category	Value
1.	Student A	96	Very Understanding
2.	Student B	92	Very Understanding
3.	Student C	92	Very Understanding
4.	Student D	96	Very Understanding
5.	Student E	96	Very Understanding

Table 3 shows the test instrument has clear readability so the language used in the items is very easy to understand and does not experience multiple interpretations. These questions were tested on 26 respondents . The scoring used is the Graded Response Model (GRM) because the generated data is in the form of politomus. Scoring guides from the answers to the questions done by students can be seen in Table 4.

Table 4. Guidelines for Scoring

Stage (1)	Stage (2)	Stage (3)	Score	Category
True	True	Sure	7	Understand the concept
True	False	Sure	6	Misconception (+)
False	True	Sure	5	Conception (-)
False	False	Sure	4	Misconception
True	True	Not Sure	3	Guess, lack of confidence
True	False	Not Sure	2	Don't understand the concept
False	True	Not Sure	1	Don't understand the concept
False	False	Not Sure	0	Don't understand the concept

The three tier knowledge assessment instrument diagotic test that was tested on 26 respondents obtained the results of validity and reliability test. The results of the final analysis of the development of a HOTS knowledge-based three-tier diagnostic test instrument consisting of 20 items are presented in **Table 5.**

Table 5. Summary of Final Analysis Results

Analysis	Output	Result
Respondent	Reliability	0.91
	Separation index	3.19
	Separation of strata	4.59
	Lowest logit value	-0.79 (04L)
	The highest logit value	1.94 (08L)
Item of question	Reliability	0.89
	Separation index	2.83
	Separation of strata	4.11
	Lowest logit value	-0.78(S1)
	The highest logit value	0.96 (S20)
Instrument		0.93

The results of product evaluations carried out for science teachers of Junior High School (SMP) In MGMP can be seen in Table 6.

Table 6. Product Evaluation Test Results

No	Indicator	Response
1	Are the items developed that are clear and easy to understand?	Yes
2	Does the formulation of the question not use sentences or words that give rise to multiple perceptions or multiple interpretations?	Yes
3	Can the images, numbers, letters and symbols used in the question be clearly read?	Yes
4	Are the terms of science used in the problem is the term familiar to you?	Yes
5	whether the question uses language that is easily understood by students?	Yes
6	Can the questions presented diagnose student understanding?	Yes

The results of product evaluation based on Table 6 state that the product is good to use. Beside that, there are open questions about the product being developed related to the advantages and disadvantages of the product being developed.

The needs analysis phase carried out in this development research was used to illustrate the reasons for the development of HOTS-based three-tier diagnostic test instruments. The types of questions given in evaluating learning at school still do not fully train students' high-level thinking skills. It is necessary to innovate to meet the criteria of the 21st century, where each person must have a high level of thinking skills in order to be able to become the next generation capable and skilled in their fields. There are several discussions in this development research as follows :

Item Parameters Question Instrument Knowledge Assessment Three Tier Diagnostic Test Based on HOTS

The knowledge assessment instrument of the three tier diagnostic test has been tested on a limited number of 27 students from SMP NEGERI 1 SIDOMULYO. This activity was carried out on October 10, 2018. The criteria for the item developed are shown in Table 7.

Table 7. Criteria for Research Item

Knowledge Dimension	Science Physics			Science Chemistry			Science Biology			Total
	C_4	C_5	C_6	C_4	C_5	C_6	C_4	C_5	C_6	
PF	S2	S4, S5	-	S8	-	-	S14	-	S21	6
PK	S1	S3	S7	S9	-	S12	-	S16, S17	S20	8
PP	-	S6	-	S10	S11	S13	S15	S18	-	6
Total	2	4	1	3	1	2	2	3	2	20

Based on Table 7, the results show that there are 6 factual knowledge questions, 8 conceptual knowledge questions, and 6 procedural knowledge questions that are integrated in HOTS cognitive (C4, C5, and C6). The material used, namely measurement, classification of living things, energy, and heat. The results of the trial on the

limited test obtained the results of data analysis using Winstep 3.73 which is 20 questions declared valid. This is because all items have met the three criteria set by Boone [21], so it is concluded that all questions are valid and all questions are fit. All fit questions mean that all items about normal functioning in measuring and there is no misconception in students about the questions. The item difficulty index is negative and positive, so that all items can vary in either category [22].

In addition, the results of item analysis of test questions in the form of three tier diagnostic tests have a reliability value of 0.93 which is included in the special category. Standard assessment criteria (reliability, validity, and rating scale) HOTS test questions developed have met good standards. Questions are reliable because the items developed have value person reliability in excellent categories (respondents answer the whole item seriously and do not carelessly) or have good criteria [7]. The factors that influence the value of alpha chonbrach, that is the items developed according to the procedure, using the right reference, through the content validation stage, and empirically tested with respondents who worked seriously and supervised carefully [23].

The results of the analysis of the separation index have very large analysis results in terms of overall respondents and items. Separation index values in terms of respondents and items have a value of 4 which indicates that heterogeneous respondents' grouping and items carefully. All questions have a valid option because the separation index value is more than 0.7 which indicates that the question has adequate consistency [19], so the respondent is able to understand the choices given at each level of difficulty in the item.

The item criteria are good because they have different power about good questions too. Different power of questions in research and development is known as Rasch discrimination power or grain score correlation and Rasch score (Pt Measure Corr). If the value of Pt. Measure Corr is positive, then the item tested is in accordance with the construct, whereas if the negative sign is in the value of Pt Measure Corr, then all items are misleading [24]. Value of Pt. Measure Corr is good (positive), its means that all students who take the ability test are low because they answer items that are wrong and have high ability because they answer the item correctly. This is because students with tests have high ability to answer incorrectly. Questions with negative correlation values must be examined to see whether the answer key is wrong, needs to be revised or removed from the test [25].

The result of Pt Measure Corr analysis in this study shows a positive value which means there are no questions that are misleading so the question does not need to be revised or deleted. This is consistent with the research conducted by Hayati & Lailatussaadah [11], that all items have a positive.

Pt Measure Corr value which means there is no conflict between items and constructs measured. The level of difficulty of the problem can be seen from its distribution through wright maps, while the level of students' ability to answer questions can be seen from the level of ability. The level of ability of students is given sequentially from the lowest to the highest based on the logit value of each person in the measure column.

Knowledge assessment instrument for three tier diagnostic tests based HOTS can diagnose student understanding

The diagnosis of students 'understanding based on the interpretation of the results of the students' answers to the three tier diagnostic test using the reference proposed by Arslan [26]. If the pattern of answers of students at the first, second, and third levels in a row is correct, correct, and sure, then the students are categorized as understanding the concept. Furthermore, if the pattern of students' answers at the first, second, and third levels are right, wrong, and sure, the students are categorized as misconceptions (+). The answer pattern of students at the first, second, and third levels in a row is wrong, correct, and sure, so the students are categorized as misconceptions (-). The cause of this misconception (-) occurs, one of which is the carelessness that students make in choosing answers [27].

The pattern of answers for students at the first, second, and third levels in a row is wrong, wrong, and sure, so the students are categorized as having misconceptions. Misconception shows a condition where students express a concept that is different from what was suggested by experts. The pattern of answers for students at the first, second, and third levels in a row is correct, correct, and unsure, so the students are categorized as guessing / lacking in confidence. Then, if the pattern of answers of students at the first, second, and third levels in a row is right / wrong, right / wrong, and not sure, then the students are categorized as not understanding the concept. Students in this condition can be said that they have an understanding mixed with misconceptions [28]. Based on the student's answer pattern, the teacher can diagnose students' understanding through the knowledge assessment instrument of three-tier diagnostic tests based on HOTS that have been developed.

The development of three-tier diagnostic assessment instruments based on HOTS has the advantaged of being able to diagnose student understanding and can improve students' high-level thinking skills (HOTS) integrated with the knowledge dimension. Although the research on developing the three-tier diagnostic assessment instrument already exists, researchers have integrated HOTS indicators (analyzing, evaluating, and creating) and the knowledge dimensions (factual, conceptual, and procedural) in this developed assessment instrument. Although the research on developing the three-tier diagnostic assessment instrument already exists, researchers have integrated HOTS indicators (analyzing, evaluating, and creating) and the knowledge (factual, conceptual, and procedural) dimensions in this developed assessment instrument. This is because some HOTS instrument development only focuses on measuring the ability of the knowledge dimension alone or developing a three tier diagnostic multiple choice test that only focuses on diagnosing student understanding (such as the development of instruments carried out by Mubarak [18]). Unlike the products that have been developed

previously, the products developed in this development research are HOTS-based three-tier diagnostic test knowledge instruments, where these products are not only used to diagnose student understanding, but are also used to improve students' higher-order thinking skills that are integrated with the knowledge dimension in this development product

Knowledge assessment instruments of three-tier diagnostic tests based on HOTS can reduce student misconceptions

The results of diagnostic tests on students are used as a basis for evaluation in reducing student misconceptions. Misconception is a mental representation of a concept that is not in accordance with the scientific theory currently held [14]. Misconceptions that occur in students are three criteria, namely misconception (+), misconception (-), and misconception. Items that experience misconceptions can be seen in Table 8.

Table 8. Analysis of Problems Experiencing Misconception

Number Item	Knowledge Dimension	Level Cognitive	Material	Misconception Criteria	Total
S9	PK	C_4	Characteristics of substances	Misconception (+)	10
S10	PP	C_4	Characteristics of substances	Misconception (+)	8
S12	PK	C_6	Characteristics of substances	Misconception (-)	11
S11	PP	C_5	Characteristics of substances	Misconception (-)	10
S14	PF	C_4	Classification of living things	Misconception	7
S16	PK	C_5	Classification of living things	Misconception	7

The results of the analysis of the questions that experienced misconceptions based on Table 8 found that as many as 10 students experienced misconceptions (+) in question number 9 and as many as 8 students experienced misconceptions (+) also in question number 10. A total of 11 students experienced misconceptions (-) in question number 12 and as many as 10 students experienced a misconception (-) in question number 11. In addition, questions with number 14 and 16 experienced misconceptions with the same number of students as many as 7 students.

This misconception (+) can be overcome by reason of the answers chosen by the students themselves. The knowledge assessment instrument of three-tier diagnostic test based on HOTS can reduce misconceptions in students by knowing the right reasons for the answers that choose correctly [28]. Misconception (-) occurs because of an analogy or reason that can reinforce the concept of wrong. This can be overcome by knowing the correct answer or the correct concept based on the exact reasons of the correct answer [25] (in this case it can be seen from one-tier and two-tier). In contrast to misconceptions (+) and misconceptions (-), misconceptions can be seen from the pattern of incorrect answers in one tier and two-tier, but in the three-tier the answer is sure [16].

This instrument can reduce well based on the correct answers at the first and second levels, so students can find out the misconceptions their experience. In addition, students can change the wrong concept to be correct through this HOTS-based knowledge assessment instrument of three HOTS. Based on this, it can be said that the HOTS-based knowledge assessment instrument of three tier diagnostic tests caould reduced student misconceptions

CONCLUSION

The conclusion in this development research is that the HOTS parameters of the three tier diagnostic test items have good quality. This instrument can diagnose student understanding through the student's own answer pattern, and can reduce misconceptions that occur in students.

ACKNOWLEDGMENTS

Acknowledgments are addressed to the supervisors in the research development of this knowledge assessment instrument of three HOTS-based diagnostic tests and to all parties involved in this development research.

REFERENCES

[1] OECD. (2016). *PISA 2015 Results in Focus*. [Online]. Tersedia: http://www.oecd. org/pisa/keyfindings/pisa-2015-results-overview.pdf.

[2] Hazrul, I. 2016. *html*. Retrieved 1 15, (2018), from http://www.ubaya.ac.id/2014 /content/ articles_detail/230/Sekelumit-Dari-Hasil-PISA-2015-Yang-Baru-Dirilis

[3] Anderson, J. O., Huann-Shyang Lin, D. F. Treagust, S.P. Ross, & L. D. Yore. (2007). Using Large-Scale Assessment Datasets For Research In Science and Mathematics Education : Programme For International Student Assessment (PISA). *International Journal of Science and Mathematics Education*. Taiwan. 5, 591-614

[4] Zohar, A. & Y.J. Dori. (2003). Higher Order Thinking Skills and Low-Achieving Students: Are They Mutually Exclusive?. *The Journal of The Learning Sciences*, 12(2), 145-181

[5] Ramadhan, D & Wasis. (2013). Analisis Perbandingan Level Kognitif dan Keterampilan Proses Sains dalam Standar Isi (SI), Soal Ujian Nasional (UN), dan Soal Programme For International Student Assessment (PISA). *Jurnal Inovasi Pendidikan Fisika*. 2 (1), 20-25

[6] Canas, A. J., P. Reiska, & A. Möllits. (2017). Developing Higher-Order Thinking Skills With Concept Mapping: A Case of Pedagogic Frailty. *Knowledge Management & E-Learning: An International Journal (KM&EL)*. 9 (3), 348-365

[7] Tajudin, N.M. & M. Chinnappan. (2016). The Link between Higher Order Thinking Skills, Representation and Concepts in Enhancing TIMSS Tasks. *International Journal of Instruction*. 9(2), 199-214

[8] Dubas, J.M. & S.A. Toledo. 2016. Taking Higher Order Thinking Seriously: Using Marzano's Taxonomy In The Economics Classroom. *International Review of Economics Education*, 12-20

[9] Heong, Y. M.,Othman, W.D.,Md Yunos, J., Kiong, T.T., Hassan, R., & Mohamad, M.M. (2011). The Level of Marzano Higher Order Thinking Skills Among Technical Education Students . *International Journal of Social and humanity,* 1 (2), 121-125

[10] Kertayasa, I. K. (2018). *Pengembangan Soal Model PISA Berbasis Online*. Retrieved from www.indonesiapisacenter.com/2014/03/tentang-website.html.

[11] Hayati, S. & Lailatussaadah. (2016). Validitas dan Reliabilitas Instrumen Pengetahuan Pembelajaran Aktif, Kreatif, dan Menyenangkan (PAKEM) Menggunakan Model Rasch. *Jurnal Ilmiah Didaktika*, Aceh, 16 (2), 169-179

[12] Saldo, G.M., S. Siraj, A.B.B. Nordin, & O. S. Al_Amedy. (2015). Higher Order Thinking Skills Among Secondary School Students in Science Learning. *The Malaysian Online Journal of Educational ScienceI, MalaysiaI*. 3(3)

[13] Rosidin, U. 2016. *Penilaian Otentik (Authentic Assessment).* Yogyakarta: Media Akademia

[14] Sunyono, Tania, L., & Saputra, A. (2016). A Learning Exercise Using Simple and Real-Time Vizualization Tool To Counter Misconceptions About Orbitals and Quantum Numbers. *Journal Of Baltic Science Education,* 15 (4), 453-463.

[15] Putra, I. E., Adlim, & Halim, A. (2016). Analisis Miskonsepsi dan Upaya Remediasi Pembelajaran. *Jurnal Pendidikan Sains Indonesia,* 4 (2), 13-19 Lailly, N. R., & Wisudawati, A. W. 2015. Analisis Soal Tipe Higher Order Thinking Skills (HOTS) dalam Soal UN Kimia SMA Rayon B Tahun 2012/2013. *Kaunia,* 11 (1), 27-39.

[16] Pesman, H., & Eryilmaz, A. (2010). Development of a Three-Tier Test. *The Journal of Educational Research*, 208-222.

[17] Bunawan, W., Setiawan, A., Rusli, A., & Nahadi. (2014). Pengembangan instrumen Tes Diagnostik Pilihan Ganda Tiga Tingkat untuk Mengakses Kemampuan Mahasiswa Calon Guru Fisika. *EDUSAINS,* 6 (2), 138-144.

[18] Mubarak, S., E. Susilaningsing, & E. Cahyono. (2016). Pengembangan Tes Diagnostik Three Tier Multiple Choice untuk Mengidentifikasi Miskonsepsi Siswa Kelas XI, *Journal of Innovative Science Education (JISE),* 5 (2), 101-110

[19] Misajon, RA., J. Pallant, & A-M. Bliuc. (2016). Rasch Analysis of the Personal Wellbeing Index. *Springer International Publishing Switzerland*, Australia. 25, 2565-2569

[20] Aydeniz, M., Bilican, K., & Kirbulut, Z. D. (2017). Exploring Pre-Service Elementary Science Teachers' Conceptual. *International Journal of Education in Mathematics, Science and Technology,* 5 (3), 221-234.

[21] Boone, W. J., Staver, J. R., dan Yale, M. S. (2014). *Rasch Analysis in the Human Science*. Dordrecht: Springer.

[22] Cowlishaw, S., S.S. Merkouris, N.A. Dowling, S. Rodda, A. Suomi, & S.L Thomas. (2019). Locating gambling problems across a continuum of severity: Rasch analysis of the Quinte Longitudinal Study (QLS). *Elsevier, Journal Homepage: www.elsevier.com/locate /addictbeh*, 92, 32-37

[23] Istiyono, E., D. Mardapi, & Suparno. 2014. Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PysTHOTS) Siswa SMA. *Jurnal Penelitian dan Evaluasi Pendidikan*. *Indonesia*,18 (1)

[24] Mohamad, M.M., N.L. Sulaiman, L.C. Sern, & K.M. Salleh. (2015). Measuring the Validity and Reliability of Research Instruments. *Procedia Social and Behavioral Sciences*, Malaysia. 204, 164-171

[25] Smiley, J. (2015). Classical Test Theory or Rasch: A Personal Account From A Novice User. *SHIKEN*. 1 (2), 16-31.Saptono, S., Rustaman, N., & Wibowo, S. D. 2016. Memfasilitasi Higerh Order Thinking Skills dalam Perkuliahan Biologi Sel Melalui Metode Integrasi Atribut Asesmen Formatif. *Unnes Science Education Journal*, 5 (3)

[26] Arslan, H. O., Cigdemoglu, C. & Moseley, C. (2012). A Three-Tier Diagnostic Test to Assess PreService Teachers' Misconceptions about Global Warming, Greenhouse Effect, Ozone Layer Depletion, and Acid Rain. *International Journal of Science Education*. 34 (11), 1667–1686.

[27] Syahrul, D. A. & Setyarsih, W. (2015). Identifikasi Miskonsepsi dan Penyebab Miskonsepsi Siswa dengan Three Tier Diagnostic Test Pada Materi Dinamika Rotasi. *Jurnal Inovasi Pendidikan Fisika*. 4. (03), 67-70 Sumintono, B., dan Widhiarso, W. 2014. *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*. Cimahi: Trimkomunikata.

[28] Abraham, M. R. Grzybowski, E. B., Renner, J. W. & Marek, E. A. (1992). Understanding and Misunderstanding of Eigth Grader of Five Chemistry Concept Found in Textbook. *Journal of Research in Science Teaching*. 29 (2), 105-120