JURNAL PEMBELAJARAN FISIKA

http://jurnal.fkip.unila.ac.id/index.php/JPF

Vol 9 (2), 2021, 129-141 ISSN: 2302-0105 (p); 2684-9828 (e)

Development of HOTS Assessment Instruments on Newton's Law of Motion by Using Multiple Types of Problems in High School Physics Lessons

Novinta Nurul Sari *, Undang Rosidin, Eneng Vera Dwi Indriyani

University of Lampung, Bandar Lampung, Indonesia

** e-mail: novinta.nurulsari@fkip.unila.ac.id*

Received: December 27, 2021 *Accepted: December 30, 2021* *Published: December 30, 2021*

Abstract: This study aims to determine the criteria for questions with the type of matchmaking, reasoned plural choice, and causality that can increase HOTS on Newton's Law material about high school physics motion. This research was conducted at MAN 2 Tangerang by using class X, which consisted of two classes. The product developed from this research is the Higher Order Thinking Skills (HOTS) test instrument using Research and Development. The test instruments used were matchmaking, reasoned plural choice, and causality on Newton's Laws of motion in high school physics lessons. Data collection techniques in this development use a questionnaire and Newton's Law of Motion test. The data obtained in the study were analyzed using the Rasch model with the help of Ministep 4.5.1 software. The results showed that the test instrument for higher order thinking skills on Newton's Law of motion material uses information, pictures, graphics, or data to solve problems, describe information in detail, relate one concept to another, and look for links from different information. can train students' higher order thinking skills.

Keywords: HOTS, Rasch Model, Newton's Laws of Motion

DOI: http://dx.doi.org/10.23960/jpf.v9.n2.202101

INTRODUCTION

Trianto (2013: 1) explains that education that is able to support future development is education that is able to develop the potential of students, so that those concerned are able to face and solve life problems they face. Currently, schools in Indonesia are already using the revised 2013 curriculum. Bima (2019) states that the 2013 curriculum has four aspects of assessment, namely spiritual aspects (KI-1), social aspects (KI-2), knowledge aspects (KI-3), and skills aspects (KI4). The implementation of the 2013 curriculum is carried out as an effort to harmonize education with the times, the demands of technological progress and the abilities of students, and refers to the abilities needed in the 21st century based on Higher Order Thinking Skills (HOTS). Bialik, Bogan, & Fadel (2015) explain that the abilities that students must possess in the 21st century are creativity, critical thinking, communication, and collaboration where these four aspects are part of HOTS.

Measuring students' HOTS abilities requires question instruments and question assessment instruments which can be developed every year according to the demands of the times and curriculum by researchers. Physics learning does not escape from working on assignments in the form of questions. Based on the results of interviews with 5 teachers and 15 students at MAN 2 Tangerang, it was found that teachers still rarely gave HOTS questions on Physics, Newton's Laws of Motion. This causes students to still have difficulty working on HOTS physics questions. Working on physics questions requires an instrument to measure the ability of students. The development of instruments to measure students' HOTS by teachers is important in learning in this competitive 21st century (Retnawati, et al. 2018). The lack of space for students in developing HOTS is due to the lack of understanding of educators regarding the HOTS instrument (Heru & Suparno, 2019; Retnawati, et al, 2018). HOTS measurement requires an instrument that not only measures the ability of students, but is also able to train students' HOTS (Hamdi, Suganda, & Hayati, 2018). The use of the right question instrument in knowledge competence is very dependent on the behavior to be measured.

Previous development studies used several types of questions, such as a combination of description and multiple choice (Sinaga, 2018), multiple choice (Wibowo & Cholifah, 2018), a combination of multiple choice and constructed response test or multiple choice reasoned (Ku, 2009). The type of multiple choice questions is often used because measurements related to objectivity will be easier in data collection and data management. However, the type of multiple choice questions is not good, because it cannot reveal the ability of students to reason, give reasons, and synthesize problems (Putri, Istiyono, & Nurcahyanto, 2016). Each question has its own advantages and disadvantages.

METHOD

Research Design & Procedures

The product developed from this research is the Higher Order Thinking Skills (HOTS) test instrument using Research and Development. The test instruments used were matchmaking, reasoned plural choice, and causality on Newton's Laws of motion

in high school physics lessons. The purpose of developing the HOTS test instrument is to measure students' higher order thinking skills.

The product developed is based on the Borg & Gall (1983) development model which consists of 10 development steps. In this development research using only 7 steps, namely (1) research and information gathering, (2) planning, (3) initial product development, (4) limited trial, (5) initial product revision, (6) field trial , and (7) final product revision.

Population and Sample

Research on the development of test instruments analyzed using the Rasch Model was carried out at MAN 2 Tangerang Regency. The population in this study, namely students of class X MIA MAN 2 Tangerang in the 2020/2021 school year. The sample of this study was taken randomly using one class which was used as the experimental class.

Data Collection and Instrument

Data collection techniques in this development using questionnaires and tests. The questionnaire method was used to obtain information on needs analysis addressed to class X students at MAN Tangerang. The questionnaire contains questions about the completeness of school facilities, the use of school facilities in learning, the use of learning media, the effectiveness of learning methods, and learning resources used, as well as the difficulties faced by students in the material developed. This questionnaire method is also used to measure the validity of product development, including construct expert test and content expert test. The revised test instrument was then piloted to class X MIA MAN 2 Tangerang in the even semester of the 2020//2021 academic year. The data from the test results were to determine the validity, reliability, higher-order thinking skills of students, and the level of individual suitability. The test questions were used by researchers to determine their effect on HOTS.

Data Analysis

The results of trials conducted by 53 students were then analyzed using the Rasch model with the help of Ministep 4.5.1 software by means of the data that has been obtained, tabulated in the Ms. software. Excel is then converted and analyzed with the help of Ministep 4.5.1 software in the Windows 10 operating system. The data analysis was carried out to determine the validity, reliability, and difficulty level of the questions, as well as to determine the level of students' higher-order thinking skills.

Methods can be written in sub-sections, with sub-subheading. Subtitles do not need to be given a notation, but are written in lowercase letters beginning with a capital letter, Times New Roman-12 unbold, left flat. For example, you can see the following.

RESULT AND DISCUSSION

In the validity analysis, the researcher identified according to the type of questions or items, which amounted to 9 questions with the type of reasoned multiple choice, 8 questions with the type of cause and effect, and 8 questions with the type of multiple reasoned reasoning. The validity or level of suitability of the items (item fit) is used to identify whether the items can function normally in measuring or not. If there are questions that do not fit, then there is an indication that there is a misconception among students about the question. The results of the analysis of the suitability of items for the type of reasoned plural choice, cause and effect, and reasoned matchmaking can be seen in Table 1-2.

Table 1. Conformity Analysis of Reasoned Multiple Choice Question Types

MEASURE	OUTFIT		PT-MEASURE	ITEM
	MNSQ	ZSTD	CORR	
-.18	.95	-.14	.47	S1
-.46	1.56	2.66	.28	S2
.04	1.27	1.06	.24	S3
-1.00	1.05	.30	.41	S4
.27	.99	.05	.31	S5
-.90	.80	-.98	.62	S6
-.18	.95	-.14	.47	S7
-1.04	1.28	1.16	.43	S8
-.11	.90	-.36	.50	S9

In Table 1, almost all reasonable multiple choice questions have met the three criteria for the suitability of the items according to Boone et al. (2014), namely: (1) the outfit mean square value received is 0.5 < MNSQ < 1.5; (2) accepted OUTFIT Z-standard (ZSTD) value -2.0 < ZSTD < +2.0; (3) the value of Pt Mean Corr is accepted: 0.4 < Pt Measure Corr < 0.85. There are only a few questions that do not meet one of the criteria, such as question number 2 which has an OUTFIT MNSQ and ZSTD scores of 1.56 and 2.66, respectively; and questions number 2, 3, and 5, which have a Pt Measure Corr value of .28, respectively; .24; and 31; question number 2 does not meet the three criteria, but the other three questions, namely numbers 3 and 5 still meet the MNSQ and ZSTD criteria so questions number 3 and 5 are maintained, but number 2 is changed or replaced. Based on this, it can be said that 8 multiple choice questions are valid and do not need to be changed or replaced. 1 question number 2 needs to be changed or replaced.

Table 2. Conformity Analysis of Cause-and-Effect Types

MEASURE	OUTFIT		PT-MEASURE	ITEM
	MNSQ	ZSTD	CORR	
.06	1.07	.34	.31	S10
.53	.65	-1.06	.34	S11
.24	1.09	.39	.24	S12
-.53	1.41	2.07	.42	S13
.90	1.36	.95	.12	S14
.60	.98	.07	.20	S15
1.36	.55	-1.13	.26	S16
.39	.52	-1.74	.49	S17

In Table 2, almost all reasonable multiple choice questions have met the three criteria for the suitability of the items according to Boone et al. (2014), namely: (1) the outfit mean square value received is 0.5 < MNSQ < 1.5; (2) accepted OUTFIT Z-standard (ZSTD) value -2.0 < ZSTD < +2.0; (3) the value of Pt Mean Corr is accepted: 0.4 < Pt Measure Corr < 0.85. There are only a few questions that do not meet one of the criteria, such as question number 13 which has an Outfit ZSTD value of 2.07; and questions number 10, 11, 12, 14, 15, and 16 which have a Pt Measure Corr value of .31 respectively; .34; .24; .12; .20; .26. Question number 13 does not meet the Outfit ZSTD value criteria but meets two other criteria, namely the Outfit MNSQ value and Pt Measure Corr. Other questions, namely numbers 10, 11, 12, 14, 15, 16 still meet the MNSQ and ZSTD criteria. Questions number 10-16 can be defended because they still meet two criteria. Based on this, it can be said that the 8 matching questions are valid and do not need to be changed or replaced.

Table 3. Conformity Analysis of the Types of Reasonable Matching Questions

MEASURE	OUTFIT		PT-MEASURE	ITEM
	MNSQ	ZSTD	CORR	
.85	1.00	.27	.19	S18
.57	.76	-.64	.19	S19
.06	.67	-1.37	.54	S20
-.13	1.28	1.24	.37	S21
.08	1.35	1.30	.30	S22
.06	.69	-1.25	.52	S23
-.58	.78	-1.28	.59	S24
-.64	.79	-1.21	.67	S25

In Table 3, almost all reasonable multiple choice questions have met the three criteria for the suitability of the items according to Boone et al. (2014), namely: (1) the outfit mean square value received is 0.5 < MNSQ < 1.5; (2) accepted OUTFIT Z-

standard (ZSTD) value -2.0 < ZSTD < +2.0; (3) the value of Pt Mean Corr is accepted: 0.4 < Pt Measure Corr < 0.85. There are only a few questions that do not meet one of the criteria, such as questions number 18, 19, 21, and 22 which have a Pt Measure Corr value of .19 in a row; .19; .37; and.30. The four questions did not meet the Pt Measure Corr score criteria, but still met the MNSQ and ZSTD criteria. Questions 18, 19, 21, and 22 can be defended because they still meet two criteria. Based on this, it can be said that the 8 matching questions are valid and do not need to be changed or replaced.

Cronbach's alpha value is used to measure reliability, namely the interaction between the person (respondent) and the item (item) as a whole. The person reliability for the 25 questions that have been made can be seen in Table 3.

Table 4. Person Reability Question

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	48.3	25.0	.64	.22	1.05	-.1	1.01	-.1
P.S.D.	10.5	.0	.53	.05	.92	1.3	.84	1.2
MAX.	69	25.0	1.31	.43	7.29	4.80	6.36	4.50
MIN.	26	25.0	-.94	.18	.31	-2.41	.34	-1.57
REAL RMSE .27		TRUE SD .	45	SEPARATION 1.69		PERSON RELIABILITY .74		
S.E. Of Person MEAN = .22								

Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY= .88

In Table 4, it is known that the average value of INFIT MNSQ and OUTFIT MNSQ, respectively, is 1.05 and 1.01, meaning that the value is getting better because the value is close to the ideal, which is 1.00. The average value of INFIT ZSTD and OUTFIT ZSTD, respectively, is -.17 and -1, meaning that the quality of the person is getting better because the value is close to the ideal, which is .0. The value of person reliability is .74 which indicates that the consistency of the answers from the respondents is good, meaning that the respondents do all the questions seriously and not carelessly. The reliability items for questions with the type of reasoned plural choice, cause and effect, and reasoned matchmaking can be seen in Table 5.

Table 5. Item Reliability Question

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	102.3	53.0	.00	.15	.98	.0	1.01	.1
P.SD.	30.6	.0	.60	.04	.24	1.3	.29	1.2
MAX.	163	53.0	1.36	.29	1.29	2.06	1.56	2.66
MIN.	64	53.0	-1.04	.12	.55	-2.07	.52	-1.74
REAL RMSE .16		TRUE SD .	45	SEPARATION .58		ITEM RELIABILITY .92		
S.E. Of *Item* MEAN = .15								

In Table 5, it is known that the average value of INFIT MNSQ and OUTFIT MNSQ, respectively, is 0.98 and 1.01, meaning that the value is getting better because the value is close to the ideal, namely 1.00. The average value of INFIT ZSTD and OUTFIT ZSTD, respectively, is .0 and .1, meaning that the quality of the items is getting better because the value is close to the ideal, which is 0.0. The item reliability value is 0.92 which indicates that the quality of the items is very good, meaning that the items on the test instrument can measure what is being measured.

From the analysis using the Rasch model with the help of Ministep 4.5.1 software, it can also provide information about the level of difficulty of the questions (item measure). The level of difficulty of the questions in the test instrument is seen from the logit value of each item contained in the measure column. A high logit value indicates the highest level of problem difficulty. The logit value and Standard Deviation (SD) for each item in detail can be seen in item measures in Appendix 10-12, and for the distribution of questions, see the Wright map in Appendix 13-15. The items on the test instrument can be grouped into four levels of problem difficulty based on their logit value, for questions with the type of reasoned plural choice, cause and effect, and reasoned matching questions in Table 6-8.

Table 6. Difficulty Level of Reasoned Plural Type Questions

Measure	Category	No	Sum
> 1,03	Very hard	-	-
0 – 1,03	Hard	3, 5,	2
-1,03 - 0	Easy	1, 4, 6, 7, 9	5
< -1,03	Very easy	8	1
	Sum of question		8

Table 7. Difficulty Level of Cause and Effect Type Problem

Measure	Category	No	Sum
> 1,03	Very hard	16	1
0 – 1,03	Hard	10, 11, 12, 14, 15, 17	6
-1,03 - 0	Easy	13	1
< -1,03	Very easy	0	0
	Sum of question		8

Table 8. Difficulty Level of Reasonable Matchmaking

Measure	Category	Number	Sum
> 1,03	Very hard		0
0 – 1,03	Hard	18, 19, 20, 22, 23	5
-1,03 - 0	Easy	21, 24, 25	3
< -1,03	Very easy		0
	Sum of question		8

The test instrument for higher order thinking skills on Newton's Law of motion which consists of 24 questions and is divided into types of reasoned plural choice

questions, cause and effect, and reasoned matchmaking was tested on 53 students of class X IPA MAN 1 Tangerang. The scores obtained by students from the test results are then processed into values that can be seen in detail in Appendix 9. These values are then sorted and information about the level of students' higher-order thinking skills will be obtained as in Table 9.

Table 9. Level of Students' Higher Order Thinking Ability

Value	Category	Sum of students	Persentase (%)
100-76	Very good	22	41,5 %
75-51	Good	10	18,86 % 32,1 %
50-26	Enough	17	7,55 % 100 %
25-1	Not enough	4	
Sum		53	

(Lewi & Aisyah, 2009)

A test instrument for higher order thinking skills on Newton's Law of motion has been developed for high school students. The test instrument developed has the characteristics of higher order thinking, where the test instrument uses information, pictures, graphs, or data to solve problems, describe information in detail, relate one concept to another, and look for links from different information. This development research uses questions of the type of reasoned plural choice, reasoned cause and effect, and multiple responses which are analyzed using the Rasch Model. Previous development studies have used several types of questions, such as a combination of description and multiple choice (Sinaga, 2018), multiple choice (Wibowo & Cholifah, 2018), a combination of multiple choice and constructed response test or multiple choice reasoned (Ku, 2009). . The type of multiple choice questions is often used because measurements related to objectivity will be easier in data collection and data management. However, the type of multiple choice questions is not good, because it cannot reveal the ability of students to reason, give reasons, and synthesize problems (Putri, Istiyono, & Nurcahyanto, 2016). Each question has its own advantages and disadvantages.

There are also very few previous studies that have developed instruments using the Rasch Model. The developed questions can be measured using the Rasch Model because they have met the objective measurement. Objective measurement produces data that is free from the influence of the type of subject, the characteristics of the assessor and the characteristics of the measuring instrument. The estimation and calibration techniques used in the modeling have eliminated the influence of these three factors.

The use of the Rasch Model has been used in research (Nirwana, Rochman, & Zukmadini, 2019) which has the advantage of being able to determine the validity of the reliability of an instrument of various types, the level of difficulty per question, and the question characteristic curve. The questions contained in the developed test instrument

display contextual stimuli in everyday life, but still relate to the concepts being studied. This is in line with Fanani's opinion (2018) which says that in the context of higher-order thinking, the stimulus presented should be contextual and interesting. Contextual and interesting stimuli are useful for attracting students' attention to read the stimulus to completion.

The question indicators on this test instrument were developed from KD 3.7 in the revised 2013 curriculum which refers to the higher-order thinking indicators of Bloom's taxonomy which have been revised by Anderson & Krathwohl. Based on Bloom's taxonomy which has been revised, higher order thinking skills involve aspects of analyzing (C4), evaluating (C5) and creating (C6) (Anderson & Krathwohl, 2001). Question indicators on the test instrument developed using three cognitive levels, namely analyzing, evaluating, and creating; and four categories of knowledge dimensions, namely factual, conceptual, procedural, and metacognitive as shown in Table 10.

Table 10. Developed Higher Order Thinking Indicators

Cognitive Level	Knowledge Dimension	Question Number	Type Question	Sum
C4	Conceptual	2	Reasonable multiple choice	1
		9 & 10	Reasonable cause and effect	2
		17, 18, 19, & 20	Reasonable multiple response	4
	Factual	1	Reasonable multiple choice	1
		11 & 12	Reasonable cause and effect	2
C5	Conceptual	5	Reasonable multiple choice	1
		22	Reasonable multiple response	1
	Factual	13 &14	Reasonable cause and effect	2
	Metacognitive	2 & 4	Reasonable multiple choice	2
		21	Reasonable multiple response	1
C6	Conceptual	6, 7, & 8	Reasonable multiple choice	3
		15 & 16	Reasonable cause and effect	2
		23	Reasonable multiple response b	1
	Factual	24	Reasonable multiple response	1
Sum				24

The specifications of the high-level thinking ability test instrument for high school students on Newton's Law of motion, namely: (1) the test instrument developed refers to the indicators of higher-order thinking skills according to Anderson & Krathwohl (2001) covering questions with analyzing, evaluating, and creating skills. ; (2) the purpose of developing a higher-order thinking test instrument on Newton's Law of motion is to produce a product that can measure and train students' higher-order thinking skills in physics learning, and can be used by teachers as a student evaluation tool on Newton's Law of motion; (3) operational verbs in the indicator questions developed include C4 (analyze, differentiate, and find), C5 (connect, examine, and assess), and C6 (make); (4) the test instrument consists of 24 items which are divided into 8 items of reasoned plural choice questions, 8 types of cause and effect questions, and 8 reasoned matching questions; (5) the processing time for 25 items of the higher-order thinking ability test instrument is 60 minutes.

The high-order thinking ability test instrument on Newton's Law of motion material developed has met the standards for assessment, because the test instrument has good validity and reliability, where the questions contained are questions that describe real phenomena in everyday life but still relate to the physics concepts that have been studied by students. This is in accordance with the opinion of Nuswowati, et al (2010) that a test can be said to be good as a measuring tool if it meets the requirements of a good test, including valid and reliable. The test instrument has been developed and has been validated by two expert lecturers and 1 expert teacher which was then tested on 53 students of class X IPA MAN 2 Tangerang. Furthermore, the test data were analyzed to obtain information about the validity, reliability, level of difficulty of the questions, and the level of students' higher-order thinking skills.

CONCLUSION

The test instrument for higher order thinking skills on Newton's Law of motion material that uses information, pictures, graphs, or data to solve problems, describes detailed information, relates one concept to another, and looks for connections from different information can train thinking skills high level of students. The items on the higher-order thinking ability test instrument developed have different levels of difficulty, including 1 very difficult question, 13 difficult questions, 9 easy questions, and 1 very easy question.

REFERENCES

Amalia, A.N. & Widayati, A. (2012). Analisis Butir Soal Tes Kendali Mutu Kelas XII SMA Mata Pelajaran Ekonomi Akuntansi di Kota Yogyakarta Tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia, 10*(1): 1-26.

Anderson, L.W., dan Krathwohl, D.R. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom’s Taxonomy of Educational Objectives. New York: Addison Wesley Longman, In.

Annuuru, T. A., Johan, R.C., & Ali, M. (2017). Peninkatan Kemampuan Berpikir Tingkat Tinggi dalam Pelajaran Ilmu Pengetahuan Alam Peserta Didik Sekolah Dasar Melalui Model Pembelajaran Treffinger. *Edutcehnologia, 3*(2): 136-144.

Arikunto, S. (2012). *Dasar-dasar Evaluasi Pendidikan Edisi 2*. Jakarta: Bumi Aksara. 344 hlm.

Arikunto, Suharsimi. (2002). Prosedur Penelitian Suatu Pendekatan Praktek. Jakarta : PT Rineka Cipta. Hlm. 129

Bialik, M., Bogan, M., Fadel, C., & Horvathova, M. (2015). Character education for the 21st century: What should students learn. *Boston, Massachusetts: Center for Curriculum Redesign*.

Bima, A. (2019). *SISTEM INFORMASI PENILAIAN K13 DI SMP NEGERI 2 WEDI* (Doctoral dissertation, Universitas Widya Dharma).

Boone, W.J., Staver, J.R., & Yale, M.S. (2014). Rasch Analysis in the Human Science. Dordrecht: Springer.498 p.

Borg, W.R & Gall, M. D. (1983). *Educational Research:* An Introduction,. Fifth Edition. Newyork: Longman.

Crawford., Caroline, M., Evelyn.(2002). Focusing Upon Higher Order Thinking Skills: Webquests and The Learner-Centered Mathematical Learning Environment. Texas: ERIC. 16 hlm.

Daryanto. (2010). Evaluasi Pendidikan. Jakarta: Rineka Cipta. 236 hlm.

Hamdi, S., Suganda, I. A., & Hayati, N. (2018). Developing higher-order thinking skill (HOTS) test instrument using Lombok local cultures as contexts for junior secondary school mathematics. *REiD (Research and Evaluation in Education), 4*(2), 126-135.

Heru, M., & Suparno, S. (2019). The Development of Reasoned Multiple Choice Test in Interactive Physics Mobile Learning Media (PMLM) of Work and Energy Material to Measure High School Students' HOTS. *Formatif: Jurnal Ilmiah Pendidikan MIPA, 9*(2).

Ku, K. Y. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking skills and creativity*, *4*(1), 70-76.

Lewy, Z., & Aisyah, N. 2009. Pengembangan Soal untuk Mengukur Kemampuan Berpikir Tingkat Tinggi Pokok Bahasan Barisan dan Deret Bilangan di Kelas IX Akselerasi SMP Xaverius Maria Palembang. *Jurnal Pendidikan Matematika*. 8(2): 15-28.

Mardapi, D. (2017). Pengukuran, Penilaian, dan Evaluasi Pendidikan. Yogyakarta: Parama Publishing

Matondang, Z. (2009). Validitas dan Reliabilitas Suatu Instrumen Penelitian. *Jurnal Tabularasa, 6*(1): 87-97.

Nirwana, N., Rochman, S., & Zukmadini, A. Y. (2019, April). An assessment of Higher Order Thinking Skills (HOTS) Based on Rasch Models of Student in Physics Learning. In *International Conference on Educational Sciences and Teacher Profession (ICETeP 2018)*. Atlantis Press.

Putri, F. S., Istiyono, E., & Nurcahyanto, E. (2016). Pengembangan instrumen tes keterampilan berpikir kritis dalam bentuk pilihan ganda beralasan (politomus) di DIY. *UPEJ Unnes Physics Education Journal*, *5*(2), 76-84.

Rahmawati, N.D., Amintoko, G., & Faizah, S. (2018). Kemampuan Berpikir Tingkat Tinggi Mahasiswa dalam Memecahkan Masalah Fungsi Pembangkit. *Jurnal Elektronik Pembelajaran Matematika, 5*(1): 21-31.

Retnawati, H., Djidu, H., Kartianom, A., & Anazifa, R. D. (2018). Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problems of Education in the 21st Century, 76*(2), 215.

Rofiah, E.,Aminah, N.S., & Ekawati, E.Y. (2013). Penyusunan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Fisika pada Siswa SMP. *Jurnal Pendidikan Fisika, 1(*2):17-22.

Sinaga, T. (2015). Pengembangan soal model PISA mata pelajaran ilmu pengetahuan alam terpadu konten fisika untuk mengetahui penalaran siswa kelas IX. *Jurnal Inovasi dan Pembelajaran Fisika*, *2*(2), 194-196.

Solekhah, Fitri M., Maharta, N., & Suana, W. 2018. Pengembangan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi pada Materi Hukum Newton tentang Gerak. *Journal of Physics and Science Learning*. 2(1): 17-26.

Sudaryono. (2013). Pengembangan Instrumen Penelitian Pendidikan. Yogyakarta: Graha Ilmu.

Sudijono, A. (2011). Pengantar Evaluasi Pendidikan. Jakarta: Raja Grafindo Persada. 504 Hlm.

Sudjana, N. 2010. Penilaian Hasil Proses Belajar Mengajar. Bandung: PT Remaja Rosdakarya. 180 hlm

Sumintono, B., & Widhiarso, W. (2015). Aplikasi Permodelan Rasch Pada Assessment Pendidikan. Cimahi: Trimkomunikata. 148 Hlm.

Susetyo, B. (2015). Prosedur Penyusunan & Analisis Tes: Untuk Penilaian Hail Belajar Bidang Kognitif. Bandung : Refika Aditama. 438 hlm.

Trianto. (2013). Mendesain Model Pembelajaran Inovatif, Progresif, Konsep, Landasan, dan Implementasinya pada Kurikulum Tingkat Satuan Pendidikan (KTSP). Jakarta: Kencana Predana Media Group.

Utari, J. I., & Ermawati, F. U. (2018). Pengembangan Instrumen Tes Diagnostik Miskonsepsi Berformat Four-Tier untuk Materi Suhu, Kalor dan Perpindahannya. Inovasi Pendidikan Fisika, 07(03), 434–439.

Wardhani, D. F., dan Putra, A. P. (2016). Pengembangan Instrumen Tes Standar Kognitif pada Mata Pelajaran IPA Kelas 7 SMP Di Kabupaten Banjar. In Proceeding Biology Education Conference: Biology, Science, Enviromental, and Learning (Vol. 13, No. 1, pp. 75-82).

Wibowo, A., & Cholifah, T. N. (2018). Instrumen tes tematik terpadu kurikulum 2013 berbasis PISA's literacy bagi siswa sekolah dasar. *JIPVA (Jurnal Pendidikan IPA Veteran), 2*(2), 209-221.

Widhy, P. (2013). Integrative Science untuk Mewujudkan 21st Century Skill dalam Pembelajaran IPA SMP. In *Makalah disampaikan pada Seminar Nasional MIPA UNY*.

Yee, M. H., Yunos, J. M., Othman, W., Hassan, R., Tee, T. K., & Mohamad, M. M. (2015). Disparity of learning styles and higher order thinking skills among technical students. Procedia-Social and Behavioral Sciences, 204:143-152.