



Development of Assessment Instruments for Higher Order Thinking Skills (HOTS) of Regulatory System Material in High School

Fajrin Nabila[✉], Krispinus Kedati Pukan¹

¹Biology Department, FMIPA, Universitas Negeri Semarang, Indonesia

Article Info

Article History:

Received: 2021

Accepted: 2021

Published: December 2021

Keywords:

Assessment Instruments,
Higher Order Thinking Skills,
Concept Understanding,
Critical Thinking Skills

Abstract

The students' comprehension on the regulatory system material from the National Examination at MAN Purworejo for the 2016/2017 to 2018/2019 academic year shows that the results decreased. Low comprehension shows understanding of students concepts is low. Efforts to increase understanding of the concept by familiarizing students with HOTS questions. HOTS assessment instruments are able to improve students understanding of concepts and higher order thinking skills. This study aimed to analyze the feasibility of HOTS assessment instruments and measure the conceptual understanding and higher order thinking skills of MAN Purworejo students on the material of the regulatory system. This study uses the *Research and Development* (R&D) method with the ADDIE (Analysis, Design, Development, Implementation, Evaluation) development model. The product was tested on 62 students, namely grade XII MIPA 1 and XII MIPA 2 at MAN Purworejo. The data collection instruments used were the material expert validation questionnaire instruments, the learning evaluation expert validation questionnaire instruments teacher and student response questionnaire instruments. The results showed that the score of the material experts 0.75, learning evaluation experts 0.75, teachers 0.83, so the instrument was declared very feasible. In addition, the result of the analysis of the responses of teachers and students obtained scores of 95.83% and 95% with very practical criteria.

© 2021 Universitas Negeri Semarang

[✉] Correspondence Address:
D6 Building 1st Floor Jl Raya Sekaran Gunungpati Semarang
E-mail: fairinnabila602@gmail.com

p-ISSN 2252-6579
e-ISSN 2540-833X

INTRODUCTION

Learning is the process of interaction between students, teacher, and learning resources in a learning environment (Law No. 20 of 2003 on the National Education System). In the learning process the factors that influencing that is: students, teachers, learning goals, materials, learning methods, and assessments. The success of the assessment is influenced by teacher quality (Lawrence & Pallrand, 2000). Assessment is the process of gathering information to determine achievement of learning targets (Nitko & Brookhart, 2014). In assessment process, tools are used in the form of assessment instruments. The assessment instrument used in learning is able to influence students thinking abilities (Istiyono, Mardapi, & Suparno, 2014).

The success of a learning process can be seen from the results of the National Examination. The absorption of regulatory system material on the results of the National Examination at MAN Purworejo for the 2016/2017 to 2018/2019 shows that the results decrease was 61,39; 36,51%; and 29,21%. The low absorption of students indicates that students understanding of concepts is not optimal. Research by Ningsih *et al.* (2018) shows that 92,5% of the 2016/2017 UN questions are *higher order thinking skills* (HOTS). To complete the HOTS assessment, higher order thinking skills are needed. Good higher order thinking skills have a positive effect on improving learning outcomes (Fayakun & Joko, 2015). Higher order thinking skills can be learned and taught to students and, if trained, can be improved (Thomas & Thorne, 2011).

The problem in MAN Purworejo is teachers' ability to develop HOTS assessment instruments is low. This is in accordance with the interviews result on April 13, 2020, that the biology teacher MAN Purworejo have difficulty in making HOTS questions with stimulus attractive and easy to understand. The observation result at MAN Purworejo shows that 85,72% of the total questions used by the teacher evaluation were classified as LOTS (C1, C2, and C3). The number of items with cognitive levels C1, C2, C3, C4 respectively is 28,57%; 35,71%; 21,42%; and 14,28%. In addition, 70% of the questions used by the teacher in the evaluation were not equipped with interesting and contextual stimulus. Item question contains the essence of the question without being preceded by a stimulus. The LOTS instrument only requires the ability to remember and understand facts, symptoms, and existing concepts (Saïdo, Si raj, Nordin, & Al-Amedy, 2015).

Higher Order Thinking Skills is one of the curriculum objectives that are contained in the Basic Framework and Structure of the SMA / MA Curriculum (Mendikbud, 2013). The HOTS test instrument is able to encourages the improvement of higher order thinking skills (HOTS) in the form of deep conceptual understanding but also encourages the improvement of students' low order thinking skills (LOTS) (Jensen, 2014). HOTS question is able to train students to developing higher order thinking skills, namely at the level of analysis, evaluation, and creation (Suryapusparini *et al.*, 2018).

The instrument is a measuring tool used to collect/retrieve data objectively so that it can draw conclusions that are objective (Purwanto, 2010). Assessment is the process of gathering information to help teachers decide student achievement towards learning goals/targets (Nitko & Brookhart, 2014). The basic principles in developing HOTS assessment instruments are: (1) using a stimulus or introduction, (2) using new and contextual material, (3) consider complexity and cognitive difficulties (Brookhart, 2010).

Multiple choice item can be used to measure students ability to interpret cause and effect relationships (Miller *et al.*, 2009). Multiple choice questions are objective and if they are of good quality, it can accurately distinguish high and low ability students (Schuwirth and van der Vleuten, 2003). Short stuffing questions are questions that require students to fill in short answers with words, phrases, numbers, or symbols.

Higher order thinking skills are *thinking skills* that occur when a person takes information and new information that stored in memory and connect the information to decide what to believe, what to do, and create new ideas (Lewis & Smith, 1993). Higher order thinking skills help a person to use previous information to solve problems (Heong *et al.*, 2011). HOTS questions have the following characteristics: (1) transferring one concept to another, (2) processing and applying information, (3) connect a variety of different information, (4) using information to solve problems, (5) critically examining ideas and information (Kemendikbud, 2019). According to Kemendikbud (2017) HOTS questions have characteristics, namely: (1) measuring high-level thinking skills, (2) based on contextual problems, (3) using various question forms, and (4) using a cognitive level of analyzing, evaluate, and create. Brookhart, (2010) defines HOTS as a transfer process which includes skills to analyze, evaluate, and create. Riswanda conducted research (2018) with the title of research is Problem Development Based on *Higher order Thinking Skills* (HOTS) and its implementation at SMA Negeri 8 Palembang. From the test results it is known that all items have very good category reliability (0.93). While the results of the reliability of the person showed that HOTS person's ability was classified as sufficient (0.72).

Based on the information above, it is necessary to develop a HOTS assessment instrument to analyze the feasibility of the instrument and to measure students understanding of concepts and critical thinking skills in the regulatory system material.

RESEARCH METHOD

This study is a Research and Development (R&D) research with the ADDIE (Analysis, Design, Development, Implementation, Evaluation) development model. The model consists of five steps: (1) analysis (*analyze*), (2) design (*design*), (3) development (*develop*), (4) implementations (*implement*), and (5) evaluation (*Evaluate*) (Branch, 2009).

The small-scale trial process was carried out at MAN Purworejo in the 2019/2020. Students are selected using *purposive sampling technic*. The research subjects were grade XII IPA 1 and XII MIPA 2 students and biology teachers at MAN Purworejo. The data obtained were the validity of the HOTS assessment instrument, the profile of the ability to analyze, the profile of the ability to evaluate, the profile of the ability to create, and the responses of teachers and students who were analyzed descriptively quantitatively.

RESULTS AND DISCUSSION

Characteristics

The results of the research on the development of HOTS assessment instruments, the regulatory system material is summarized in three main parts, namely instrument characteristics, expert validation results, and students HOTS profile. Characteristics of the HOTS assessment instrument of the development results consisted of 48 multiple choice items and 4 short answer items containing HOTS indicators (analyzing, evaluating, and creating). Each indicator has a different number that is contained in the instrument. The following is data on the composition of indicators for high-order thinking skills in products developed before and after going through the expert validation process can be seen in Table 1.

Table 1 Composition of High Order Thinking Skills Indicators on HOTS Assessment Instruments

HOTS indicator	Before being Validated		After being Validated			
	Number of Questions	Percentage	First Revision		Second Revision	
			Number of Questions	Percentage	Number of Questions	Percentage
Analyze	25	46,29%	32	59,26%	31	59,61%
Evaluate	25	46,29%	18	33,34%	17	32,69%

Create	4	3,70%	4	7,40%	4	7,69%
Total	54	100%	54	100%	52	100%

The HOTS assessment instrument consists of 48 multiple choice questions and 4 short answers questions. The number of items with analyzing indicators on development products was the most (59.61%) compared to the number of items with evaluating indicators (32.69%) and creating indicators (7.69%). This is because Basic Competencies 3.3 with operational verbs analyze consists of 3 sub-materials, namely nervous system, hormonal system, and sensory system while Basic Competencies 3.4 with operational verbs evaluates consists of one material, namely drugs, so the number of questions with indicators analyzes is more than indicators evaluate and create. Meanwhile, the smallest number of items is the number of items with create indicator. This is because the regulatory system material covers 2 Basic Competencies, namely Basic Competencies 3.3. with operational verbs analyze and Basic Competencies 3.4 with operational verbs evaluate, so it is difficult to develop items with creating indicators. This is supported by the opinion of Kemendikbud (2017) that not all Basic Competencies can be made into HOTS questions so in this study, both Basic Competencies are easier when HOTS questions are made with indicators of analyzing and evaluating compared to indicators of creating. In addition, creating indicators in the regulatory system material are developed in the form of short questions, because questions with creating indicators require complex answers, so require the appropriate scoring type. In choosing the type of question, should consider the following matters: (1) suitability with the subject matter, (2) suitability with evaluation goal, (3) suitability with scoring, (4) suitability with processing of evaluation results, (5) conformity with test administration, (6) suitability with funds and practicality (Yusrizal, 2016).

In the HOTS assessment instrument as a result of development, there was change in the amount for each HOTS indicator based on material expert validation. There were 2 items that were discarded, namely item number 9 with analyzing indicators and item number 48 with evaluating indicators. In analyzing indicators, there was a percentage increase of 0.35%. The percentage of items with the evaluating indicator decreased of 0.65%, and there was an increase in the percentage of items with the creating indicator of 0.29%.

Validity of HOTS Assessment Instruments

Instrument validated by material experts, learning evaluation experts, and MAN Purworejo teachers. Material expert validation assesses the content aspects of the questions. Validation by learning evaluation experts includes aspects of construction and language. The validation results from material experts obtained a value of 0.75 with medium validity criteria. The results of the material expert validation can be seen in Table 2.

Table 2 Results of the HOTS Assessment Instrument Material Validation

Aspect	Result			V	Criteria
	1	2	3		
I. Content					
A. Material suitability with basic competencies	12	11	11	0.76	Medium
B. Material depth	4	3	3	0.78	Medium
C. Material accuracy	15	13	14	0.67	Medium
D. Answer choices accuracy	12	12	11	0.79	Medium
Average of Validation Result				0.75	Medium

Criteria of test validity:

$V < 0,4$	= Poor
$0,4 \leq V \leq 0,8$	= Medium
$0,8 < V$	= High (Retnawati, 2016)

The results of the material validator assessment indicate that the questions are appropriate to the material aspect. Daryanto (2008) says that good questions in terms of material have the following characteristics: (1) The questions are in accordance with the basic competency indicators, (2) the boundaries of questions and answers are clear, and (3) the material asked is in accordance with the level, type school, and class.

The evaluation of the learning evaluation experts includes aspects of construction and language. The score of the validator assessment on construction and language aspects obtained a value of 0.75 with medium validity criteria. The results of the learning evaluation expert assessment can be seen in table 3.

Table 3 Results of the HOTS Assessment Instrument Construction and Language Validation

Aspect	Result		V	Criteria
	1	2		
I. Construction				
A. Question construction accuracy	14	16	0.73	Medium
B. Answer choices accuracy	8	7	0.79	Medium
C. Conformity with HOTS question characteristics	13	13	0.6	Medium
Average			0.71	Medium
II. Language				
D. Conformity with Indonesian Language Rules	8	8	0.86	High
E. Straightforward	6	8	0.71	Medium
F. Communicative	3	4	0.83	High
Average			0.8	High
Average of Validation Result			0.75	Medium

Criteria of test validity:

$V < 0,4$	= Poor
$0,4 \leq V \leq 0,8$	= Medium
$0,8 < V$	= High (Retnawati, 2016)

The instrument of development results has qualified of the construction and language aspect. Based on the results of the assessment of the learning evaluation expert validator, the validity criteria are medium (0.75). In the construction aspect, it was obtained a value of 0.71 with medium validity criteria. Feasibility in the construction aspect indicates that the questions are feasible in construction. Questions that are good in construction should qualify the following requirements: (1) the subject matter is formulated clearly and firmly, (2) the main questions and answers only contain the statements that are needed, (3) the subject matter does not give any indication that there is a correct answer, (4) does not contain multiple negative sentences, (5) the answer choices are homogeneous and logical, (6) the answer

choices do not contain "All of the above choices are wrong" or "All of the above choices are correct", (7) the length of the answer options is relative the same, (8) pictures, graphs, tables, diagrams contained in the questions must be clear and functional, (9) the question items do not depend on the answers to the previous questions (Yusrizal, 2016).

In the language aspect, the value of the two validators is 0,8 with high validity criteria. This shows that the question instrument has qualified the standards so students easy to understand. Feasibility on the language aspect shows that the questions have used appropriate Indonesian rules, did not use regional languages, and used communicative sentences (Kemendikbud, 2017).

Developed instrument validated by material experts and educational evaluation experts was also validated by the teacher as an instrument user. Validation by teacher's instrument contains three aspects: material aspect, construction aspects, and language aspect. The results of the instrument validation by the teacher can be seen in Table 4.

Table 4 Results of the Validation of HOTS Assessment Instruments by the Teacher

Aspect	Result	V	Criteria
I. Material			
A. Material suitability with basic competencies	12	0.82	High
B. Material depth	4	1	High
C. Material accuracy	15	0.73	Medium
		0.85	High
II. Construction			
D. Question construction accuracy	16	0.8	High
E. Answer choices accuracy	8	0.86	High
F. Conformity with HOTS question characteristics	14	0.67	Medium
		0.77	Medium
III. Language			
G. Comformity with Indonesian Language Rules	8	0.86	High
H. Straightforward	7	0.71	Medium
I. Communicative	4	1	High
		0.86	High
Average		0.83	High

Criteria of test validity:

- V < 0,4 = Poor
- 0,4 ≤ V ≤ 0,8 = Medium
- 0,8 < V = High (Retnawati, 2016)

In the material aspect, it was obtained a value of 0.85 with high validity criteria. The feasibility of the material aspect makes it easy for the teacher because the questions used are in accordance with the concept of the material, as well as the depth of the material is right for grade XII SMA. In the construction

aspect, it was obtained a value of 0.77 with medium validity criteria. These results indicate that the question construction have qualified the standards. A good question construction is able to measure the ability of students accurately because students are not given the opportunity to guess the answer by looking at the length of the answer choices or the inhomogeneity and illogical choice. Whereas in the language aspect based on the teacher assessment, it was obtained a value of 0.86 with high validity criteria. The proper language aspect of the question instrument makes it easier for students to understand the meaning of the question. Feasibility on the material aspects, construction aspects, and language aspects shows that the development of the questions has followed the following guidelines:

1. questions using new and relevant material,
2. question contains only one right answer,
3. question content is independent,
4. the question stem contains a problem and is as short as possible,
5. the main idea is in the question stem and not in the choices,
6. choices must be grammatically consistent,
7. independent choices and should not overlap,
8. homogeneous choices and the same length,
9. right answers appear for each choice /option in the same number, and are randomly distributed,
10. reasonable distractor (Miller et al., 2009; Haladyna, 2004).

Teachers and Students Responses to the HOTS Assessment Instrument Regulatory System Material

Questionnaire for teacher and student responses aims to determine the responses of teachers and students as users of development results instruments. The results of teacher responses can be seen in Table 5.

Table 5 Teacher Responses

Statement	Percentage
1. An assessment instrument developed using an affordable cost	100%
2. Ease of assessment instruments developed to be implemented in learning evaluation	100%
3. Ease of the developed assessment instrument for administration	100%
4. Ease of developed assessment instruments for general use	100%
5. The assessment instrument is equipped with instructions for use to make it easier for students to use the assessment instrument	100%
6. The assessment instrument is equipped with scoring guidelines to make it easier for teachers to assess students' abilities	75%
Average	95.83%
Criteria	Very Practical

The table shows score of the instrument practicality indicator assessment is 95% with very good category. The teacher response questionnaire regarding the practicality of the product consists of 6 indicators, namely: (1) the cost used to develop is affordable, (2) the ease of the instrument to be implemented in the evaluation of learning, (3) the ease of the instrument to be administered, (4) the ease of the instrument for general use, (5) completeness of instructions for using instruments, and (6) completeness of scoring guidelines. The teacher assessment score is 95% with very good criteria. The high practicality aspect means that the instrument is the result of practical development for reuse by the teacher in evaluating learning. The practicality of the instrument is the possibility of an instrument to be used by

teachers in measuring learning success (Purwanto, 2010). Kunandar (2014) said that an assessment instrument is practical if it is easy to use administratively and technically.

Student questionnaire responses were given to 62 students via *google form*. The results of student responses can be seen in Table 6.

Table 6 Students' Responses

Respondents	Percentage									
	1	2	3	4	5	6	7	8	9	10
Small Scale Trials	78%	97%	86%	94%	87%	92%	90%	95%	98%	97%
Average Percentage	91%									
Criteria	Very Worthy									

Based on the results of students' questionnaire via *google form* on a small-scale test, score of 91% was obtained with very feasible criteria. The lowest score is 78% on the 1st indicator about on the aspects of attractiveness of questions stimulus and students interest in questions. The highest score obtained was the 9th indicator about on the ability of questions in developing students critical thinking skills.

The student response questionnaire aims to know how students respond to the HOTS questions developed. The score of students responses was 91% with very good criteria. The indicators on the students responses consist of 10 indicators, namely: (1) the attractiveness of the question stimulus and the students interest in the questions, (2) the clarity of the work instructions, (3) cases on the questions encountered in everyday life, (4) easy to understanding the questions, (5) clarity of the language used, (6) there are no sentences that have multiple meanings, (7) the suitability of the types of questions used for the regulatory system material, (8) the questions test the students understanding, (9) the ability of question to develop students critical thinking skills, and (10) questions able to develop higher-order thinking skills (analyzing, evaluating, and creating). Most of the students gave positive responses to instrument products about the development results. Students interest in a material is proportional to students understanding (Ardiansyah, Irwandi, & Murniati, 2016).

Higher Order Thinking Skills Profile

Comparison of the average score on the indicators of analyzing, evaluating, and creating can be seen in Table 7. The profiles of students' higher order thinking skills can be seen in Table 8.

Table 7 Comparison of Students Higher Order Thinking Skills

Aspect	Score	Maximum score	Percentage	Criteria
Analyze	0.398	1	39.8%	less
Evaluate	0.3359	1	33.59%	less
Creating	0.7314	1	73.14%	good
Average	0.4884	1	48.84%	poor

Table 8 Profile of Students Higher Order Thinking Skills

Value interval	Frequency	Percentage	Criteria
X < 30%	0	0%	very less
30% < X ≤ 50%	41	66%	less
50% < X ≤ 70%	21	34%	sufficient
70% < X ≤ 90%	0	0%	good
90% < X	0	0%	very good
Average Value		48,84%	poor

High-order thinking skills can be divided into 3, namely analyzing, evaluating, and creating. Based on the table above, it is known that the ability to analyze and evaluate is in the low category with respectively scores is 39.8% and 33.59%. The ability to analyze is less because students are not accustomed to complete analysis questions. Evaluation questions on the regulatory system material used by teachers in evaluating learning are LOTS. LOTS questions were not able to develop students analytical skills. Analysis questions used by the teacher to evaluate only 14% of the total number of questions in evaluation activity. Questions with the aim of measuring the ability to analyze (C4) can be solved by involving the ability to break down parts or connect several concepts. This is in accordance with the opinion of Anderson and Krathwohl (2001) that analysis skills include 3 things, namely: (1) skills to divide information into smaller sections to determine patterns or relationships, (2) skills to distinguish between causal and effect factors of a case, and (3) identify/formulate questions. The students are able to complete the questions with analyzing indicators if they have ability to remember (C1), understand (C2), and apply (C3). The questions with analyzing indicators demands the ability of students to specify aspects/elements, describe, organize, compare, and find implied meanings (Kemendikbud, 2017). The low ability to evaluate is caused by students rarely complete the questions with evaluations indicator. This is in accordance with the analysis results of teacher evaluation questions, that there are no questions with evaluating indicators. The questions to measure the ability to evaluate (C5) can be solved by involving the ability to making a judgement of a solution to the problem or being able to determine the right or wrong statement on the questions. This is supported by the opinion of Anderson & Krathwohl (2001) that evaluation skills cover several things, namely: (1) evaluating solutions, or ideas based on appropriate assessment criteria or standards, (2) making hypotheses, or criticisms, and (3) decide to accept or reject a statement based on predetermined criteria. Item with evaluating indicators demand the ability of students to know (C1), understand (C2), apply (C3), analyze (C4) then evaluate a statement presented about the regulatory system material. This is supported by Irawati (2018) that the low ability to evaluate is influenced by the ability to think at a level below is the ability to analyze. The low ability to evaluate indicates that students are not accustomed to evaluate, justifying or blaming. The same thing was also said by Kemendikbud (2017) that the process of evaluating thinking requires students to making a hypothesis, criticism, prediction, and evaluate, test, justify or blame a statement on a case.

Based on the results of small-scale trial, it is known that the ability of students to creates in the high category with an average of 73.14%. The score on the creating indicators is higher than the questions with analyzing and evaluating indicators. This is because the short stuffing questions with creating indicators easy to solve. Most of the questions only contain commands for making diagrams graphs based on the data that available in the questions. A high score of the ability to create indicates that students are able to design, making solutions in this case that is present data in graphical form. This is supported by the Kemendikbud opinion (2017) that in the dimension of creating, learners are required to design, develop, produce, discover, renew, enhance. The ability to create is the ability of students to plan and produce something that is new, and produce *multiple solutions* (Brookhart, 2010). In questions with creating indicators (C6), students are asked to make graphs/diagrams based on the data that has been available on the questions. The questions to measure the ability to create can be solved by involving the ability to form a new structure/new solution that comes from the elements. The same thing was also conveyed by Anderson & Krathwohl (2001) that the create skills include: (1) generalizing, (2) making solutions, and (3) organizing elements into new structures.

Based on the results of the study, it is known that students high order thinking skills are classified as low. The low level of thinking skills of students is influenced by teacher method to teaches the material and the how students learn, and the skills of students in solving problems on the item of questions (Akmala, Suana, & Sesunan, 2019). Students who are accustomed to working on LOTS questions will have difficulty solving HOTS questions causing the scores obtained to be low. Higher order thinking skills

(HOTS) in students can be improved through activities that get the brain working to analyze, conclude, and evaluate (Kuswana, 2011).

Most of the HOTS abilities of students are in the low category and a small part are in the medium category. The absence of students with high HOTS abilities is because students do not understand the material of the regulatory system.

CONCLUSION

The HOTS assessment instrument for regulatory system consists of 48 multiple choice questions and 4 short stuffing question. The HOTS assessment instrument for regulatory system is declared feasible to measure higher order thinking skills based on the results of the material expert's validation (0.75), result of the learning evaluation expert's validation (0.75), result of the teacher's validation (0.83), and teacher responses (95.83%), and the students responses (95%).

REFERENCES

- Akmala, N.F., Suana, W., Sesunan, F. 2019. Analisis Kemampuan Berpikir Tingkat Tinggi Siswa SMA pada Materi Hukum Newton tentang Gerak. *Titian Ilmu: Jurnal Ilmiah Multi Sciences*. 11(2): 67-72.
- Anderson, L.W. & Krathwohl, D.R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Pearson Education.
- Ardiansyah, A.A.I, Irwandi, D., & Murniati, D. 2016. Analisis Literasi Sains Siswa Kelas XI IPA pada Materi Hukum Dasar Kimia di Jakarta Selatan. *EduChemia (Jurnal Kimia dan Pendidikan)*. 1(2), 149-161.
- Branch, R.M. (2009). *Instructional Design: The ADDIE Approach*. London: Springer.
- Brookhart, S.M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. Virginia: ASCD Publications.
- Daryanto. (2008). *Evaluasi Pendidikan*. Jakarta: Rineka Cipta.
- Fayakun, M. & Joko, P. 2015. Efektivitas Pembelajaran Fisika Menggunakan Model Kontekstual (CTL) dengan Metode Predict, Observe, Explain) Methods Toward High-Order Thinking Skill. *Jurnal Pendidikan Fisika Indonesia*. 11(1), 49-58.
- Haladyna, T.M. (2004). *Developing and Validating Multiple-Choice Test Items (3rd ed)*. Mahwah: Lawrence Erlbaum Associates.
- Heong, Y.M., Othman, W.D., Yunos, J.B., Kiong, T.T., Hassan, R.B., & Mohamad, M.M.B. 2011. The Level of Marzano Higher Order Thinking Skills Among Technical Education Students. *International Journal of Social and Humanity*. 1(2), 121-125.
- Irawati, T.N. 2018. Analisis Kemampuan Berpikir Tingkat Tinggi Siswa SMP dalam Menyelesaikan Soal Pemecahan Masalah Matematika pada Materi Bilangan Bulat. *Jurnal Gammath*. 3(2): 1-7.
- Istiyono, E., Mardapi, D., & Suparno. 2014. Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PysTHOTS) Peserta Didik SMA. *Jurnal Penelitian dan Evaluasi Pendidikan*. 18(1), 1-12.
- Jensen, J.L., Mc Daniel, M.A., Woodard, S.M., & Kummer, T.A. 2014. Teaching to the Tes or Testing to Teach: Exams Requiring Higher Order Thinking Skills Encourage Greater Conceptual Understanding. *Educational Psychology Review*. 26, 307-329.
- Kemendikbud. (2017). *Modul Penyusunan Soal Higher Order Thinking Skill (HOTS)*. Jakarta: Direktorat Pembinaan SMA Direktorat Jenderal Pendidikan Dasar dan Menengah Menengah Departemen Pendidikan dan Kebudayaan.
- Kemendikbud, (2019). *Buku Penilaian Beorientasi Higher Order Thinking Skills*. Jakarta: Direktorat Jenderal Guru dan Tenaga Kependidikan Kementerian Pendidikan dan Kebudayaan.
- Kunandar. (2014). *Penilaian Autentik (Penilaian Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013) Suatu Pendekatan Praktis Disertai Contoh*. Jakarta: Rajawali Pers.
- Kuswana, W.S. (2011). Taksonomi Berpikir. Bandung: Rosda Karya.
- Lawrence, M. & Pallrand, G. J. 2000. A Case Study of The Effectiveness of Teacher Experience in The Use of Explanation-based Assessment in High School Physics. *School Science and Mathematics*. 100(1), 36-47.
- Lewis, A. & Smith, D. 1993. Defining Higher Order Thinking. *Theory Into Practice*. 32(3), p. 131-137.
- Mendikbud. (2013). Peraturan Menteri Pendidikan dan Kebudayaan Nomor 69 Tahun 2013 tentang Kerangka Dasar dan Struktur Kurikulum Sekolah Menengah Atas/Madrasah Aliyah.
- Miller, M. A., Linn, R., & Gronlund, N. E. (2009). *Measurement and Assessment in Teaching (10th ed)*. New Jersey: Pearson Education, Inc.
- Ningsih D.L., Marpaung, R.R.T., & Yolida, B. 2018. Analisis Soal Ujian Nasional Biologi Sekolah Menengah Atas. *Jurnal Biomedik*. 6(6), 1-10.
- Nitko, A.J. & Brookhart, S.M. (2014). *Educational Assessment of Students (6th ed)*. Edinburgh: Pearson Education Limited.
- Purwanto. (2010). *Metodologi Penelitian Kuantitatif*. Yogyakarta: Pustaka Pelajar.
- Riswanda, J. 2018. Pengembangan Soal Berbasis Higher Order Thinking Skills (HOTS) serta Implementasinya di SMA

- Negeri 8 Palembang. *Jurnal Penelitian Pendidikan Biologi*. 2 (1), 49-58.
- Saido, G.A.M., Siraj, S., Nordin, A.B., & Al-Amedy, O.S. 2015. Teaching Strategies for Promoting Higher Order Thinking Skills: A Case of Secondary Science Teachers. *Malaysian Online Journal of Educational Management*. 3(4), p. 16-30.
- Schuwirth, L.W.T., van der Vleuten, C.P.M., 2003. ABC of Learning and Teaching in Medicine: Written Assessment. *BMJ* 326 (7390), 643–645.
- Suryapuspitarini, B.K., Wardono, & Kartono. 2018. Analisis Soal-Soal Matematika Tipe Higher Order Thinking Skill (HOTS) ada Kurikulum 2013 untuk Mendukung Kemampuan Literasi Siswa. Prosiding 1st Seminar Nasional Matematika. Semarang: Universitas Negeri Semarang.
- Thomas, A. & Thorne, G. 2011. How to Increase Higher Level Thinking. Online at <https://www.cdl.org/how-to-increase-high-order-thinking/> [Diakses pada 10 Januari 2020].
- Undang-Undang Republik Indonesia No. 20 Tahun 2003. *Sistem Pendidikan Nasional*. 8 Juli 2003. Lembaran Negara Republik Indonesia Tahun 2003 Nomor 4301. Jakarta.
- Yusrizal. (2016). *Tanya Jawab Seputar Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Banda Aceh: Syiah Kuala University Press.