

Higher Order Thinking Skills: Student Profile Using Two-Tier Multiple Choice Instrument

Submitted 24 March 2022, Revised 23 April 2022, Accepted 24 April 2022

Indri Andriyatno¹, Zulfiani Zulfiani^{2*}, Yuke Mardiaty³

¹²³Department of Biology Education, Faculty of Tarbiya and Teaching Sciences,
Universitas Islam Negeri Syarif Hidayatullah Jakarta, Tangerang Selatan, Indonesia
Corresponding Email: *zulfiani@uinjkt.ac.id

Abstract

This research aimed to analyze students' Higher Order Thinking Skills (HOTS) using a two-tier multiple choice (TTMC) test instrument. This study used a descriptive quantitative method with a purposive sampling technique. The instruments used are multiple-choice tests with free reasoning, the TTMC test for quantitative data, interview guides, and learning observations for qualitative data. Anderson and Krathwohl's Bloom taxonomy revision (2001) cognitive instrument used to classify HOTS. The results showed that the higher-order thinking skill profile has an average score of 63,21 and dominated good. Students' cognitive level of analyzing/C4 (74,35%) is more dominant compared to cognitive evaluating/C5 (68,90%) and creating/C6 (63,24%). The hormonal regulatory subconcept gets the highest correct answer average with a percentage of 89,74%, and the contraceptive method subconcept gets the lowest correct answer average with 35,90%. One factor that influences the high and low percentage of the correct answer for each subconcept is the level of difficulty of the question; medium (enough) in the contraceptive method and too tricky in the contraceptive method. Analysis of answer patterns based on the results of a TTMC test shows that students are better able to answer questions on the first tier than answering on the second tier.

Keywords: Higher Order Thinking Skill (HOTS), Two-Tier Multiple Choice Instrument, Human Reproductive System

INTRODUCTION

Higher-order thinking skill (HOTS) is the ability to transform knowledge and experience in an organized manner to solve problems. HOTS can also be interpreted as finding explanations or solutions in difficult and perplexing situations. By integrating new information with information stored in earlier memories, linking to each other, unifying the information, and applying it as a new solution, students require a large number of facts (Mislia et al., 2019; Zulfiani et al., 2015); students expect to apply skills and knowledge to produce a conclusion (Brookhart, 2010); increase the qualities and education of students (Pratiwi & Mustadi, 2021). By thinking at a high level, students are expected to understand complicated things more clearly.

The value of sciences' Indonesian students in TIMMS 2015 was 397 and ranked 44th out of 47 countries (TIMSS & PIRLS, 2015). Meanwhile, the scientific literacy score in PISA 2018 was 396 and ranked 70st from 78 countries, down from 2015 with a score of 403 (OECD, 2019). Both values obtained are below the international average. The research of question items in 26 subjects in 34 provinces revealed that most of the 1,779 questions studied were at level 1 (knowledge and understanding) and level 2 (application). In the 136 high schools investigated, 27 schools compiled HOTS questions that accounted for up to 20% of all USBN problems, 84 schools for less than 20%, and 25 schools indicated they had no idea

whether the problems compiled by HOTS or not (Mukhtar & Haniin, 2019). The test questions that teachers to students are often the test is a type of questions at the level of cognitive processing C1-C3, which includes a low level and students tend to memorize learning materials, less trained to develop skills (Noma et al., 2016). In the classroom, improving students' creativity and critical thinking skills has been a low focus (Anazifa & Djukri, 2017).

Problem levels 1 and 2 tend to be less able to measure thinking ability, so students will need help solving problems and finding the right solution. Problems with low cognitive levels are generally incapable of seeing the relationship between knowledge gained in the classroom and real context in life. Students' low ability to solve HOT problems indicates that students need to be accustomed to thinking at a high level. HOTS needs to be given to prepare competencies to face the challenges of the 21st century (Pratiwi & Mustadi, 2021); learners at all levels of education are expected to have a high HOTS score (Saltan & Divarci, 2017); superb aptitude that is useful in all element of life (Wartono et al., 2018).

The learning process is an activity that supports students in maximizing higher-order thinking. The HOTS-based learning process will stimulate students to think at higher levels and have the ability to solve problems related to disciplines and real life. HOTS-based learning makes students accustomed to processing information based on knowledge. In addition, one alternative that can be used to improve HOTS is to prepare the right evaluation instrument that can contain complex problem-solving. One way to improve higher-order thinking is to do tests that will impact good problem-solving skills (Bhattacharya & Mohalik, 2021; Widana, 2017). HOTS can be measured by various instruments, such as modified multiple choice, short answer construction, long answer construction, and reasoned essays (Nofiana et al., 2016; Zulfiani et al., 2015). An example of an instrument that can be used to measure HOTS is Two-Tier Multiple Choice (TTMC).

The first level (tier I) in the TTMC is a matter of material concepts, while the second level (tier II) is the reason for the answers for level I. HOTS can be improved through second-level questions because questions do not directly ask the concepts tested but instead have to be answered with more complex thinking. Thus, the question in the second level is generally more complicated than the first level. The research results on the profile of HOTS using the TTMC instrument in 5 schools in Tangerang Selatan showed that the ability of students to answer questions about aspects of analyzing (C4) is higher than evaluating (C5) (Lesmana, 2016). TTMC has validity with enough interpretation and high reliability (Nofiana et al., 2016); feasible and meets the criteria as a question good results with content validity (CV) of

1,00, has an average test reliability of 0,92 which is classified as very high (Shidiq et al., 2014).

The advantage of TTMC compared to the choice of conventional forms is that the percentage of students guessing the answer is only 4%, while the use of conventional multiple-choice makes students guess the answer by up to 20% (Tuysuz, 2009). Educators can find alternative concepts and misconceptions of the topics taught (Septiana et al., 2014; Treagust, 1998). TTMC instruments can measure high-order thinking levels with 41,6% consistency (Adodo, 2013; Shidiq et al., 2014). TTMC is useful in uncovering understanding and misunderstandings so that motivated students find the right answers after carrying out the TTMC test (Bayrak, 2013; Nabilah et al., 2014; Treagust, 2006; Tsui & Treagust, 2010).

METHOD

This quantitative descriptive study aims to determine the profile of students' higher-order thinking skills using a two-tier multiple-choice test instrument. This research was conducted at one of the senior high schools in Indonesia, with the research population being all students of class XI science. The research sample used was three classes with a total of 39 students. The sampling technique was carried out by purposive sampling with the criteria; (1) all students have studied the concept of the human reproductive system, (2) all students come from regular classes so that they have the same abilities, and (3) class XI students can think abstractly so that the profile of higher-order thinking abilities can be measured by good.

The instruments used are test and non-test. There are two test instruments, i.e., free-reasoned multiple-choice and two-tier multiple-choice instruments. The development of the two test instruments was carried out according to the guidelines of Cengiz Tuysuz (2009), as follows; (a) Researchers conducted a literature study on higher-order thinking skills, TTMC instruments, and made questions on the concept of the human reproductive system, (b) Researchers conducted interviews with nine students. Answers from interviews are used as answer options for level I questions (tier I) so that one level of Multiple Choice questions is obtained with free reasons, and (c) Multiple Choice Questions of one level are tested on 38 students. The free reasons written were used as answer options for level II (tier II) questions so that a complete TTMC instrument was obtained. Then the construct validity was carried out by two learning experts so that 26 valid questions were obtained. The instrument trial was conducted on 39 students, and the Cronbach Alfa reliability value was 0.73 (reliable). Different power questions obtained 11 questions categorized very well, five right questions, and three fewer questions. The difficulty problem level obtained six questions categorized as too difficult, 12 questions as enough (medium), and 1 question as easy.

The non-test instruments used were interviews and learning observations. The type of interview used is free which aims to confirm student answers. The questions in the interview guide ask about the questions being tested, such as the easiest and most difficult sub-concepts, the learning resources used, and the number of the most difficult questions. At the same time, observation aims to observe the learning process in the classroom directly. The observation sheet was developed based on three aspects: preliminary activities, core activities such as mastery of the material, learning strategies, relationships with students, and closing activities.

Data analysis techniques in this study include; (a) two-tier multiple-choice test data scoring, (b) higher-order thinking ability profile, and (c) learning observation results. Students are given a score of one if they answer correctly on both levels (both tiers) and are given a score of zero for all other options (Tuysuz, 2009) (Table 1).

Table 1. Two-Tier Multiple Choice Instrument Scoring Guidelines

Criteria	Score
There are no correct answers on the first and second tier	0
Answer with one correct answer in the first/second tier	0
Answering with two correct answers on the first and second tier	1

Profile of students' higher-order thinking skills was obtained from the scores obtained when working on two-tier multiple-choice test questions. The higher-order thinking ability profiles are grouped into four categories which are shown in Table 2.

Table 2. Category of Higher Order Thinking Ability Profile

Student' Scores	Profile of students' higher-order thinking skills category
100-76	Excellent
75-51	Good
50-26	Enough
25-1	Less

Then the learning observation data will be used to determine the relationship between the learning activities carried out on the profile of students' higher-order thinking abilities. The observation sheet contains several aspects that must be observed and are given a checklist (√) on the yes or no choice. The checklist in the "yes" column is given a score of 1 (one) if the learning aspect is implemented and the checklist (√) in the "no" column is given a score of 0 (zero) if the learning aspect is not implemented. The criteria of learning observation results are classified into five categories. Average percentage of learning observation results (%); 0 – 20 = Very less; 21 – 40 = Less; 41 – 60 = Good enough; 61 – 80 = Good; 81 – 100 = Excellent.

RESULTS AND DISCUSSION

Results

The data described are (1) profiles of students' higher-order thinking skills, (2) Percentage correct answer TTMC tests, and (3) results of observations of the learning process.

Higher order thinking skill profile of students

The HOTS profile can be known by looking at the average score of the TTMC test against the maximum score. The results of high-order thinking skill profile are contained in Table 3.

Table 3. High Order Thinking Skills Profile

Score	Mean	Frequency	Percentage (%)	Category
$76 < N \leq 100$	79,53	9 students	23,07	Excellent
$51 < N \leq 75$	66,27	27 students	69,23	Good
$26 < N \leq 50$	43,85	3 students	7,70	Enough
$1 < N \leq 25$	0	0 students	0	Less
Total/ Mean	$\bar{x} = 63,21$	39 students	100	Good

Data in Table 3 indicates that the high level of HOTS profile is dominated in good categories with an average score of 63,21.

Percentage correct answer TTMC test

The percentage of correct answers to the TTMC test will be outlined in two explanations; the correct answers are based on the cognitive level and average of correct answers in each subconcept. The results showed that high levels of thinking ability at the cognitive level of C4 (analyzing) were higher than C5 (evaluating) and C6 (creating). The percentage of correct answers based on cognitive levels is found in Table 4.

Table 4. Percentage of Correct Answers Based on Cognitive Level

Cognitive Level		Question Number	Correct Answers on Both Tier (%)	Mean (%)	Average of All Cognitive Levels (%)
C4 (Analyzing)	Attributing	1	43,59	65,38	74,35
		2	79,48		
		7	53,84		
		13	84,61		
	Organizing	9	79,49		
C5 (Evaluating)	Checking	4	87,17	82,05	68,90
		5	71,80		
		10	87,17		
	Detecting	15	87,17	88,46	
		6	76,92		
	Judging	17	100	58,97	
		16	35,90		
19		82,05			

Cognitive Level	Question Number	Correct Answers on Both Tier (%)	Mean (%)	Average of All Cognitive Levels (%)
C6 (Creating)	Critiquing	14	76,92	63,24
		12	15,38	
	Planning	3	89,74	
	Generating	11	20,51	
		18	69,22	
		8	20,51	
Mean (Category)		66,39 (Good)	68,85 (Good)	68,83 (Good)

Data in Table 4 indicates that analyzing (C4) obtained the highest average with a percentage of 74,35%. This proves that students have been able to parse a concept into parts and can connect between them. Students can also determine how to organize a concept. Students' ability to solve the problem of evaluating categories (C5) is lower than solving the analyze category problem (C4). The percentage of cognitive levels evaluated was 68,90%. The percentage indicates that most students need help assessing whether a conclusion or solution is based on the data presented. The lowest average at all cognitive levels is at the creating cognitive level (C6) with a percentage of 63,24%. The percentage indicates that students still find combining parts into a new product challenging. Students have been unable to find the most reasonable hypothesis when presented with a problem. The two-tier multiple-choice test results also showed that the average correct answer in each subconcept varies. The average of correct answers in each subconcept is contained in Table 5.

Table 5. Average Correct Answers on Each Subconcept

No.	Subconcept	Question Number	Problem Difficulty Level	Correct Answers on Both Tier (%)	Mean (%)
1.	Structure of Reproductive Organs	1	Difficult	43,59	61,53
		2	Difficult	79,48	
2.	Hormonal Regulation	3	Medium (Enough)	89,74	89,74
3.	Gametogenesis	4	Medium (Enough)	87,17	79,48
		5	Medium (Enough)	71,80	
4.	Fertilization	6	Medium (Enough)	76,92	57,69
		7	Medium (Enough)	53,84	
		8	Difficult	20,51	
		9	Medium (Enough)	79,49	
5.	Pregnancy	10	Medium (Enough)	87,17	51,91
		11	Difficult	20,51	
		12	Difficult	15,38	
		13	Medium (Enough)	84,61	
6.	Birth	14	Medium (Enough)	76,92	82,05
		15	Easy	87,17	

No.	Subconcept	Question Number	Problem Difficulty Level	Correct Answers on Both Tier (%)	Mean (%)
7.	Contraceptive Methods	16	Difficult	35,90	35,90
8.	Reproductive Organ	17	Medium (Enough)	100	83,75
	Abnormalities	18	Medium (Enough)	69,22	
		19	Medium (Enough)	82,05	
		Mean		66,39	67,76

The data in Table 5 shows that the average correct answer is the highest, namely on the subconcept of hormonal regulation, with a percentage of 89.74%, moderate (enough). The lowest is the contraceptive method subconcept, with a percentage of 35.90%, which is too difficult.

Observation of the learning process

The results of the student profile are in line with the observations of learning in Table 6 shows that the average implementation of the learning process is categorized good with an average percentage of 72,22%.

Table 6. Learning Observation Results

No.	Observed Aspects	Implementation (%)	Mean (%)	Category
I.	Preliminary Activities	75	75	Good
II.	Core Activities			
A.	Mastery of Learning Materials	100	91,67	Excellent
B.	Learning Strategies	100		
C.	Relationships with students	75		
III.	Closing Activities	50	50	Good enough
	Mean		$\bar{x} = 72,22$	Good

Discussion

The profile of students' higher-order thinking skills dominated in the good category with an average value of 63,21. The results above are in line with the results of learning observations. The motivation for learning of students during learning is very high. All students always observe and listen to the teacher's instructions. Some students also ask questions about the information that needs to be understood. The questions asked showed a great critical attitude and curiosity toward the concepts studied. Teachers always plan lessons that encourage students to think at a high level.

The learning method is scientifically based and student-centered. At the beginning of the lesson, the teacher presents contextual phenomena with images. Contextual learning proved successful in training HOTS (Fayakun & Joko, 2015). The phenomenon presented makes students think and formulate questions. Noma et al. (2016) mentioned that one way to improve the ability to analyze (C4) is by conducting activities to identify phenomena and

formulate questions. Learning activities are continued by guiding students to make hypotheses and relevant reasons. Then the teacher gave students the worksheet and practiced questions to associate various information related to the material concept and find the linkages between subconcepts. Must-have questions discussed usually relates to material relevant to real life. Completing tasks independently allows students to evaluate the learning materials that have been obtained.

The explanation above, based on the observations, shows that during the learning process, teachers always carry out activities that can improve HOTS, such as formulating questions, giving opinions or hypotheses along with relevant reasons, and completing tasks independently in solving problems. The results of learning observations show that implementing learning stages to improve the ability to be able to think at a high stage has been met. The learning process's implementation is in line with the high level of thinking ability profile of good learners because students are used to analyzing, evaluating, and creating activities during the learning process. The scientific approach taken by teachers supports achieving high levels of thinking ability. HOTS could be realized through a scientific approach such as observing, requesting, gathering information, associating, and communicating (Wahyuni & Arief, 2015); can be improved by activities such as involving students in non-routine problem-solving activities, providing opportunities for students to construct their knowledge, and improving their ability to analyze, evaluate, and create (Apino & Retnawati, 2017). HOTS also allows students to actively develop concepts, laws, or principles (Yuniarti et al., 2018); the instructor's engagement in guiding the debate is a critical factor in encouraging students' higher-order thinking methods (Mandernach et al., 2009).

The research results show that the questions in the low category are both at the cognitive level of analyzing, evaluating, and creating because students do not understand the meaning of the questions well. It isn't easy to learn concepts abstract basis, the many uses of scientific terms and language in a subconcept, as well as the complexity of the questions given. For example, the pregnancy concept (question number 12) gets the lowest correct answer average with a percentage of 15,38%, and the problem's difficulty is too difficult (Table 5). The indicator of question number 12 is that students can criticize the influence of external to pregnancy. All students can answer correctly on the first level with a percentage of 100%. That is, students know some of the external influences on pregnancy and their affect in the 1st trimester of pregnancy. The percentage of students who answered correctly decreased after being tested back to the second level. Students assume that the embryo is vulnerable to damage such as radiation or drugs during the first trimester.

The low selection of students' correct answers is due to students needing to understand the meaning of the question well. Learners need to understand the meaning and purpose of these principles to apply the same principles and interpretations directly to the problem. It is also supported based on the findings of student interviews, which stated that several questions are not well understood, one of which is number 12 about the stage of pregnancy. Errors occur in students because of students' overgeneralization of the concept of pregnancy. Learners need to understand the essence of the actual theory to generalize concepts. This can be referred to as fixation, that students apply the same principle directly without considering its meaning and ignore the actual concept (Talanquer, 2006).

High-category questions both at the cognitive level of analyzing, evaluating, and creating are caused by teachers' creativity in teaching, classroom management that is good, and material that is relevant to life, as well as readiness and motivation for good students in receiving the material. The reproductive organ abnormalities concept (question number 17) gets the highest correct answer average with a percentage of 100% with a problem difficulty level medium (enough) (Table 5). The indicator for question number 17 is that students can predict abnormalities of reproductive organs with the fetus's condition. The matter presents a description of reproductive organ disorders accompanied by pictures. The pictures shown allow all students to answer correctly on both levels with a percentage of 100%. All students can relate the problem description to the picture presented and know the cause of pregnancy because they can answer the reason on the second level.

Reproductive organ disorders cause the selection of the correct answer by all students is a material that is quite relevant to everyday life. Then, the teacher presents an example of a picture of the disorder during the learning process so that there is a match between knowledge owned by students accompanied by actual pictures. The teacher doesn't just explain knowledge by memorizing some materials but can relate the material to real-world situations and their application in life so that it is simple to comprehend for students. The teacher also shows good classroom management, one of which is by conducting questions and answers, especially with students sitting in the back so that the interest and attention of students remain focused during learning. This is in line with research Aini & Ridwan (2021), which states that if the teacher explains the learning objectives related to everyday life and provides a stimulus for students to be active during learning, it will significantly assist students in understanding the material provided. Retnawati et al. (2018), teachers' active participation in planning, implementing, and assessing HOTS-oriented instruction can help students develop their HOTS skills.

According to the findings, students are better at working on topics at the cognitive level of analyzing (C4) than at the cognitive level of evaluating (C5) and creating (C6). Based on the analysis of answer patterns at both levels, the percentage of correct answers on first-level (tier I) questions is always more significant than the percentage of second-level answers (tier II). This is because questions at the first level ask the concept tested to be easier to answer. So it can be seen that the second level question is the reason for the first level question. Thus, the study results have proven that second-level questions on TTMC instruments can improve HOTS. Tier II facilitates testing students' learning at higher levels of thinking because students must know why they chose the answers at the first level.

Cullinane (2011) states the inclusion of reasons at the second level of the two-tier multiple-choice question form can be used to assess students' capacity to give reasons and develop higher-order thinking skills. TTMC was used as an alternative formative assessment to examine students' understanding by requiring them to apply higher-order thinking abilities in providing explanations in the second tier and discover any misconceptions they may have (Cullinane & Liston, 2011; Sesli & Kara, 2012). Teachers and the learning environment need to pay attention to high levels of thinking ability, especially in biology subjects, because biology is a science connected with various systems of living things and complicated phenomena (Zulfiani et al., 2018). HOTS are necessary for a country's educational system because they aid in the promotion of long-term learning, provide advantages in future global competition, and compete with future challenges (Aini & Ridwan, 2021; Rahayu et al., 2020). High-level thinking is also indispensable to adapt in the future as the times grow more modern. The educational challenges of the 21st century require students to think in an organized manner and be able to use information and knowledge in life. To confront the challenges of 21st-century learning, students must be capable of critical thinking skills, information literacy, as well as learning and innovating skills (Frydenberg & Andone, 2011).

CONCLUSION

HOTS profile of students achieved an average score of 63,21 and dominated the good category. The cognitive rate analyzed/C4 (74,35%) was more dominant than evaluating/C5 (68,90%) and create/C6 (63,24%). The results of learning observations show that the implementation of the stages of learning to improve HOTS has been fulfilled with an average value of 72,22 (Good). Thus, a high-order thinking skill profile is influenced by high learner-learning motivation. Students are accustomed to analyzing, evaluating, and creating during the learning process, right teacher learning strategies, HOTS-based learning processes, and conducive learning environments. The results also prove that the second-level questions on the TTMC instrument make students think at a higher level and have reasoning abilities. The

limitation of the research was many students who are not used to doing tests with the TTMC instrument, and the time for making the instrument was quite long. The advice that can be given is that students should be trained and accustomed to working on TTMC questions, and teachers can provide learning that comes from problem analysis. Future researchers are expected to be able to prepare the time for making the instrument long enough so that the results obtained are more in-depth.

REFERENCES

- Aini, K., & Ridwan, M. (2021). Students' Higher Order Thinking Skills Through Integrating Learning Cycle 5e Management With Islamic Values in Elementary School. *Al-Tanzim: Jurnal Manajemen Pendidikan Islam*, 05(03), 142-156. <http://doi.org/10.33650/al-tanzim.v5i3.3042>
- Adodo, S. O. (2013). Effects of Two-Tier Multiple Choice Diagnostic Assessment Items on Students' Learning Outcome in Basic Science Technology (BST). *Academic Journal of Interdisciplinary Studies*, 2(2), 201-210. <http://dx.doi.org/10.5901/ajis.2013.v2n2p201>
- Anazifa, R. D., & Djukri, D. (2017). Project-Based Learning and Problem-Based Learning: Are They Effective to Improve Student's Thinking Skills?. *Jurnal Pendidikan IPA Indonesia*, 6(2), 346-355. <https://doi.org/10.15294/jpii.v6i2.11110>
- Anderson, L.W. & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assesing: A Revision of Bloom's Taxonomy of Educatioanl Objectives*. New York: Addison Wesley Longman.
- Apino, E., & Retnawati, H. (2017). Developing instructional design to improve mathematical higher order thinking skills of students. *Journal of Physics: Conference Series*, 812, 1–7. doi:10.1088/1742-6596/755/1/011001.
- Bayrak, B. K. (2013). Using Two Tier Test to Identify Primary Student's Conceptual Understanding and Alternative Conception in Acid Base. *Mevlana International Journal of Education*, 3(2), 19-26. <http://dx.doi.org/10.13054/mije.13.21.3.2>
- Bhattacharya, D., & Mohalik, R. (2021). Factors Influencing Students' Higher Order Thinking Skills Development. *Education India Journal : A Quarterly Refereed Journal of Dialogues on Education*, 10(1), 349–361.
- Brookhart, S. M. (2010). *How to Assess Higher Order Thinking Skills in Your Classroom*. Alexandria: ASCD.
- Cullinane, A. (2011). Two-tier Multiple Choice Questions (MCQs) - How effective are they. *International Journal of Science & Technology Education*, 7(1), 611–624.
- Fayakun, M. & Joko, P. (2015). Efektivitas Pembelajaran Fisika Menggunakan Model Kontekstual (CTL) dengan Metode Predict, Observe, Explain Terhadap Kemampuan Berpikir Tingkat Tinggi. *Jurnal Pendidikan Fisika Indonesia*, 11(1), 49-58. <https://doi.org/10.15294/jpfi.v11i1.4003>

- Frydenberg, M. E., & Andone, D. (2011). Learning for 21st Century Skills. *IEEE's International Conference on Information Society*. 314–318. <https://doi.org/10.1109/ISOCIETY18435.2011.5978460>
- Lesmana, A. D. (2016). *Identifikasi Profil Kemampuan Berpikir Tingkat Tinggi Siswa SMA Menggunakan Instrumen Two-Tier Tes pada Mata Pelajaran Fisika*, Skripsi UIN Syarif Hidayatullah Jakarta.
- Mandernach, B. J., Forrest, K. D., Babutzke, J. L., & Manker, L. R. (2009). The role of instructor interactivity in promoting critical thinking in online and face-to-face classrooms. *Merlot Journal of Online Learning and Teaching*, 5(1), 49-62. https://jolt.merlot.org/vol5no1/mandernach_0309.pdf
- Mislia, T. S., Indartono, S., & Mallisa, V. (2019). Improving Critical Thinking among Junior High School Students through Assessment of Higher Level Thinking Skills. *Advances in Social Science, Education and Humanities Research*, 323, 326–333. <https://doi.org/10.2991/icosce-icsmc-18.2019.58>
- Mukhtar, M. & Haniin, K. (2019). *Modul Penyusunan Soal Keterampilan Berpikir Tingkat Tinggi (Higher Order Thinking Skills)*. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- Nabilah., Andayani, Y., & Laksmiwati, D. (2014). Analisis Tingkat Pemahaman Konsep Siswa Kelas XI IPA SMAN 3 Mataram Menggunakan *One Tier* dan *Two Tier Test* Materi Kelarutan dan Hasil Kali Kelarutan. *Jurnal Pijar MIPA*, 8(2), 64-69. <https://doi.org/10.29303/jpm.v8i2.8>
- Nofiana, M., Sajidan., & Puguh. (2016). Pengembangan Instrumen Evaluasi *Higher Order Thinking Skills* pada Materi Kingdom Plantae. *Jurnal Pedagogi Hayati*, 1(1), 46-53. <https://doi.org/10.31629/ph.v1i1.37>
- Noma, L. D., Prayitno, A. B., & Suwarno. (2016). PBL untuk Meningkatkan Kemampuan Berpikir Tingkat Tinggi Siswa Kelas X SMA. *Bioedukasi*, 9(2), 62-66.
- Organization for Economic Cooperation and Development. (2019). *Country Note: Results from PISA 2018 (Indonesia)*. (<https://www.eocd.org.edu/pisa>).
- Pratiwi, N. & Mustadi, A. (2021). Hots-Based Learning in 2013 Curriculum: Is it Suitable?. *Jurnal Pendidikan Indonesia*, 10(1), 128-135. <http://dx.doi.org/10.23887/jpi-undiksha.v10i1.22781>
- Rahayu, D., Sabar, N., Achmad, F., Laili, E. R., & Choiriyah, W. (2020). Pembentukan Karakter Siswa Berorientasi Higher Order Thinking Skills (HOTS) di Sekolah Dasar. *ELSE (Elementary School Education Journal)*, 4(1), 109–118. <http://dx.doi.org/10.30651/else.v4i1.4071>
- Retnawati, H., Djidu, H., Kartianom, Apino, E., Anazifa, R. D. Teachers' Knowledge About Higher-Order Thinking Skills And Its Learning Strategy. *Problems of Education in the 21st Century*, 76(2), 215-230.

- Saltan, F. & Divarci, O. F. (2017). Using blogs to improve elementary school students' environmental literacy in science class. *European Journal of Educational Research*, 6(3), 347–355. <https://doi.org/10.12973/eu-jer.6.3.347>
- Septiana, D., Zulfiani., & Noor, M. F. (2014). Identifikasi Miskonsepsi Siswa pada Konsep Archaeobacteria dan Eubacteria Menggunakan *Two-Tier Multiple Choice*. *EDUSAINS*, VI(2), 190-200. <https://doi.org/10.15408/es.v6i2.1151>
- Sesli, E., & Kara, Y. (2012). Development and application of a two-tier multiple-choice diagnostic test for high school students' understanding of cell division and reproduction. *Journal of Biological Education*, 46(4), 214-225. <https://doi.org/10.1080/00219266.2012.688849>
- Shidiq, A. S., Masykuri, M., & Susanti, E. (2014). Pengembangan Instrumen Penilaian *Two-Tier Multiple Choice* untuk Mengukur Keterampilan Berpikir Tingkat Tinggi (*Higher Order Thinking Skills*) pada Materi Kelarutan dan Hasil Kali Kelarutan untuk Siswa SMA/MA Kelas XI. *Jurnal Pendidikan Kimia (JPK)*, 3(4), 83-92.
- Talanquer, V. (2006). Commonsense Chemistry: A Model for Understanding Students' Alternative Conceptions. *Journal of Chemical Education*, 83. <http://dx.doi.org/10.1021/ed083p811>
- TIMSS & PIRLS International Study Center. (2015). *TIMSS 2015 International Results in Science, Chapter I: Student Achievement*. Boston: The TIMSS & PIRLS International Study Center, Boston Collage. <https://timss2015.org/download-center>
- Treagust, D. F. (2006). Diagnostic Assessment in Science as a Means to Improving Teaching, Learning, and Retention. *UniServe Science Assessment Symposium Proceedings*, 1-9.
- Treagust, D. F. (1988). Development and Use of Diagnostic Test to Evaluate Student's Misconception in Science. *Journal of Science Education*, 10(2), 159-169. <https://doi.org/10.1080/0950069880100204>
- Tsui, C.Y. & Treagust, D. (2010). Evaluating Secondary Students' Scientific Reasoning in Genetics Using a Two-Tier Diagnostik Instrumen. *International Journal of Science Education*, 32(8), 1073-1098. <https://doi.org/10.1080/09500690902951429>
- Tuysuz, C. (2009). Development of Two Tier Diagnostic Instrument and Assess Students' Understanding in Chemistry. *Scientific Research and Essay*, 4(6), 626-631.
- Wahyuni. D. E. & Arief, A. (2015). Implementasi Pembelajaran *Scientific Approach* dengan Soal *Higher Order Thinking Skill* pada Materi Alat-alat Optik Kelas X di SMA Nahdlatul Ulama' 1 Gresik. *Jurnal Inovasi Pendidikan Fisika (JIPF)*, 4(3), 32-37. <https://doi.org/10.26740/ipf.v4n3.p%25p>
- Wartono, W., Hudha, M. N., & Batlolona, J. R. (2018). How Are The Physics Critical Thinking Skills of The Students Taught by Using InquiryDiscovery Through Empirical and Theorethical Overview?. *Eurasia Journal of Mathematics, Science and Technology Education*, 14(2), 691-697. <https://doi.org/10.12973/ejmste/80632>

- Widana, I. W. (2017). Higher Order Thinking Skills Assessment (Hots). *Jisae: Journal of Indonesian Student Assesment and Evaluation*, 3(1), 32–44. <https://doi.org/10.21009/jisae.031.04>
- Yuniarti, N. N., Susanto, & Irvan, M. (2018). The Implementation of Scientific Approach with Project Based Learning Model to Improve Students' Analytical Skill and Responds. *International Journal of Advanced Engineering Research and Science*, 5(10), 90-93. <https://dx.doi.org/10.22161/ijaers.5.10.11>
- Zulfiani., Yunistika R., & Juanengsih, N. (2018). Enhancing Students' Higher-Order Thinking Skills Through Guided and Free Inquiry-Based Learning. *Advances in Social Science, Education and Humanities Research "3rd International Conferences on Education in Muslim Society"*, 115. <https://dx.doi.org/10.2991/icems-17.2018.6>
- Zulfiani., Herlanti, Y., & Nurrachman, L. (2015). Development of a Higher-Order Thinking Instrument for Use with X Grade Senior High School Students. *Proceedings of the Eighth International Conference on Science, Mathematics, Engineering and Technology Education*, 427-433.