



## Development of Higher Order Thinking Skill Qualified Test on Colligative Properties of Solutions

Rahielanamy\*, Agus Kamaludin

Department of Chemistry Education, Faculty of Tarbiyah and Teacher Training, State Islamic  
University Sunan Kalijaga Yogyakarta, Yogyakarta, Indonesia

\*Email: rahielanamy.uinsuka@gmail.com

DOI: 10.24815/jpsi.v10i2.23968

### Article History:

Received: December 23, 2021  
Accepted: March 10, 2022

Revised: February 16, 2022  
Published: March 25, 2022

**Abstract.** Higher-order thinking skills (HOTS) become one of the major focuses of 21st-century learning. Those skills can be trained through stimulus in the learning process and assessment that uses hots-based questions. The research aims to generate higher-order thinking skills qualified questions on colligative properties of solutions that fulfill validity and reliability instrument. This research uses the Borg & Gall development model that consists of data collection, product planning, expert validation, limited trial, and final product improvement stages. The data collection procedure consisted of a study of documentation, validation sheets, and tests. Validation sheets are used to carry out validation tests and product quality assessments by experts and reviewers. The test was used to conduct a limited trial by 40 students of class XII MAN Temanggung. The test subjects were selected using a purposive sampling technique. The results of this study were 20 HOTS-qualified multiple-choice questions on the colligative properties of solutions that have very good quality by the experts and reviewers' judgment. Item analysis based on the results of the limited test on 40 students showed 17 items (85%) were valid with a reliability value of 0.82 (very strong); the difficulty level of the items is 15% easy, 75% medium, 10% difficult and; the distinguishing power of questions obtained 85% good categories. It can be concluded that the HOTS-qualified questions for the material of Colligative Properties of Solutions have good quality and are suitable for use as an assessment instrument.

**Keywords:** question, higher-order thinking skills (hots), colligative properties of solutions

## Introduction

Higher-order thinking skills (HOTS) become an indicator of the success of learning in the 21<sup>st</sup> century (Tan & Halili, 2015). HOTS is the main component in 21st-century learning that has to be owned by students (Monari, 2020; Cox, 2019; Chalkiadaki, 2018; Quieng et al., 2015). HOTS is the ability to solve the problem through the association of new knowledge and old knowledge by connecting several facts and then converting them into new solutions (Nursalam et al., 2018). Through HOTS, students can distinguish ideas clearly, deliver arguments well, solve problems, construct explanations, make hypotheses, and understand complexities, even preparing students to become superior and successful individuals in their careers (Dinni, 2018; Widiawati et al., 2018). However, the HOTS of students in Indonesia still needs to be improved. Based on Programme for International Student Assessment (PISA) which uses HOTS-oriented questions as to its instrument, in 2018, Indonesia was ranked 72<sup>nd</sup> out of 77 participating countries (OECD, 2019).

The implementation of the 2013 Curriculum is an effort by the Indonesian government to improve the quality of students, especially HOTS (Kurniawan & Lestari, 2019). There are improvements to the curriculum components in the 2013 curriculum, such as content standards and assessment standards which focus on HOTS (Astutik, 2018). Assessment standard in the 2013 curriculum is established by adapting the international standard assessment model, one of which is the HOTS-oriented assessment model (Setiawan et al., 2021; Widana, 2017; Chinedu et al., 2015; Wang & Wang, 2014). By the existence of HOTS, it is expected that students can have critical and creative thinking skills and understand learning materials more deeply (Rintayati et al., 2020; Widana, 2017). The government has implemented the application of HOTS-based assessment by increasing the standard of HOTS-based national examination questions starting in 2016 (Sumaryanta, 2018). The average result of the national examination in 2019 for high school majoring in science was 53.16 (Puspendik, 2019). The National Examination average of 53.16 is still very low (D category), so it needs to be increased. This is also done with research conducted by Ichsan et al., (2019), which shows that most students have low HOTS abilities, so they need to be improved through various media developments, materials, models, and HOTS-based learning strategies. However, it is still a big challenge for some teachers to find a suitable learning model in creating HOTS in students (Husamah, et al., 2018).

HOTS can be practiced and developed through learning activities in the classroom, starting from the learning process to learning evaluation (Wiyaka et al., 2020; Abosalem, 2015; Chinedu et al., 2015). Tajularipin et al., (2017) stated that HOTS is an ability that must exist in every learning. In this case, the teacher plays an essential role in stimulating students' thinking by learning and questions that can improve students' HOTS abilities (Kurniawan & Lestari, 2019; Tajularipin Sulaiman et al., 2017). Teachers should often provide HOTS practice questions to improve students' higher-order thinking skills (Putri et al., 2019). Students' HOTS abilities must also be practiced by providing HOTS-based learning resources, planning and implementing cooperative and collaborative learning (Widyaningsih et al., 2021; Warner & Kaur, 2017). Driana & Ernawati (2019) stated that the success of developing students' HOTS by the teacher could be seen from the results of the assessment used by the teacher. However, most teachers are still not optimal in developing HOTS-based assessments because they lack understanding of HOTS, and they also have difficulty formulating indicators used in HOTS-based assessment instruments (Widyaningsih et al., 2021; Retnawati et al., 2018; Salirawati et al., 2017). Based on interviews with chemistry teachers in Temanggung Regency, to compile HOTS-loaded questions, references are required as a matrix for developing questions used in the assessment. According to the findings of Wiyaka et al., (2020), there is little research and literature on HOTS-based evaluation that can be utilized as a reference. As a result, teachers in schools tend to use questions that test the aspect of memory so that students are less trained to solve questions that measure higher-order thinking skills (Yuliati & Lestari, 2018; Budiman & Jailani, 2014). Therefore, developing an assessment instrument that can measure students' higher-order thinking skills is very needed (Ichsan et al., 2019; Tanujaya, 2015).

The HOTS-based assessment instrument is an assessment instrument that involves the ability to analyze (C4), evaluate (C5), and create (C6) (Setiawan et al., 2021; Hanifah, 2019). HOTS-qualified questions can stimulate students to understand the lesson more deeply so that they are not only able to apply, analyze and evaluate but will also understand the factual knowledge being taught better (Jensen et al., 2014). These HOTS-qualified questions can help train and improve students' HOTS, as well as increase students' self-potential (Giatri et al., 2021; Setiawan et al., 2021). HOTS question is certainly different

from questions in general, HOTS question has some characteristics, such as measuring high-level abilities and it is based on contextual problems (Widana, 2017).

The most prominent characteristic of the HOTS question is contextual problem-based stimulus. This means that the question can refer to real phenomena in life so that students can apply learning concepts in class to solve problems (Widana, 2017). As one of the sciences that is very close to life, chemistry is one of the sciences that require HOTS because it deals with the representation of phenomena at the macroscopic level. Macroscopic is a chemical representation obtained from real observation of a phenomenon that is seen and perceived by the five senses or it can be in the form of everyday experiences (Yanto et al., 2013). One of the materials that are close to everyday life and involve macroscopic representation is the colligative properties of solutions. The proximity of the colligative properties of solutions with everyday life makes this material important to be studied by using a deeper level of thinking.

These problems can be solved by developing a valid and reliable HOTS-based test instrument. The results of research conducted by Hardiani (2017) states that the process of assessing student learning outcomes requires valid and reliable instruments so that learning objectives can be achieved optimally. The application of the development of the HOTS test instrument can develop HOTS and provide examples of HOTS form questions to teachers. This research was carried out to generate HOTS-qualified questions on the colligative properties of solutions that are valid and reliable and also can be used as an assessment instrument to measure the level of thinking of students.

## Methods

The research is development research (R & D) that adapts the Research and Development model by Borg & Gall which is modified according to the objectives and interests of the research. Five stages were applied in the research, such as data collection, product design followed by product development, expert validation test, limited tryout, and improvement of the final product.

The product in the form of HOTS-qualified questions on the colligative properties of the solution was validated by one material expert lecturer and one media expert lecturer, then improvements were made based on experts' advice and inputs. The first improved product was assessed by reviewers, five high school chemistry teachers, then the second improvement was carried out. The second improved product was then tested restrictedly towards 40 students of class XII MIPA of MAN Temanggung, Central Java, Indonesia. The limited tryout in this research was conducted on April 20, 2021, online via Google Forms due to the Covid-19 pandemic.

The data collection procedure in this study consisted of a study of documentation, validation sheets, and tests. Validation sheets were used to carry out validation tests and product quality assessments by experts and reviewers. Meanwhile, the test was used to conduct a limited trial by 40 students of class XII MAN Temanggung. The test instrument is in the form of 20 multiple choice questions with five answer choices each about the colligative properties of the solution. The test subjects in this study were selected using a purposive sampling technique. According to Sugiyono (2015), purposive sampling is a sampling technique with certain considerations. In this study, sampling was assisted by a chemistry teacher for class XII at MAN Temanggung.

Data analysis was used to obtain a valid and reliable HOTS-based assessment instrument. The data analysis technique in this study consisted of qualitative data analysis from validation sheets and quantitative analysis of HOTS questions based on test results. Validation of data from experts and reviewers using a qualitative descriptive method. This means that input in the form of expert advice and reviewers is used as a reference for product improvement before limited testing is carried out. A qualitative analysis of the expert assessment was carried out to determine the feasibility and quality of HOTS questions in terms of material, construction, language, and characteristics of HOTS questions. The data obtained is in the form of a quality assessment in the form of letters (Very Poor (VP) - Very Good (VG)) which is then converted into numbers using the provisions of the Likert scale with a score of 1 to 5, with value interpretation; one very poor; two less; three sufficient; four is good; and five is very good (Setiawan et al., 2021). Then the average value is calculated to determine the feasibility of the questions. The feasibility category for the product is interpreted according to Table 1. (Arikunto, 2013; Ernawati, 2017):

Table 1. Product feasibility criteria

| No. | Score (%) | Feasibility Category |
|-----|-----------|----------------------|
| 1   | <21       | Very poor            |
| 2   | 21 – 40   | Poor                 |
| 3   | 41 – 60   | Sufficient           |
| 4   | 61 – 80   | Good                 |
| 5   | 81 – 100  | Very good            |

The data was obtained from the reviewers' assessment (five high school chemistry teachers) in the form of numbers 1 - 5. The data obtained were analyzed using Aiken's V formula to calculate the content validity of an item. Aiken's V formula is as follows (Aiken, 1985):

$$V = \frac{\sum s}{[n(c-1)]}$$

Remark:

- s = r – I<sub>o</sub>
- r = the number given by the rater
- I<sub>o</sub> = the minimum validity score
- c = the maximum validity score
- n = number of raters

The range of V is 0 to 1. The higher the number of V (approaching 1 or equal to 1), the higher the validity of the item (Aiken, 1985). According to Retnawati (2016) if the number of  $V \geq 0,8$  means the item is accepted and is considered very valid.

In further, analysis quantitative of HOTS questions based on the results of limited tryout was carried out with the help of the IBM SPSS 20 program and the AnBuSo 8.0 program. The analysis of questions was carried out to find out the characteristics of the HOTS questions which included:

#### Validity test

The validity of the product, HOTS questions totaling 20 items, was calculated by using the IBM SPSS 20 program. After that, it was interpreted with rtable at a significance level of 5% with N = 40. The decision is taken with the condition that if  $r_{count} > r_{table}$ , the

item is considered valid. Whereas, if  $r_{\text{count}} < r_{\text{table}}$ , the item is considered invalid (Sugiyono, 2015).

#### Reliability test

The reliability test of this instrument used an internal consistency estimation technique by using the Cronbach-alpha formula, assisted by the SPSS IBM 20 application. The questions are considered to be reliable if the alpha value is  $\geq 0,6$  (Sugiyono, 2015).

#### Level of difficulty

The difficulty level of the items aimed to find out how big the difficulty level of the items that have been made is. To measure the difficulty level of the items, three categories of questions, which are easy, medium, and hard were made is. The difficulty level is calculated by the formula of difficulty index (Brookhart & Nitko, 2011):

$$(P) = NP / N$$

Remark:

(P) = difficulty index

NP = the number of students who answered the questions correctly

N = number of students taking the test

The analysis was carried out with the assistance of the AnBuSo 8.0 program. Through analysis, the question criteria are made based on difficulty index (P) which is provided the information that 0.00 – 0.30 is categorized hard; 0.31 – 0.70 is categorized medium; and 0.71 – 1.00 is categorized easy (Setiawan et al., 2021).

#### Distinguishing Power

Analysis of distinguishing power of multiple-choice questions can be calculated by the formula (Brookhart & Nitko, 2011),

$$DP = \frac{JB_B - JB_A}{n}$$

Remark:

DP = differentiating question

$JB_B$  = number of students wrong answer from under the group

$JB_A$  = number of students wrong answer from one group

n = 27%  $\times$  N

After the calculation, the questions were categorized into items that are accepted, revised, and rejected. The category of distinguishing power of questions was made based on the following criteria; if  $DP > 0.3$ , the question is good /accepted; if  $0.2 < DP < 0.3$ , the question is fair /accepted with improvement; and if  $DP < 0.2$ , the question is poor/rejected. The analysis to determine the distinguishing power was carried out with the assistance of the AnBuSo 8.0 program.

## Results and Discussion

The development of HOTS qualified chemistry questions on the colligative properties of solutions is designed using the Borg & Gall model which is limited to five stages of development. The five stages consist of data collection, product design, expert validation, limited trial, and final product improvement.

Literature study, field study, and material analysis were carried out at the data collection stage (Sukmadinata, 2017). A literature study was carried out by reviewing relevant previous research literature, and the results showed that HOTS is the main goal in education and is one of the top five variables that can improve student achievement (Pratama & Retnawati, 2019). According to Arifin & Retnawati (2017), the potential for higher-order thinking of students can be developed through assessment activities using HOTS-based questions. In a field study conducted by interviewing three high school chemistry teachers in Temanggung Regency, HOTS-based questions are still not widely used because of difficulties in compiling them. It coincides with the results of research by Saraswati & Agustika (2020), that assessment instruments commonly used in schools such as daily tests or grade ascension exam, questions at level C1 - C3 (LOTS) still dominate. The results of research conducted by Widarta & Artika (2021) regarding the analysis of the characteristics of champions on evaluation instruments stated that most of the questions made by the teacher did not have a stimulus, the distribution of cognitive process dimensions on the questions was uneven, and the questions made were not HOTS questions. Kurniasi, et al., (2019) said that the difficulty of compiling HOTS questions is caused by the lack of availability of references about HOTS. Therefore, more HOTS-based assessment instruments needed to be developed (Sa'adah, et al., 2019; Nurulwati, 2019). Material analysis by analyzing opportunities for basic competencies to form HOTS questions seen from operational verbs in basic competencies. Basic competencies 3.1 and 3.2 on the material colligative properties of solutions use the verb "analyze and distinguish" both of which belong to the cognitive aspect of analyzing (C4) according to Bloom's taxonomy. Therefore, the colligative properties of the solution have a high chance of forming HOTS questions. It coincides with research conducted by Khaldun, et al. (2019) that basic competencies with the cognitive level of analyzing are competencies that require high-level thinking skills so HOTS questions can be made.

The HOTS questions produced in this study amounted to 20 multiple-choice questions according to the characteristics of the HOTS questions. Multiple-choice questions can be used as a HOTS-qualified instrument to measure micro-dimensional critical thinking skills such as making the most reasonable assumptions, seeing the author of the goal, choosing the most appropriate conclusions. The HOTS questions are arranged based on the grid that has been made (King, et al., 2010; Mustika, et al., 2018). The grid is important in making questions because it can make it easier to write the questions and improve the quality of questions as an assessment instrument (Werdayanti, 2008; Leksono, et al., 2013; Fahdini, et al., 2014; April, 2019). After the product have been compiled are then assessed by validators (material experts and media experts) to determine their feasibility. The level of product feasibility can be determined based on the assessment of experts (Ramadhan, et al., 2019). After the expert's validation, the researcher made improvements to the draft questions based on suggestions or input from the validator. The results of the draft improvement were assessed by five high school chemistry teachers in Temanggung Regency for further analysis of the content validity before conducting a limited trial.

The valid HOTS questions were then tested on the 12<sup>th</sup>-grade students of MAN Temanggung who had obtained the colligative properties of the solution. The trial was

carried out on April 20, 2021, online via google form due to the pandemic. One of the developed questions in this research can be seen in Figure 1.



Radiator water is a special fluid in motorized vehicles that cool the engine temperature so as not to overheat. In areas with shallow temperatures, radiator water will quickly freeze to cause damage to motorized vehicles' components. To prevent car radiator water from freezing, adding a liquid that is difficult to freeze is necessary. A car radiator containing 5 liters of water is in a wintering area with a  $-20^{\circ}\text{C}$ . To prevent the car radiator water from freezing, the thing to do is.... ( $K_f$  of water =  $1.86^{\circ}\text{C/m}$ )

- A. increase the volume of radiator water used to four times the initial volume
- B. reduce the radiator water by 2 liters
- C. add 1.5 liters of 2.5 M hydrochloric acid to the radiator water
- D. add 3 liters of ethylene glycol ( $\rho = 1.11 \text{ g/mL}$ ) to the radiator water
- E. add 1.0 mole of ethylene glycol ( $\text{C}_2\text{H}_6\text{O}_2$ ) solution in the radiator water

Figure 1. HOTS question on colligative properties of solutions

The question in Figure 1 is the HOTS question because it conforms with the HOTS criteria. The question uses contextual stimuli and it is educational for students. According to Laily & Wisudawati (2015), a stimulus needs to be present in each item to be able to bring up the HOTS ability. HOTS questions involving contextual stimuli are expected to be able to train students to solve problems using their reasoning and logic (Saraswati & Agustika, 2020). The question also measures students' abilities to process, apply, and use the information of stimuli which is in the form of reading. In addition, the development questions can be used as a test assessment instrument that measures the high-level thinking skills of high school students in the matter of colligative properties of solutions. This is reinforced by the opinion of Rahayu, et al., (2020) that a question that reflects the HOTS question must have the characteristics of being able to regulate higher-order and problem-based thinking skills that are presented contextually in the form of case studies or tables.

HOTS-qualified assessment instruments can increase the students' potency and encourage them to think more deeply about the material being studied (Jensen et al., 2014; Setiawan et al., 2021). According to Widarta & Artika (2021) student who are accustomed to resolving problems requiring hots are expected to solve life's problems. Teachers need HOTS questions because many teachers have not made HOTS questions (Armianti et al., 2020).

## Product Validation

The product consisting of HOTS-qualified questions on the colligative properties of solutions is validated by a material expert lecturer, a media expert lecturer, and five high school chemistry teachers as reviewers. The validation aims to know the feasibility of the product (Ibrahim & Ishartiwi, 2017; Dewi, et al., 2018; Putri, et al., 2022). Expert validation is an important role in developing HOTS questions before analyzing the items through limited trials. This aims to see the quality of the initial product from the HOTS qualification questions which are then improved based on input and suggestions from experts and reviewers (Sa'idah, et al., 2018). Product validation by experts was carried out once. It was validated by one material expert and one media expert. Validation data from experts, in the form of criticism and suggestions, are used as a reference for product improvement. According to the material expert, several questions need to be improved, especially on questions discussing the concentration of substances. Meanwhile, according to the media expert, all questions in the point of construction, graphics, and language are good and they do not need to be improved. In addition to criticism and suggestions, experts also provide quality assessments which were then converted into numbers. Furthermore, the percentage of the average score of each question in the assessment aspects was analyzed. The analysis results of the validation and the assessment of media and material experts can be seen in Figure 2 and Figure 3.

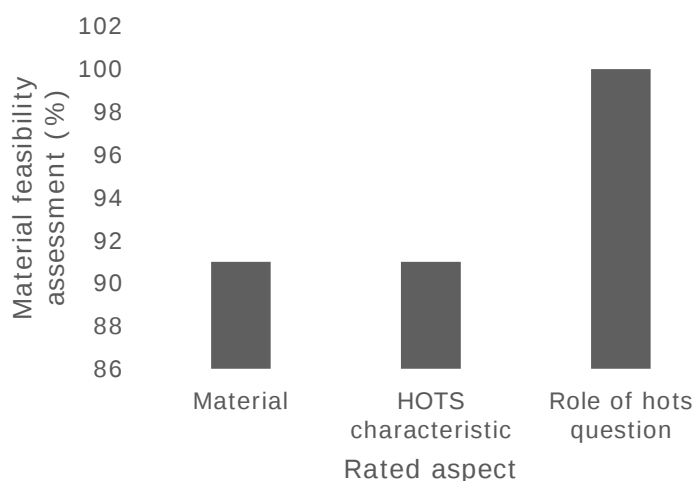


Figure 2. Analysis result of material expert's validation

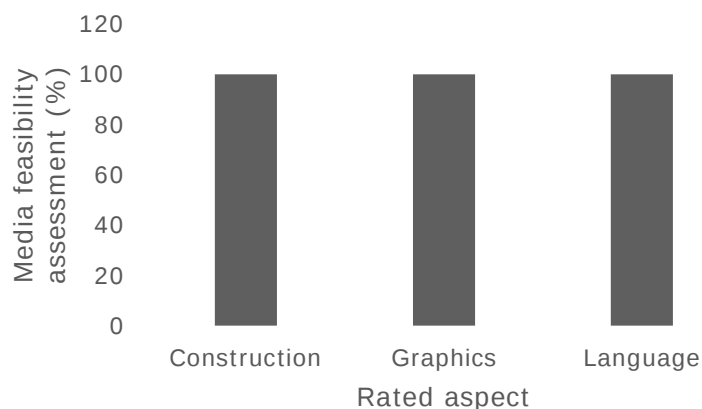


Figure 3. Analysis result of media expert's validation

Based on the material expert's assessment, the average percentage for the material aspect is 91%, the characteristic of the HOTS-question aspect is 91%, and the role of the HOTS-question aspect is 100%. Overall, an average score of 94% is obtained from validation by a material expert. Furthermore, assessment by media experts involves three aspects, those are aspects of construction, graphics, and language. In those aspects, an average score of 100% is gained. It means that the questions developed are in the very good category (Purnamasari, 2015). This shows that the product is very feasible to be used as a test instrument to measure students' high-level skills. These results are in line with research conducted by Yuliantaningrum & Sunarti (2020), the development of a HOTS question instrument to measure critical thinking, creative thinking, and problem solving is feasible to use with material, language, and construct validity values of 83.33%, 89.85 %, and 92.42% in the category of very suitable for use.

The validation and assessment of reviewers (five high school chemistry teachers in Temanggung Regency) aim to determine the content validity of the developed product. Content validity is the validity that is estimated through testing against the feasibility of test content through rational analysis through expert judgment (Azwar, 2013). The assessment is carried out by giving a score between 1 (very unrepresentative or highly irrelevant) and 5 (very representative or very relevant). The scores obtained from the results of the reviewer's assessment were analyzed using Aiken's validity analysis. The results of the analysis can be categorized as valid if they meet the coefficient limit Aiken's V. The boundary conditions for the Aiken's V coefficient for 5 rating scales and 5 raters are 0.80 with a probability of 0.40 (Aiken, 1985). The results of the reviewer's assessment analysis can be seen in Figure 4.

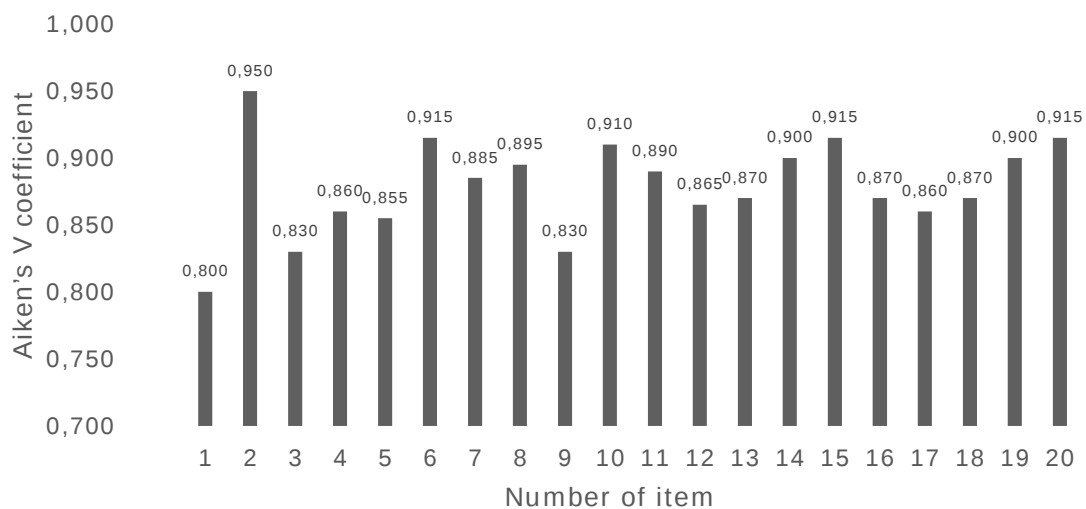


Figure 4. The result of the reviewer's assessment

The reviewer's content validation analysis shows that the overall coefficient of each item is greater than or equal to the valid criteria of 0.80. The test results show that the overall instrument items developed with 20 valid items with an average coefficient level of 0.879. Based on these results if the validity coefficient  $\geq 0.80$  means the item can be said to be valid (Retnawati, 2016). This is line with the research result of Qirom, et al., (2021) on the development of HOTS-questions, the items developed are valid with Aiken's V value 0,759.

### Analysis Data of Limited Trial

The limited trial was carried out on April 20, 2021, via Google Forms, and involved 40 students of class XII MIPA of MAN Temanggung. The data of this limited trial are used for item analysis. The research of Warju, et al., (2020) stated that item analysis holds an important role in determining the validity of the questions and producing good quality items. This is a line with Widarta & Artika (2021) that the quality of the evaluation instruments that have been made needs to be controlled through analysis on each item. The result of limited tryout was analyzed to determine the quality of the HOTS questions according to the characteristics of the questions as follows:

#### Question Validity

Analysis of multiple-choice questions was calculated by using the SPSS IBM25 program to test the validity of questions according to the limited trial which has been held. Arifin (2017) said that a good and appropriate assessment instrument must have two important requirements, valid and reliable. Furthermore, the validity was interpreted by  $r_{table}$  at a significance level of 5% with  $N = 40$ . Then, a  $r_{table}$  of 0.312 was obtained and there were 17 (85%) valid questions and 3 (15%) invalid items. The invalid items are questions number 3, 9, and 10. These invalid items need to be improved to be used as test questions (Khalidun, et al., 2019). The result of question validation is shown in Figure 5.

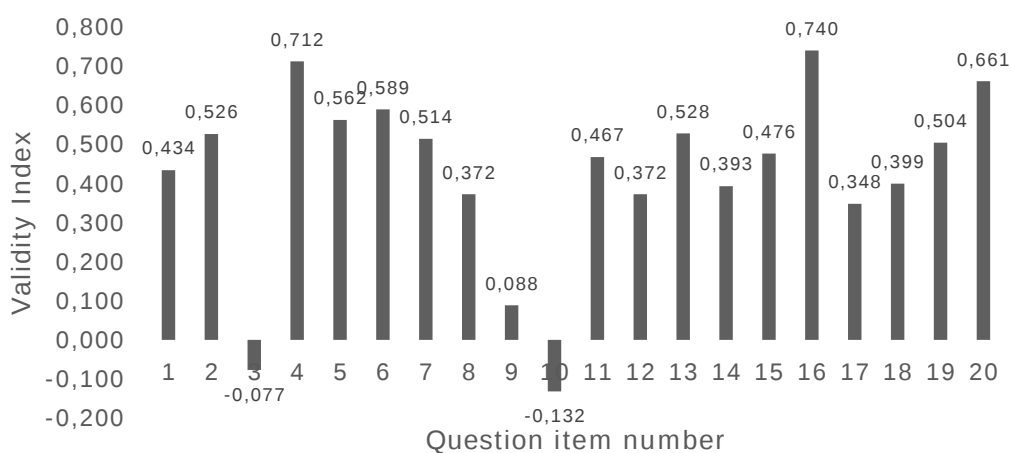


Figure 5. The result of the validity test

#### Question Reliability

Instrument reliability was calculated using the Cronbach-Alfa formula with the help of SPSS IBM 25. According to Hayati & Lailatussaadah (2016), if the value of Cronbach-Alfa is more than 0.7, it means that the instrument can be used and is reliable. The calculation results show that the question reliability is 0.82 on 17 valid questions, which means that the questions have a very strong level of reliability (Sudijono, 2013). These results are in accordance with the results of research by Khalidun, et al., (2019), the development of computer-based higher order thinking skills questions, has a reliability index of 0.763 which is included in the very high category. An assessment instrument with a high-reliability index can minimize errors so that the assessment becomes more accurate (Jamillah, 2021).

### Difficulty Level

The difficulty level of the developed HOTS question was analyzed by using the AnBuSo 8.0 program. The analysis shows that 3 (15%) questions have a difficulty index  $\leq 0,3$  which means that questions classified as the hard question; 15 questions (75%), that have a difficulty index  $0,31 - 0,70$ , are classified as medium and; 2 questions (10%) have a difficulty index  $\geq 0,71$  are easy questions (Setiawan, et al., 2021). The average level of difficulty in the medium category (0,55). The level of difficulty of the question is defined as a good question is a question with a medium category (Nuryanti, et al., 2018). In addition, Dewi, et al., (2021) the questions that are too difficult will make it difficult for students to answer and they do not want to try to solve the question, while questions that are too easy cannot provide motivation and stimulation for students to think deeply. The analysis result can be seen in Figure 6.

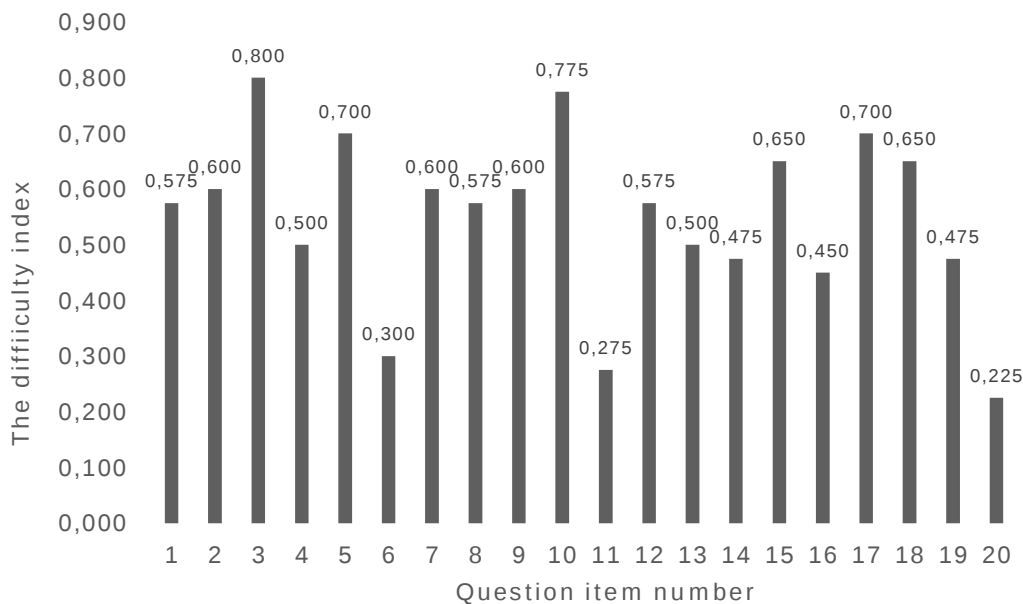


Figure 6. The analysis result of difficulty level

### Distinguishing Power

The distinguishing power of questions was analyzed by using the AnBuSo 8.0 program. According to Setiawan, et al., (2021), a question with a distinguishing power index of more than 0,3 has good distinguishing power,  $0,2 - 0,2$  means the distinguishing power is quite good, and  $< 0,2$  have poor distinguishing power. The analysis shows that 17 questions have a good distinguishing power index. It means the questions are accepted well without any improvement. Furthermore, three questions are rejected. Bhakti (2015) in his research stated that the distinguishing power is low because multiple-choice questions make students tend to guess the answer. Items that have good distinguishing power mean that they can be used to distinguish between high-ability students and low-ability students (Dewi, et al., 2021). The analysis result can be seen in Figure 7.

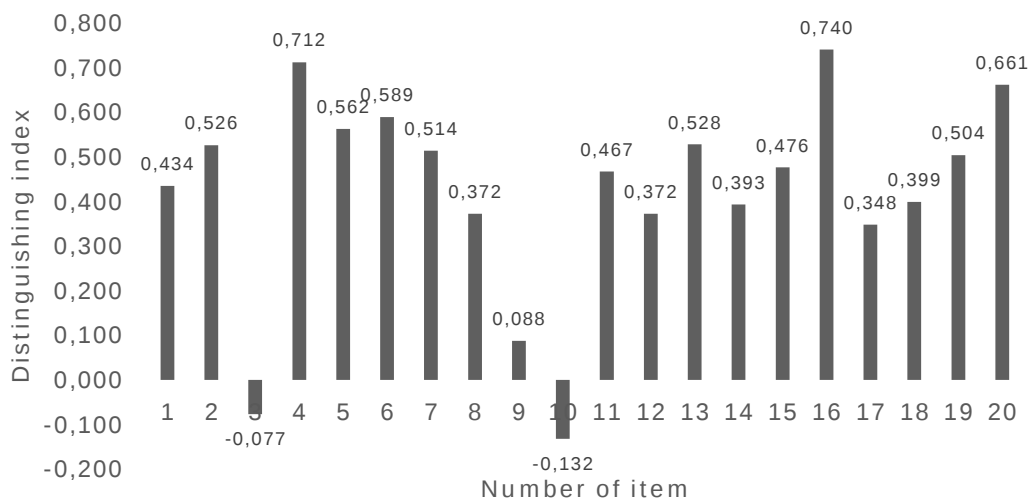


Figure 7. The analysis result of distinguishing power

This research has limitations, one of which is that the product development trial has not yet reached the stage of a wider field trial. This research is only tested on the limited trial stage involving one school, MAN Temanggung. The suggestion for further research is that a field trial involving more respondents should be conducted. In addition, research on the development of HOTS questions is still limited to one of the basic competencies of chemistry in class XII so further research is required to be able to develop a HOTS assessment instrument on other basic competencies. Any effort to support the improvement of students' higher-order thinking skills is very important to be done.

## Conclusion

Based on the results of research and data analysis on the development of a HOTS qualified assessment instrument on the material "colligative properties of solutions" it can be concluded that the feasibility of HOTS questions is based on expert and reviewer ratings, respectively, by 94% from material experts, 100% from media expert assessments. The content validity of the product based on the reviewer's assessment which was processed using the Aiken formula obtained an average Aiken's  $v$  coefficient of 0.879 which means the items developed are very feasible / very valid. This shows that the HOTS item is feasible to be used as an assessment instrument. Quantitative analysis on the HOTS questions that have been tested shows that there are three invalid questions out of the 20 questions with an average reliability level of 0.82 (very strong category). The HOTS questions on the colligative properties of the solution have met the standard of eligibility in terms of validity, reliability, difficulty level with an average score of 0.55 (medium), and distinguishing power with an average score of 0.42 (good). The HOTS questions can be used by students as training materials for higher-order thinking skills and as a reference for teachers in compiling and developing HOTS-based questions independently which will be implemented in learning.

## References

- Abosalem, Y. 2015. Assessment techniques and students' higher-order thinking skills. ICSIT 2018 - 9th International Conference on Society and Information Technologies, Proceedings, March, 61–66. <https://doi.org/10.11648/j.ijsedu.20160401.11>.
- Aiken, L.R. 1985. Three coefficients for analyzing the reliability and validity of ratings, Educational and Psychological Measurement, 45(1), 131-142. <http://doi.org/10.1177/0013164485451012>.
- April, S. 2019. Meningkatkan kemampuan guru dalam menyusun kisi-kisi soal dengan metode pendampingan pola "OCF" di SDN Yanti Jogoroto. JDMP (Jurnal Dinamika Manajemen Pendidikan), 4(1):17-24.
- Arifin, Z. 2017. Kriteria instrumen dalam suatu penelitian. Jurnal Theorems (The Original Research of Mathematics), 2(1):28–36.
- Arifin, Z. & Retnawati, H. 2017. Pengembangan instrumen pengukur higher order thinking skills matematika siswa SMA kelas X. Pythagoras: Jurnal Pendidikan Matematika, 12(1):79-85.
- Arikunto, S. 2013. Prosedur penelitian suatu pendekatan praktik. Jakarta: Rineka Cipta.
- Armianti, Subhan, M., Nasution, M.L., Azis, S.A., Rani, M.M., Rifandi, R., & Harisman, Y. 2020. Profesionalisme guru dalam membuat soal higher order thinking skills. Jurnal Pendidikan Matematika, 4(1):75-84. <http://dx.doi.org/10.33603/jnpm.v4i1.2587>.
- Astutik, P.P. 2018. Integrasi penguatan pendidikan karakter (PPK) dan higher order thinking skills (HOTS) dalam pembelajaran tematik SD. Journal of Education, 1(12):343–354.
- Azwar, S. 2014. Metode penelitian. Yogyakarta: Pustaka Pelajar.
- Bhakti, Y.B. 2015. Pengaruh jumlah alternatif jawaban dan teknik penskoran terhadap reliabilitas tes. Jurnal Formatif, 5(1):1–13.
- Borg, W.R. & Gall, M.D. 1983. Educational research: An introduction. New York: Longman Inc.
- Brookhart, S.M. & Nitko, A.J. 2011. Strategies for constructing assessments of higher order thinking skills. Assessment of Higher Order Thinking Skills, 1:327–59.
- Budiman, A. & Jailani, J. 2014. Pengembangan instrumen asesmen higher order thinking skill (HOTS) pada mata pelajaran matematika SMP kelas viii semester 1. Jurnal Riset Pendidikan Matematika, 1(2), 139. <https://doi.org/10.21831/jrpm.v1i2.2671>.
- Chalkiadaki, A. 2018. A systematic literature review of 21<sup>st</sup>-century skills and competencies in primary education. International Journal of Instruction, 11(3):1–16. <https://doi.org/10.12973/iji.2018.1131a>.

- Chinedu, C.C., Tun, U., Onn, H., & Kamin, Y. 2015. Strategies for improving higher order thinking skills in teaching and learning of design and technology education. *Journal of Technical Education and Training*, 7(2):35–43.
- Cox, J. 2019. Teaching strategies that enhance higher-order thinking. <http://www.teachhub.com/teaching-strategies-enhance-higher-order-thinking>.
- Dewi, P., Elvia, R., & Elvinawati, E. 2021. Pengembangan butir soal hots untuk menguji kemampuan berpikir tingkat tinggi siswa di MA Negeri 2 Kota Bengkulu. *Alotrop*, 5(2), 141-148.
- Dinni, H.N. 2018. HOTS (High Order Thinking Skills) dan kaitannya dengan kemampuan literasi matematika. *Prisma*, 1:170–176.
- Driana, E. & Ernawati. 2019. Teachers' understanding and practices in assessing higher order thinking skills at primary schools. *Acitya: Journal of Teaching & Education*, 1(2):110–118.
- Ernawati, I. 2017. Uji kelayakan media pembelajaran interaktif pada mata pelajaran administrasi server. *Elinvo (Electronics, Informatics, and Vocational Education)*, 2(2): 204-210.
- Fahdini, R., Mulyadi, E., Suhandani, D., & Julia. 2014. Identifikasi kompetensi guru sebagai cerminan profesionalisme tenaga pendidik di Kabupaten Sumedang. *Mimbar Sekolah Dasar*, 1(1):33–42.
- Giatri, R., Rochmiyati, & Pargito. 2021. Development of assessment instruments based on higher order thinking skills in thematic learning. *The International Journal of Social Sciences World*, 3(1):165-175.
- Hanifah, N. 2019. Pengembangan instrumen penilaian higher order thinking skill (hots) di sekolah dasar. *Conference Series*, 1(1):1–8. <http://ejournal.upi.edu/index.php/crecs/article/view/14286>.
- Hardiani, I. N. 2017. Pengembangan instrumen penilaian sikap sosial pembelajaran ips kelas iv sd. *E-journal Mitra Pendidikan*, 1(60):615-628. <https://doi.org/10.1017/CBO9781107415324.004>.
- Hayati, S. & Lailatussaadah. 2016. Validitas dan reliabilitas instrumen, pengetahuan pembelajaran aktif, kreatif, menyenangkan (pakem) menggunakan model Rasch. *Jurnal Ilmiah Didaktika*, 16(2):169-179.
- <http://puspendik.kemdikbud.go.id/hasil-un/>. Diakses pada tanggal 21 September 2019 pukul 22.29 WIB.
- Husamah, H., Fatmawati, D., & Setyawan, D. 2018. OIDDE learning model: Improving higher order thinking skills of biology teacher candidates. *International Journal of Instruction*, 11(2):249-264.

- Ibrahim, N. & Ishartiwi, I. 2017. Pengembangan media pembelajaran mobile learning berbasis android mata pelajaran IPA untuk siswa SMP. *Refleksi Edukatika: Jurnal Ilmiah Kependidikan*, 8(1):48-55.
- Ichsan, I.Z., Sigit, D.V., & Miarsyah, M. 2019. Environmental learning based on higher order thinking skills: a needs assessment. *International Journal for Educational and Vocational Studies*, 1(1):21-24.
- Jamillah, J. 2021. Analisis butir, pilihan, dan reliabilitas tes prestasi bahasa Inggris selama covid-19. *Jurnal Pendidikan*, 30(3):547-556.
- Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. 2014. Teaching to the test...or testing to teach: exams requiring higher order thinking skills encourage greater conceptual understanding. *Educ Psychol Rev*, (26):307–329. <https://doi.org/10.1007/s10648-013-92489>.
- Khaldun, I., Hanum, L., & Utami, S.D. 2019. Pengembangan soal kimia higher order thinking skills berbasis komputer dengan wondershare quiz creator materi hidrolisis garam dan larutan penyangga. *Jurnal Pendidikan Sains Indonesia*, 7(2):132-142.
- Kurniasi, E. R., Yopa, Y., & Karennisa, F. 2020. Analisis soal ulangan harian matematika kelas IX SMP Negeri 1 Toboali. *Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara Jakarta*, 12(1):43-52.
- Kurniawan, R.Y. & Lestari, D. 2019. The development assessment instruments of higher order thinking skills on economic subject. *Dinamika Pendidikan*, 14(1):102–115. <https://doi.org/10.15294/dp.v14i1.19226>.
- Lailly, N.R. & Wisudawati, A.W. 2015. Analisis soal tipe higher order thinking skill (HOTS) dalam soal un kimia sma rayon b tahun 2012/2013. *Jurnal Kaunia*, 11(1):27-39.
- Leksono, S. ., Rustaman, N., & Redjeki, S. 2013. Kemampuan profesional guru biologi dalam memahami dan merancang model pembelajaran konservasi biodiversitas di SMA. *Cakrawala Pendidikan*, 32(3):408–419.
- Monari, Johnson M. 2020. Higher order questions for higher order thinking skills. *International Journal of Education and Research*, 8(9):1–8.
- Mustikasari, V.R., Munzil, M., & Lestari, L.P. 2018. Pengembangan instrumen penilaian kemampuan berpikir tingkat tinggi materi sistem pendengaran dan sonar SMP. *Jurnal Eksakta Pendidikan (JEP)*, 2(2):116-122.
- Nursalam, Angriani, A.D., Darmawati, Baharuddin, & Aminuddin. 2018. Developing test instruments for measurement of students' high-order thinking skill on mathematics in junior high school in makassar. *Journal of Physics: Conference Series*, 1028(1). <https://doi.org/10.1088/1742-6596/1028/1/012169>.
- Nurulwati. 2019. Teacher training for making higher order of thinking test. *Asian Journal of Science Education*, 1(1):1–9.

- Nuryanti, S., Masykuri, M., & Susilowati, E. 2018. Analisis Iteman dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik Sekolah Menengah Kejuruan. *Jurnal Inovasi Pendidikan IPA*, 4(2):224-2
- OECD. 2019. What Students Know and Can Do: Indonesia. Oecd, 1–10. [https://www.oecd-ilibrary.org/education/pisa-2018-results-volume-iii\\_bd69f805-en%0Ahttps://www.oecd-ilibrary.org/sites/bd69f805-en/index.html?itemId=/content/component/bd69f805-en#fig86](https://www.oecd-ilibrary.org/education/pisa-2018-results-volume-iii_bd69f805-en%0Ahttps://www.oecd-ilibrary.org/sites/bd69f805-en/index.html?itemId=/content/component/bd69f805-en#fig86)
- Prabayanti, N.M.D., Sudiana, I.K., & Wiratini, N.M. 2018. Analisis tes ulangan kenaikan kelas buatan guru mata pelajaran kimia. *Jurnal Pendidikan Kimia Indonesia*, 2(1):25-31.
- Pratama, G.S. & Retnawati, H. 2018. Urgency of higher order thinking skills (hots) content analysis in mathematics textbook. *Journal of Physics, Conf. Series* 1097,1-8
- Purnamasari, A. 2015. Pengembangan alat evaluasi pembelajaran berbasis teknologi informasi dan komunikasi dengan wondershare quiz creator materi sistem penilaian persediaan. *Jurnal Pendidikan Akutansi*, 3(1):1-9.
- Putri, N.E., Yoserizal, Y., Fajri, Y.F., & Fitria, Y. 2019. Analysis of student capacity ability about higher order thinking skills (HOTS) in elementary school. *International Journal of Science and Research (IJSR)*, 8(1):371-373.
- Qirom, M.S., Sridana, N., & Prayitno, S. 2021. Pengembangan soal higher order thinking skills matematika lingkup materi ujian nasional SMP/MTs di SMPN 1 Mataram tahun ajaran 2019/2020. *Jurnal Ilmiah Pendidikan Indonesia*, 3(1):61–70.
- Quieng, M.C., Lim, P.P., & Lucas, M.R.D. 2015. 21-st century-based soft skills: spotlight on non-cognitive skills in a cognitive-laden dentistry program. *European Journal of Contemporary Education*, 11(1):72–81. <https://doi.org/10.13187/ejced.2015.11.72>.
- Rahayu, S., Suryana, Y., & Pranata, O.H. 2020. Pengembangan soal high order thinking skill untuk meningkatkan kemampuan berpikir tingkat tinggi matematika siswa Sekolah Dasar. *Pedadidaktika: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, 7(2):127-137.
- Retnawati, Heri. 2016. Analisis kuantitatif instrumen penelitian. Yogyakarta: Parama Publishing.
- Retnawati, H., Djidu, H., Kartianom, Apino, E., & Anazifa, R.D. 2018. Teachers' knowledge about higher-order thinking skills and their learning strategy. *Problems of Education in the 21st Century*, 76(2):215–230. <https://doi.org/10.33225/pec/18.76.215>.
- Riduwan, M.B.A. 2016. Skala pengukuran variabel-variabel penelitian. Bandung: CV Alfabeta.
- Rintayati, P., Lukitasari, H., & Syawaludin, A. 2020. Development of two-tier multiplechoice test to assess Indonesian elementary students' higher-order thinking skills. *International Journal of Instruction*, 14(1):555–566. <https://doi.org/10.29333/IJI.2021.14133A>

- Sa'adah, S.I., Rasmiwetti, R., & Linda, R. 2019. Pengembangan soal hots dengan wondershare quiz creator sebagai media display pada materi stoikiometri kelas X. JTK (Jurnal Tadris Kimiya), 4(2):177-188.
- Sa'idah, N., Yulistiani, H.D., & Megawati, E. 2019. Analisis instrumen tes higher order thinking matematika smp. Jurnal Pendidikan Matematika, 13(1):41-54.
- Salirawati, D., Permanasari, R. L., Purtadi, S., Ratnaningtyas, A., Nugraheni, E., & Situasi, A. 2017. Pelatihan pengembangan soal HOT (higher order thinking) sebagai peningkatan kompetensi pedagogik guru. Inotek, 21(1):14-25.
- Saraswati, P. M. S., & Agustika, G. N. S. 2020. Kemampuan berpikir tingkat tinggi dalam menyelesaikan soal HOTS mata pelajaran matematika. Jurnal Ilmiah Sekolah Dasar, 4(2), 257-269.
- Setiawan, J., Sudrajat, A., Aman, & Kumalasari, D. 2021. Development of higher order thinking skill assessment instruments in learning Indonesian history. International Journal of Evaluation and Research in Education, 10(2), 545-552. <https://doi.org/10.11591/ijere.v10i2.20796>
- Sudijono, A. 2013. Pengantar evaluasi pendidikan. Jakarta: Rajawali Press.
- Sugiyono. 2015. Metode penelitian kuantitatif kualitatif dan R&D. Bandung: Alfabeta.
- Sukmadinata, & Syaodih, N. 2017. Metode penelitian Pendidikan. Bandung: PT Remaja Rosdakarya
- Sumaryanta. 2018. Penilaian HOTS dalam pembelajaran matematika. Indonesia Digital Journal of Mathematics and Education, 8(8):500-509.
- Tajularipin, S., Vickneswary, M., Diwiyah, M., Raidah, H., & Suzieleez, S.A.R. 2017. Implementation of higher order thinking skills in teaching of science. International Research Journal of Education and Sciences, 1(1):1-3.
- Tan, S.Y. & Halili, S.H. 2015. Effective teaching of higher-order thinking (HOT) in education. The Online Journal of Distance Education and E-Learning, 3(2):41-47.
- Tanujaya, B. 2015. Instrument development of higher order thinking skill in mathematics instructional on senior high school. Jisae: Journal of Indonesian Student Assessment and Evaluation, 1(1):65-72. <https://doi.org/10.21009/jisae.011.05>.
- Wang, S. & Wang, H. 2014. Teaching and learning higher order thinking. International Journal of Arts & Sciences, 7(2):179-187.
- Warju, W., Ariyanto, S.R., Soeryanto, S., & Trisna, R.A. 2020. Analisis kualitas butir soal tipe hots pada kompetensi sistem rem di sekolah menengah kejuruan. Jurnal Pendidikan Teknologi Dan Kejuruan, 17(1):95-104.
- Warner, S. & Kaur, A. 2017. The perceptions of teachers and students on a 21<sup>st</sup>-century mathematics instructional model. International Electronic Journal of Mathematics Education, 12(2):193-215.

- Werdayanti, A. 2008. Pengaruh kompetensi guru dalam proses belajar mengajar di kelas dan fasilitas guru terhadap motivasi belajar siswa. *Jurnal Pendidikan Ekonomi*, 3(1):79–92.
- Widarta, F.O. & Artika, W. 2021. Analisis bentuk stimulus, dimensi kognitif, dan karakteristik hots pada instrumen evaluasi mata pelajaran IPA karya guru. *Jurnal IPA & Pembelajaran IPA*, 5(3):197-208.
- Widana, I.W. 2017. Modul penyusunan higher order thinking skill (HOTS). In Direktorat Pembinaan SMA Direktorat Jenderal Pendidikan Dasar Dan Menengah Departemen Pendidikan dan Kebudayaan.
- Widiawati, L., Joyoatmojo, S., & Sudiyanto. 2018. Higher order thinking skills as effect of problem based learning in the 21st century learning. *International Journal of Multicultural and Multireligious Understanding*, 5(3):96–105. <https://ijmmu.com/index.php/ijmmu/article/view/223>.
- Widyaningsih, S.W., Yusuf, I., Prasetyo, Z.K., & Istiyono, E. 2021. The development of the hots test of physics based on modern test theory: question modeling through e-learning of moodle lms. *International Journal of Instruction*, 14(4):201-212.
- Wiyaka, W., Prastikawati, E.F., & Kusumo Adi, A.P. 2020. Higher-order thinking skills (hots)-based formative assessment: a proposed model for language learning assessment. In *Vision: Journal for Language and Foreign Language Learning*, 9(2): 115–130. <https://doi.org/10.21580/vjv9i25859>.
- Yanto, R., Enawaty, E., & Erlina. 2013. Pengembangan lembar kerja siswa (LKS) dengan pendekatan makroskopis-mikroskopis-simbolik pada materi ikatan kimia. *Jurnal Pendidikan dan Pembelajaran*, 2(3):1–9.
- Yonelia, V.Y.V. & Haryati, S.H.S. 2015. Analisis butir soal ujian semester genap mata pelajaran kimia kelas X IPA SMA PGRI Pekanbaru tahun ajaran 2013/2014 (Doctoral dissertation, Riau University).
- Yuliantaningrum, L. & Sunarti, T. 2020. Pengembangan instrumen soal hots untuk mengukur keterampilan berpikir kritis, berpikir kreatif, dan pemecahan masalah materi gerak lurus pada peserta didik sma. *Inovasi Pendidikan Fisika*, 9(2):62-67.
- Yuliati, S.R. & Lestari, I. 2018. Higher-order thinking skills (HOTS) analysis of students in solving hots question in higher education. *Perspektif Ilmu Pendidikan*, 32(2):181–188. <https://doi.org/10.21009/pip.322.10>