Jurnal Pendidikan Indonesia
Volume 11, Number 4, 2022 pp. 742-752
P-ISSN: 2303-288X E-ISSN : 2541-7207
Open Access: https://doi.org/10.23887/jpiundiksha.v11i4.52595

UTBK-Based Higher Order Thinking Skills (HOTS) Test Instruments on Reaction Rate Topic

Zelen Surya Minata[1*], Nur Indah Agustina[2], Annida Elfiana Citra Ardianty[3], Sri Rahayu[4], Muntholib[5]

[1,2,3,4,5] Chemistry Education Department, State University of Malang, Malang, Indonesia

ARTICLE INFO

Article history:
Received September 07, 2022
Revised September 11, 2022
Accepted December 13, 2022
Available online December 25, 2022

Kata Kunci:
HOTS, taksonomi bloom, instrument tes, laju reaksi

Keywords:
HOTS, bloom's taxonomy, test instrument, reaction rate

ABSTRAK

Kemampuan berpikir tingkat tinggi merupakan salah satu keterampilan yang diperlukan di abad ke-21. Penelitian ini bertujuan untuk mengembangkan instrumen tes pilihan ganda dan essay untuk mengukur Higher Order Thinking Skills (HOTS) peserta didik pada materi laju reaksi. Penelitian ini menggunakan metode Research and Development (R & D) dengan model pengembangan ADDIE (Analysis, Design, Development, Implementation, dan Evaluation). Responden dalam penelitian merupakan 150 siswa kelas XII MIPA yang telah mendapatkan pembelajaran materi laju reaksi. Hasil validitas isi produk menunjukkan instrumen soal 92,01% valid dengan kategori sangat baik. Hasil validitas empirik menunjukkan bahwa sebanyak 20 butir soal pilihan ganda dan 8 butir soal essay valid dan memiliki reliabilitas yang tinggi dengan nilai Cronbach Alpha soal pilihan ganda 0,723 dan soal essay 0,736. Daya beda butir soal dari kedua tipe soal dikategorikan sangat baik dengan 20 butir soal pilihan ganda dikategorikan sedang, sedangkan pada soal essay 1 butir soal dikategorikan mudah dan 7 soal dikategorikan sedang. Selanjutnya instrumen soal yang telah dikembangkan dapat digunakan untuk mengukur HOTS peserta didik.

ABSTRACT

One of the efforts that can be support students in developing 21st century skills is to implement learning based on 21st century skills. In addition, another alternative that can be taken is to create, direct, and provide HOTS-based test instruments as an evaluation of learning. This study aims to develop multiple choice test instruments and essays to measure learners' Higher Order Thinking Skills (HOTS) on reaction rate materials. This research uses the Research and Development (R & D) method with the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) development model. The respondents in the study were 150 students of class XII MIPA one of the high schools who had received learning material on reaction rate. The results of the validity of the product contents show that the question instrument is 92.01% valid with an excellent category. The results of empirical validity show that as many as 20 items of multiple-choice questions and 8 items of essay questions are valid and have high reliability with Cronbach Alpha scores of multiple-choice questions of 0.723 and essay questions of 0.736. The difference between the question items from the two types of questions is categorized very well with 20 items of multiple-choice questions categorized as moderate, while in essay questions, 1 question is categorized as easy and 7 questions are categorized as moderate. Furthermore, the question instruments that have been developed can be used to measure the HOTS of students.

1. INTRODUCTION

21st century skills are an invaluable soft skills asset and these skills are a must-have for everyone, especially in 21st century learners who will face stiff competition in 2045 when demographic bonuses occur. Assessment and Teaching of 21st Century Skills (ATC21S) organizes 21st century skills into four categories, namely ways of thinking, ways of working, tools for working and skills for living in the world. Assessment and Teaching of 21st Century Skills (ATC21S) organizes 21st century skills into four categories, namely ways of thinking, ways of working, tools for working and skills for living in the world. One of the

*Corresponding author
E-mail addresses: sri.rahayu.fmipa@um.ac.id (Zelen Surya Minata)

skills that are indispensable in the 21st century by these learners is the Ways of thinking skills which include critical, creative, and high order thinking skills (Aslamiah et al., 2021; Sulaiman & Ismail, 2020). Furthermore, previous study explained that there are five important components in 21st century skills, including critical thinking, creative thinking, cooperation, motivation, and metacognitive (Lai & Viering, 2012). Critical, creative, and high order thinking are essential to be trained and developed in students in 21st century learning. However, current conditions show that the 21st century skills possessed by students in Indonesia are still low (Amran et al., 2019; Dinni, 2018). This is supported by data from the Programme for International Student Assessment (PISA) in 2018 which shows that the results of the science literacy assessment of students in Indonesia are ranked 74th out of 79 country participants (Chamisah, 2017; Fadillah & Ni'mah, 2019). Therefore, learning innovations are needed that can facilitate learners to develop 21st century skills. One of the efforts that can be support students in developing 21st century skills is to implement learning based on 21st century skills. In addition, another alternative that can be taken is to create, direct, and provide HOTS-based test instruments as an evaluation of learning (Heong et al., 2012; Ichsan et al., 2020). This information supported by research who explained that the formation of HOTS-based questions as a method for training teacher competencies and measuring the level of high-level thinking ability of students (Kusuma et al., 2017). Teachers are required to be able to have good competence in terms of compiling HOTS questions that are relevant to problems in everyday life. But in reality, there are still many teachers who do not integrate life phenomena in the evaluation questions so that they cannot facilitate students to develop their cognitive skills (Klosterman & Sadler, 2010; Sepriyanti et al., 2022).

HOTS is one of the ultimate goals to be achieved by the national education system which is described in the 2013 curriculum. In general, cognitive processes are divided into two levels, namely Lower Order Thinking Skills (LOTS) and Higher Order Thinking Sklls (HOTS). HOTS focuses on developing students' ability to analyze evaluating, drawing conclusions, and synthesizing new information (Brookhart, 2010; Kamin et al., 2015). HOTS is the ability to think that does not just remember (recall), restate (restate), or refer without processing. Based on Bloom's taxonomy that explains that HOTS generally measures the ability of the realm to analyze (analyzing/ C4), evaluate (evaluating/C5), and creating/(C6) (Anderson et al., 2001; Susiaty & Oktaviana, 2019). C6 is the highest cognitive level of the HOTS question instrument because it requires students to be able to think creatively in finding solutions to a problem (Chalkiadaki, 2018; Damaianti et al., 2020). In general, the hope of the Higher Education Entrance Test Institute (LTMPT) to hold a Computer-Based Written Examination (UTBK) can be a prediction related to the ability of prospective students to be able to complete their studies on time in college well. Recently was found that participants in the Computer-Based Writing Examination (UTBK) complained about the difficulty of chemistry subjects, especially in reaction rate materials because the level of difficulty of the questions increased and applied Higher Order Thinking Skills (HOTS) (Harta et al., 2020; Sarah et al., 2021). This difficulty can be caused by students who are not familiar with doing chemistry questions based on HOTS (Mahanal, 2019; Narayanan & Adithan, 2015). Therefore, it is necessary to have an assessment instrument that focuses on assessing Higher Order Thinking Skills (HOTS) on chemistry questions applied to students. The application of the HOTS assessment instrument is an alternative for teachers in training and measuring the level of students' high-level thinking ability (Damaianti et al., 2020; Nugraha et al., 2020).

The implementation of high-level thinking in chemistry subjects that contain macroscopic, microscopic, and symbolic levels can be done by applying chemistry questions that are often done by students at school, one of which is in reaction rate material that is closely related to daily life. In general, students' understanding at the submicroscopic level of reaction rate material is still low when compared to the level of understanding of other chemical representations (Handayanti et al., 2015; Reza et al., 2021). It can be interpreted that the understanding of reaction rate matter at the submicroscopic level involves atoms, ions, and molecules that cannot be observed with the naked eye so that critical thinking skills are needed to understand them. Previou stsudy state reaction rate has four submaterials which include the concept of reaction rate, collision theory, factors affecting the reaction rate, order and reaction rate equation (Hakimah et al., 2021). The characteristic of reaction rate topic is focuses on understanding concepts, calculations, and is applicative. Furthermore, reaction rate is also one of the chemical concepts that are abstract and difficult for students to understand so that it often causes misconceptions (Jusniar et al., 2020; Yan & Subramaniam, 2018). Previous relevant research on the development of HOTS question instruments in chemistry learning has been carried out on thermochemistry topic (Novatania & Kamaludin, 2021). Then there are also researcher who develop HOTS questions in electrolyte and non-electrolyte solution topic (Afriani et al., 2018). Moreover there are also develop HOTS questions on solubility topic (Sarah et al., 2021). However, other topics of chemical material are still not further developed especially in reaction rate topic. Furthermore, in studying the reaction rate material, prerequisite materials are needed, namely chemical equilibrium, chemical reaction alignment, and the concept of mole so that it can be categorized as a complex chemical topic.

Facts on the ground show that HOTS assessment test instruments are still rarely implemented in learning assessments. Preliminary studies conducted at several senior high schools in Malang City showed that the test questions used in learning evaluation tend to measure students' thinking ability in the cognitive level of memorization and remembering so it cannot stimulate students to think at a high level. Through HOTS, learners will be able to give ideas clearly, argue well, be able to solve problems, construct explanations, hypothesize, and understand complex things to be clearer and more comprehensive. HOTS has an important role to play in supporting educational success and making learners more accustomed to creative and innovative thinking (Dinni, 2018; Lee & Choi, 2017). In addition, HOTS is also one of the main skills needed by students to deal with all problems and challenges in the 21st century. This is what underlies researchers to develop HOTS questions based on UTBK questions based on taxonomy bloom on reaction rate topic. This study aims to develop UTBK-based higher order thinking skills in multiple-choice and essays question on reaction rate topic that are valid and reliable. The results of the HOTS assessment that have been developed are expected to be an evaluation in learning to train and improve students' HOTS abilities.

2. METHOD

The approach used in this research is research and development (Research & Development). Research and development methods are research methods used to produce and test the effectiveness of products (Creswell, 2013). The product developed in this study is a UTBK-based Higher Order Thinking Skills chemistry question on the reaction rate material of 20 question items in multiple choice form and 8 essay items. In this study, the stages in design and development used the ADDIE framework (Analysis, Design, Development, Implementation, and Evaluation) (Aldoobie, 2015). The analysis stage is carried out with the aim of identifying and collecting information about the problems and needs of the assessment instrument product to be developed. The design stage is the preparation of the concept and framework of the question instrument which includes indicators of competency achievement, indicators of question items, and assessment rubrics. Furthermore, the development stage is carried out to test the validity of the content and the empirical validity of the question item. However, in this study only to the stage of development because it is only limited to measuring the level of validity and reliability of the questions developed. The trial sample in this study was as many as 150 students of class XII from one of the senior high schools in Malang City with the criteria of having received reaction rate material topic. Research instruments used to obtain data include content validation sheets and student answer sheets as empirical validation data. Content validation is carried out to test the quality of the feasibility of the questions reviewed from the aspects of content, construction, and language. Construct validation is performed to test the validity, reliability, differentiability, and difficulty level of the question item by using the help of the *SPSS 24.0 for Windows program.* Product content validation data is obtained from content assessment conducted by validator experts who are two senior high school teachers. Score percentages are categorized based on Table 1.

Table 1. Product Validity Criteria

Percentage (%)	Category
80 < x < 100	Very good
60 < x < 80	Good
40 < x < 60	Enough
20 < x < 40	Not good
0 < x < 20	Not very good

The question items that are categorized are very good and good, then an empirical validity test is carried out. Empirical validity includes the validity of the question item, reliability, differentiability, and degree of difficulty (Arikunto & Suharsimi, 2009). Data obtained through the results of student answer sheets. The data analysis technique of the percentage of validity of the content of the HOTS question instrument. The validity test shows the level of accuracy of the question instrument to measure what should be measured. The validity test is carried out with the help of the spss 24.0 for Windows program. The test results were compared with the r_{table} value of product-moment (0.159) at a 95% confidence level with n = 150. If the $r_{count} \geq r_{table}$ then the question item can be declared valid, while the $r_{count} \leq r_{table}$ then the question item can be declared invalid. Assessment instruments can be said to be reliable if they are able to be used repeatedly to measure the same object and produce the same data (Sugiyono, 2017). Reliability test analysis is if the cronbach alpha value > 0.60 the assessment instrument has high reliability, while if the cronbach alpha value < 0.60 the assessment instrument has low reliability (Arikunto & Suharsimi, 2009). The interpretation of the reliability data of the question can be seen in Table 2.

Tabel 2. Reliability Criteria

Cronbach Alpha (r)	Category
0,80 < r < 1,00	Very good
0,60 < r < 0,80	Good
0,40 < r < 0,60	Enough
0,20 < r < 0,40	Not good
0,00 < r < 0,20	Not very good

The differentiability shows the level of ability of the question instrument in distinguishing highly capable learners from low-ability learners. The number that shows the magnitude of the differentiation is called the discrimination index (D). The interpretation of the differential power of the question can be presented as in Table 3.

Table 3. Differential Power Criteria

Score	Category
D < 0,00 (negative)	Very low
0,00 < D < 0,20	Low
0,21 < D < 0,40	Moderate
0,41 < D < 0,70	High
0,71 < D < 1,00	Very high

Difficulty level is a number index that shows the easy level and difficulty of a question can be answered correctly by students (Sugiyono, 2017). The interpretation of the difficulty level data of the question item can be presented as in Table 4.

Table 4. Criteria for The Level of Difficulty

Score	Category
0,00 < P < 0,30	Difficult
0,31 < P < 0,70	Moderate
0,71 < P < 1,00	Easy

3. RESULT AND DISCUSSION

Result

The product developed by the researcher is in the form of an instrument about HOTS on reaction rate material consisting of four submaterials, namely the concept of reaction rate, collision theory, factors affecting the reaction rate, order and equation of reaction rate. The question instruments developed are multiple choice and essay. The development of this question instrument was strengthened by the results of needs analysis and surveys in schools that showed information that students rarely do HOTS-based chemistry questions so that the questions used to measure student learning outcomes still use Low Order Thinking Skills (LOTS) questions. In addition, references to HOTS in high school are also still limited and research that develops HOTS-based questions on chemistry topics is still small. A description of the developed HOTS indicators can be seen in Table 5 and Table 6.

Table 5. Description of HOTS Multiple Choice Item Indicator

No	Submaterial	Indicator	Cognitive Level
1	Concept of reaction rate	Analyze phenomena according to the type of reaction	C4 (Analyze)
2	Concept of reaction rate	Analyze research variables to measure reaction rate	C4 (Analyze)
3	Concept of reaction rate	Analyze the reaction rate based on	C4 (Analyze)

No	Submaterial	Indicator	Cognitive Level
		submicroscopic representations	
4	Collision theory	Analyze the exact collision theory based on a phenomenon	C4 (Analyze)
5	Collision theory	Analyze treatments that can improve effective collisions	C4 (Analyze)
6	Collision theory	Summing up the exact statement based on experimental data	C5 (Evaluate)
7	Factors that affect the reaction rate	Analyze the factors affecting the reaction rate	C4 (Analyze)
8	Factors that affect the reaction rate	Analyze the factors affecting the reaction rate with collision theory	C4 (Analyze)
9	Factors that affect the reaction rate	Analyze control variables in the experiment of factors affecting the reaction rate	C4 (Analyze)
10	Factors that affect the reaction rate	Analyze the role of catalysts in a chemical reaction	C4 (Analyze)
11	Factors that affect the reaction rate	Analyze appropriate treatment based on specified conditions and variables	C4 (Analyze)
12	Factors that affect the reaction rate	Design an experimental procedure of factors affecting the reaction rate	C6 (Create)
13	Factors that affect the reaction rate	Summing up the exact statement based on experimental data	C5 (Evaluate)
14	Factors that affect the reaction rate	Identify the tools and materials needed in designing the experiment	C6 (Create)
15	Factors that affect the reaction rate	Analyze the factors affecting the reaction rate	C5 (Evaluate)
16	Factors that affect the reaction rate	Analyze the factors affecting the reaction rate from the experimental data	C4 (Analyze)
17	Factors that affect the reaction rate	Analyze the graph of a reaction with catalysts and without catalysts	C4 (Analyze)
18	Order and equation of reaction rate	Analyze the order and equations of reaction rates	C4 (Analyze)
19	Order and equation of reaction rate	Determine the order of the reaction based on experimental data	C4 (Analyze)
20	Order and equation of reaction rate	Analyze the reaction rate due to temperature changes	C4 (Analyze)
21	Order and equation of reaction rate	Determine the rate of formation of a compound	C4 (Analyze)
22	Order and equation of reaction rate	Analyze the reaction rate equation based on experimental data	C4 (Analyze)
23	Order and equation of reaction rate	Analyze the reaction rate equation based on experimental data	C4 (Analyze)
24	Order and equation of reaction rate	Determine the order of the reaction based on experimental data	C4 (Analyze)

Table 5. Description of Hots Essay Item Indicator

No	Submaterial	Indicator	Cognitive Level
1	Reaction rate concept	Analyze the phenomenon of chemical reactions to determine the meaning of the concept of reaction rate	C4 (Analyze)
2	Reaction rate concept	Analyze the comparison of the rate of chemical reactions in life phenomena	C4 (Analyze)
3	Factors affecting the reaction rate	Design experiments on factors affecting the reaction rate	C6 (Create)
4	Factors affecting the reaction rate	Analyze the factors affecting the reaction rate of a phenomenon	C4 (Analyze)
5	Factors affecting the reaction rate	Analyze the factors affecting the reaction rate by using collision theory	C4 (Analyze)

No	Submaterial	Indicator	Cognitive Level
6	Factors affecting the reaction rate	Analyze the influence of surface area on reaction rate	C4 (Analyze)
7	Factors affecting the reaction rate	Formulate a hypothesis against the rapid slow rate of chemical reactions	C6 (Create)
8	Order and equation of reaction rate	Analyze the order of the reaction and predicting the time of occurrence of the reaction	C4 (Analyze)
9	Order and equation of reaction rate	Analyze reaction equations and graphs of a chemical reaction	C4 (Analyze)
10	Order and equation of reaction rate	Analyze the rate of combustion reactions under certain conditions	C4 (Analyze)
11	Order and equation of reaction rate	Design an experiment to measure the reaction rate	C6 (Create)

Base on Table 4 and Table 5 show a total of 24 multiple-choice questions and 11 essay questions were successfully developed according to the indicators. The question instrument that has been developed is validated by two validators who are high school teachers using a likert scale questionnaire. The results of the validator assessment showed that the average percentage score of 92.05% which was categorized as very good which could then be tested on students who met the predetermined criteria, namely having learned on the reaction rate material.

Empirical validity data obtained from learners' answers. In multiple-choice type questions, the correct answer will be given a score of 1 and the wrong answer will be given a score of 0. In essay type questions, a score of 2 is given on a complete and clear answer accompanied by evidence and strong reasons, a score of 1 if the answer is accurate but the reason is unclear or incomplete, and a score of 0 if the answer does not show a logical conclusion. The empiric validity test is performed with the help of the SPSS 24 for Windows program. r_{count} from the calculation of each question item compared to r_{table}. The r_{table} value based on the number of samples is 0.159. Based on the results of the analysis of multiple choice type questions, 20 points of valid questions were obtained and 4 points of invalid questions were obtained. The results of the validity of the question items can be seen in Table 6.

Table 6. Empirical Validity Results

Validity Index	Type Question	Question Number	Number of Question	Percentage (%)
$r_{count} > r_{table}$ (valid)	Multiple choice	2,3, 4. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, 18, 19, 20, 22, 23, 24	20	84,3%
	Essay	1, 2, 3, 4, 5, 6, 7, 10	8	72,7%
$r_{count} > r_{table}$ (invalid)	Multiple choice	1, 15, 16, 21	4	16,6%
	Essay	8, 9, 11	3	27,3%

Based on the data in Table 6, it can be interpreted that there are 20 multiple-choice questions and 8 essays that are declared valid. This is supported by the value of the r_{count} of the question grain which is higher than the value of the r_{table} ($r_{count} \geq r_{table}$) with a r_{count} range of 0.243-0.685. Meanwhile, 4 multiple choice type questions and 3 essay questions were declared invalid ($r_{count} \leq r_{table}$) so it was not feasible to do a reliability test. Valid question in multiple choice show that there are 16 items C4 question, 2 items C5 question, and 2 items C6 question. Then, valid question in essay show that there are 6 items C4 question and 2 items C6 question. Based on this result, the example of C4, C5, and C6 question can be seen in Table 7.

Table 7. Example Question of HOTS (C4, C5, and C6)

Question Level	Question Indicator	Question
C4 (Analyzing)	Presented some important information regarding the experiment of factors affecting	Carbon dioxide (CO2) is a non-colored gas. At low concentration, this gas is odorless. Whereas, at high concentrations it smells sour and sharp. Carbon dioxide is widely as an inert gas in fire machines, a supercritical solvent in the manufacture of low-caffeine coffe, an ingredient for making carbonates beverages, and as a refrigerant. The main sources of carbon dioxide gas are the processes of

Question Level	Question Indicator	Question
	the reaction rate, students can analyse the appropriate control variables	combustion, respiration of organism, fermentation and the addition of acids to limestone. The reaction equation between limestone and hydrochloric acid is as follow. $CaCO_3(s) + 2HCl(aq) \rightarrow CaCl_2(aq) + H_2O(l) + CO_2(g)$ A group of children will conduct an experiment on the effect of surface area on the reaction rate of $CaCO_3$ and HCl. If the experiment is carried out at a constant pressure, which variables should be controlled? a. $CaCO_3$ mass b. H_2O and CO_2 volume c. Volume CO_2 d. Volume H_2O e. $CaCO_3$ size
C5 (Evaluating)	Presented a description of phenomenon related to collision theory, students are asked to determine the experimental step according to phenomenon	Cantika and her practicum group wanted to prove the collision theory which states that "the wider the the surface area of the touch plane of the reagent particle, then the effective collision that occurs will be more so that the reaction rate will be faster". Here's which is the most appropriate test step to investigate the event? a. Put 100 mL of vinegar acid into beaker A and 150 mL into beaker B; prepare two tablets of ulcer drugs containing $Mg(OH)_2$ with different masses and brands; put one tablet of the drug each into beakers A and B; turn on the stopwatch; record the time it takes for the tablets to finish reacting with vinegar acid on each glass. b. Put 100 mL of vinegar acid into beaker A and 100 mL into beaker B; prepare two tablets of ulcer drugs containing $Mg(OH)_2$ with the same mass; grind one tablet of the drug until it becomes powder; insert one tablet of ulcer medicine in keeping form into beaker A and one tablet in powder form into chemical B simultaneously; turn on the stopwatch; record the time it takes for the tablets to finish reacting with vinegar acid on each glass. c. Put 100 mL of vinegar acid into each of the beakers A and B; prepare two tablets of ulcer medicine containing $Mg(OH)_2$ with different brands; put one tablet of the drug into beaker A and one tablet with another brand into beaker B simultaneously; turn on the stopwatch; record the time it takes for the tablets to finish reacting with vinegar acid on each glass d. Put 100 mL of vinegar acid into beaker A and 150 mL into beaker B; prepare two tablets of ulcer drugs containing $Mg(OH)_2$ with the same mass; put each tablet of the drug in beakers A and B; turn on the stopwatch; stir on beaker A only; record the time it takes for the tablets to finish reacting with vinegar acid on each glass e. Put 100 mL of vinegar acid into chemical class A and 150 mL into beaker B; prepare two tablets of ulcer drugs containing $Mg(OH)_2$ with the same mass; put one tablet of the drug each into beakers A and B; turn on the stopwatch; record the time it takes for the tablets to finish reacting with vinegar acid on each glass
C6 (Planning)	Presented with the tools and materials of an experiment, students can design an experiment from the tools and materials provided	One day Bintan helped mom make a cake and accidentally dropped baking soda into the water containing vinegar and gas bubbles popped into the mixture. Bintan was very curious about what happened, then he tried it again twice. Both experiments showed different reaction times, unfortunately Bintan forgot to record it. Bintan also intends to retry the experiment. Assist Bintan in designing the experiment! a. What materials will you choose for the experiment to be carried out? and determine the amount! b. If the following tools are provided:

Question Level	Question Indicator	Question
		- Plastic cups - Drip pipette - Test tubes - Knife - Measuring cups - Stopwatch - Tablespoons - Label - Teaspoon - Glove - Test tube racks What tools would you choose for the experiment to be carried out? Also determine the amount! c. Design the experimental procedure that you will do!

The question reliability test is only carried out on valid questions so that as many as 20 multiple-choice questions and 8 essay questions are carried out reliability. *The reliability test of HOTS questions on each question type is carried out with the help of the SPSS 24 for Windows program.* The results of the analysis showed that the Cronbach alpha value of multiple-choice type questions was 0.723 and essay-type questions were 0.736 so that they were categorized as having good reliability. The differentiation of the question items is determined based on the discriminant index formula which is then categorized according to the criteria. The power test of the difference in question items is carried out with the help of the Microsoft Excel program. The results of the analysis showed that as many as 20 points of multiple-choice questions and 8 items of essay questions had differentiation power with excellent categories. Therefore, these two types of questions can be used to distinguish highly capable learners from low-ability learners. The difficulty level of the question item is determined based on the difficulty level index formula which is then categorized based on the criteria. The difficulty test of the question items is carried out with the help of the Microsoft Excel program. The results of the analysis showed that as many as 20 points of multiple-choice questions had a medium level of difficulty. While in the essay questions there is 1 question item that is categorized as easy and 7 question items that are categorized as medium.

Discussion

The study implements three stages of the ADDIE development model. The first stage is analysis which is a form of observation to identify and explore information about problems that arise. Based on the results of observations, information was obtained that teachers rarely implement HOTS-based questions as an evaluation in chemistry learning, especially in reaction rate topic. The questions given as learning evaluations are only limited to low cognitive levels (Low Order Thinking Skills) C1-C3. Furthermore, students also experience difficulties in solving the problems given by the teacher. This is due to the unfamiliarity of students to practice doing HOTS-based questions. In addition, the learning resources used are less facilitating learners to think at a high level. Therefore, the development of HOTS-based assessment instruments can be an alternative in solving these problems.

The second stage is a design that is an activity to design and compile indicators of competency achievement and indicators of question items based on basic competencies (KD). The indicators that have been prepared are consulted to expert lecturers which are then developed into HOTS-based assessment questions. The third stage is development which is the core stage of the preparation of the question instrument. The question instrument that has been developed is then validated by two chemistry teachers at one of the senior high schools in Malang City who have more than 10 years of teaching experience. Several revisions were made to the question item after getting advice from validators. The next process is carried out empirical validation aimed at students of class XII who have obtained reaction rate topic. The data obtained from the answers of students are then analyzed to test validity, reliability, difference, and the level of difficulty. A good assessment instrument must fullfill five important aspects, namely validity, reliability, objectivity, practical, and economical (Arikunto, 2015; Laliyo et al., 2019). The HOTS-based assessment instrument that has been developed meets these five aspects. The validity of the question item shows that 20 questions are declared valid that the r_{count} value ≥ r_{table} so that it can be used to measure what is to be measured. This is in line with previous research which shows that questions that can be used as an instrument for assessment and collection of research data, one of which is to meet valid categories (Rintayati et al., 2020; Hairida, 2017). In this study, only valid questions can be tested for reliability. The reliability of the questions is categorized as high with a cronbach alpha value above 0.60 in both types of questions so that it has good consistency to obtain data with the same results. This is in accordance with the results of previous studies which show that reliable assessment instruments can be used to collect research data continuously (Danczak et al., 2020). Based on this information, it can be interpreted that validity and reliability are basic aspects that must be met in developing an assessment instrument that can be used to measure research variables. The differentiability of the question item shows the level of ability of the question instrument in distinguishing highly capable learners from low-ability learners (Sugiyono, 2017; Widyaningsih et al., 2021). Based on the results of data processing, all multiple-choice questions and essays have a very high differentiation power so that they can function properly as a measurement instrument. The difficulty level of the question items interprets that each question item has a level of difficulty or ease for students to answer.

The results of data processing showed that 20 questions (100%) of multiple choice questions had a moderate level of difficulty, while in essay questions, 1 question item (12.5%) was categorized as easy and 7 question items (87.5%) were categorized as medium. HOTS-based assessment instruments in the form of essays developed also have advantages that can stimulate students to think more comprehensively and deeply. Based on the results of the HOTS-based assessment instrument that has been developed, it can be interpreted to meet the categories of content validity and empirical validity in accordance with the existing theoretical framework. Assessment instruments are very important in learning because they can be used as a benchmark for learning outcomes. In learning activities, question instruments are important things needed by teachers to measure the extent of students' ability to understand chemistry learning topic. So far, conditions on the ground show that the majority of learning evaluations conducted by teachers rarely integrate HOTS. Almost all question instruments used to evaluate LOTS-based learning (Ghani et al., 2017; Widiyawati et al., 2019). Therefore, learners cannot hone critical and creative thinking skills as an important aspect of 21st century skills to solve problems (Afandi et al., 2018; Ramdani et al., 2019; Redhana, 2019). The implication of this study is produce the question instruments developed can be used as evaluations in learning and research to measure the HOTS ability of students, especially in reaction rate topic. Students who are used to getting HOTS-based learning will be better prepared to face global challenges. HOTS is an important need as the focus of this 21st century learning. The HOTS question instrument that has been developed by researchers is expected to be applied to measure the HOTS ability of students, especially in the evaluation of chemistry learning of reaction rate topic. Furthermore, the HOTS-based assessment instrument that has been developed is expected to train students to be accustomed to critical and creative thinking in solving problems and preparing themselves for computer-based national exams. This research has limitations, one of which is this study only to the stage of development because it is only limited to measuring the level of validity and reliability of the questions developed. It is hoped that future research will be able to deepen development research related to higher order thinking skills (hots) test instruments.

4. CONCLUSION

Based on the results and research and discussion, it can be concluded that the question instruments developed meet the validity of the content and empirics. Content validation carried out by two expert validators informs that the question instrument has excellent quality with very good categories. The validity test results showed that 20 items of multiple choice questions and 8 items of essay questions were declared valid. Valid question instruments have high reliability with a moderate degree of difficulty and excellent differentiation in both multiple choice and essay questions. However, the development of this question instrument has limitations, namely that it is only implemented until the third stage of the ADDIE development model (development) so that the last two stages (implementation and evaluation) have not been applied.

5. REFERENCES

Afandi, A., Sajidan, S., Akhyar, M., & Suryani, N. (2018). Pre-Service Science Teachers' Perception About High Order Thinking Skills (HOTS) in the 21st Century. *International Journal of Pedagogy and Teacher Education*, *2*(1), 107. https://doi.org/10.20961/ijpte.v2i1.18254.

Afriani, A., Setyarini, M., & Efkar, T. (2018). Pengembangan Instrumen Asesmen Pengetahuan Berbasis HOTS pada Materi Elektrolit dan Non Elektrolit. *Jurnal Pendidikan Dan Pembelajaran Kimia (JPPK)*, *7*(2), 1–12. http://jurnal.fkip.unila.ac.id/index.php/JPK/article/view/16107.

Aldoobie, N. (2015). ADDIE Model. *American International Journal of Contemporary Research*, *5*(6). www.aijcrnet.com/journals/Vol_5_No_6_December_2015/10.pdf.

Amran, A., Perkasa, M., Satriawan, M., Jasin, I., & Irwansyah, M. (2019). Assessing students 21st century attitude and environmental awareness: Promoting education for sustainable development through science education. *Journal of Physics: Conference Series*, *1157*(2). https://doi.org/10.1088/1742-6596/1157/2/022025.

Anderson, L. W., Krathwohl Peter W Airasian, D. R., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J.,

& Wittrock, M. C. (2001). *A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives* (Abridged E). Ney York: Longman.

Arikunto, & Suharsimi. (2009). *Dasar-dasar evaluasi pendidikan*. PT Bumi Aksara.

Aslamiah, A., Abbas, E. W., & Mutiani, M. (2021). 21st-Century Skills and Social Studies Education. *The Innovation of Social Studies Journal*, *2*(2), 82. https://doi.org/10.20527/iis.v2i2.3066.

Brookhart, S. M. (2010). *Assess Thinking Higher-Order Skills in Your Classroom*.

Chalkiadaki, A. (2018). A systematic literature review of 21st century skills and competencies in primary education. *International Journal of Instruction*, *11*(3), 1–16. https://doi.org/10.12973/iji.2018.1131a.

Chamisah. (2017). TIMSS and PISA-How They Help The Improvement of Education Assessment in Indonesia. *Conference Proceedings ARICIS I*, 42–56. https://doi.org/10.22373/aricis.v1i0.935

Creswell, J. W. (2013). *Educational Research Planning, Conducting, and Evaluating Quantittaive and Qualitative Research* (Fourth Edi). Pearson.

Damaianti, V. S., Abidin, Y., & Rahma, R. (2020). Higher order thinking skills-based reading literacy assessment instrument: An Indonesian context. *Indonesian Journal of Applied Linguistics*, *10*(2). https://doi.org/10.17509/ijal.v10i2.28600.

Danczak, S. M., Thompson, C. D., & Overton, T. L. (2020). Development and validation of an instrument to measure undergraduate chemistry students' critical thinking skills. *Chemistry Education Research and Practice*, *21*(1), 62–78. https://doi.org/10.1039/C8RP00130H.

Dinni, H. N. (2018). Pengembangan Mutu Pendidikan Menuju Era Global: Penguatan Mutu Pembelajaran dengan Penerapan HOTS (High Order Thinking Skills). *PRISMA I*, *1*, 170–176. https://journal.unnes.ac.id/sju/index.php/prisma/article/view/19597.

Fadillah, A., & Ni'mah. (2019). Analisis literasi matematika siswa dalam memecahkan soal matematika PISA konten change and relationship. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, *3*(2), 127–131. https://doi.org/10.31764/jtam.v3i2.1035.

Ghani, I. B. A., Ibrahim, N. H., Yahaya, N. A., & Surif, J. (2017). Enhancing students' HOTS in laboratory educational activity by using concept map as an alternative assessment tool. *Chemistry Education Research and Practice*, *18*(4), 849–874. https://doi.org/10.1039/c7rp00120g.

Hairida. (2017). Pengembanagan Instrumen untuk Mengukur Self Efficacy Siswa dalam Pembelajaran Kimia. *EDUSAINS*, *9*(1), 53–59. https://www.neliti.com/publications/177946/pengembangan-instrumen-untuk-mengukur-self-efficacy-siswa-dalam-pembelajaran-kim.

Hakimah, N., Muchson, M., Herunata, H., Permatasari, M. B., & Santoso, A. (2021). Identification student misconceptions on reaction rate using a Google forms three-tier tests. *AIP Conference Proceedings*, *2330*(March). https://doi.org/10.1063/5.0043114.

Handayanti, Y., Setiabudi, A., & Nahadi, N. (2015). Analisis Profil Model Mental Siswa Sma Pada Materi Laju Reaksi. *Jurnal Penelitian Dan Pembelajaran IPA*, *1*(1), 107. https://doi.org/10.30870/jppi.v1i1.329.

Harta, J., Rasuh, N. T., & Seriang, A. (2020). Using HOTS-Based Chemistry National Exam Questions to Map the Analytical Abilities of Senior High School Students. *Journal of Science Learning*, *3*(3), 143–148. https://doi.org/10.17509/jsl.v3i3.22387.

Heong, Y. M., Yunos, J. M., Othman, W., Hassan, R., Kiong, T. T., & Mohamad, M. M. (2012). The Needs Analysis of Learning Higher Order Thinking Skills for Generating Ideas. *Procedia - Social and Behavioral Sciences*, *59*, 197–203. https://doi.org/10.1016/j.sbspro.2012.09.265.

Ichsan, I. Z., Hasanah, R., Ristanto, R. H., Rusdi, R., Cahapay, M. B., Widiyawati, Y., & Rahman, M. M. (2020). Designing an Innovative Assessment of HOTS in the Science Learning for the 21st Century. *Jurnal Penelitian Dan Pembelajaran IPA*, *6*(2), 211. https://doi.org/10.30870/jppi.v6i2.4765.

Jusniar, J., Effendy, E., Budiasih, E., & Sutrisno, S. (2020). Misconceptions in rate of reaction and their impact on misconceptions in chemical equilibrium. *European Journal of Educational Research*, *9*(4), 1405–1423. https://doi.org/10.12973/eu-jer.9.4.1405.

Kamin, Y., Chinedu, C. ., & Olabiyi, O. . (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. https://publisher.uthm.edu.my/ojs/index.php/JTET/article/view/1081.

Klosterman, M. L., & Sadler, T. D. (2010). Multi-level assessment of scientific content knowledge gains associated with socioscientific issues-based instruction. *International Journal of Science Education*, *32*(8), 1017–1043. https://doi.org/10.1080/09500690902894512.

Kusuma, M. D., Rosidin, U., Abdurrahman, A., & Suyatna, A. (2017). The Development of Higher Order Thinking Skill (Hots) Instrument Assessment In Physics Study. *IOSR Journal of Research & Method in Education (IOSRJRME)*, *07*(01), 26–32. https://doi.org/10.9790/7388-0701052632.

Lai, E. R., & Viering, M. (2012). Assessing 21 st Century Skills : Integrating Research Findings National

Council on Measurement in Education. *Pearson*, *April*, 1–67. https://eric.ed.gov/?id=ED577778.
Laliyo, L. A. R., Botutihe, D. N., & Panigoro, C. (2019). The Development of Two-Tier Instrument Based On Distractor to Assess Conceptual Understanding Level and Student Misconceptions in Explaining Redox Reactions. *International Journal of Learning, Teaching and Educational Research*, *18*(9), 216–237. https://doi.org/10.26803/ijlter.18.9.12.
Lee, J., & Choi, H. (2017). What affects learner's higher-order thinking in technology-enhanced learning environments? The effects of learner factors. *Computers and Education*, *115*, 143–152. https://doi.org/10.1016/j.compedu.2017.06.015.
Mahanal, S. (2019). Asesmen Keterampilan Berpikir Tingkat Tinggi. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika*, *3*(2), 51. https://doi.org/10.36312/e-saintika.v3i2.128.
Narayanan, S., & Adithan, M. (2015). Analysis Of Question Papers In Engineering Courses With Respect To Hots (Higher Order Thinking Skills). *American Journal of Engineering Education (AJEE)*, *6*(1), 1–10. https://doi.org/10.19030/ajee.v6i1.9247.
Novatania, D. W., & Kamaludin, A. (2021). Development of High Order Thinking Skills (HOTS) Test Instruments on Thermochemistry Topics. *JTK (Jurnal Tadris Kimiya)*, *6*(2), 174–184. https://doi.org/10.15575/jtk.v6i2.12746.
Nugraha, V. D., Muntholib, M., Joharmawan, R., Parlan, P., Yahmin, Y., & Su'Aidy, M. (2020). The development of the acid-base chemistry test oriented to higher order thinking skills for 11th grade students. *AIP Conference Proceedings*, *2215*(April). https://doi.org/10.1063/5.0000546.
Ramdani, A., Jufri, A. W., Gunawan, G., Hadisaputra, S., & Zulkifli, L. (2019). Pengembangan Alat Evaluasi Pembelajaran Ipa Yang Mendukung Keterampilan Abad 21. *Jurnal Penelitian Pendidikan IPA*, *5*(1). https://doi.org/10.29303/jppipa.v5i1.221.
Redhana, I. W. (2019). Mengembangkan Keterampilan Abad Ke-21 Dalam Pembelajaran Kimia. *Jurnal Inovasi Pendidikan Kimia*, *13*(1). https://journal.unnes.ac.id/nju/index.php/JIPK/article/view/17824.
Reza, M., Puspita, K., & Oktaviani, C. (2021). Quantitative Analysis Towards Higher Order Thinking Skills of Chemistry Multiple Choice Questions for University Admission. *Jurnal IPA & Pembelajaran IPA*, *5*(2), 172–185. https://doi.org/10.24815/jipi.v5i2.20508.
Rintayati, P., Lukitasari, H., & Syawaludin, A. (2020). Development of Two-Tier Multiple Choice 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.
Sarah, F., Khaldun, I., & Gani, A. (2021). The Development Higher Order Thinking Skill (Hots) As Questions In Chemistry Study (Solubility And Solubility Product Constant). *Jurnal Pendidikan Sains (Jps)*, *9*(1), 51. https://doi.org/10.26714/jps.9.1.2021.51-60.
Sepriyanti, N., Nelwati, S., Kustati, M., & Afriadi, J. (2022). *The Effect Of 21st-Century Learning On Higher-Order Thinking Skills (HOTS) And Numerical Literacy Of Science Studnts In Indonesia Based On Gender*. *11*(2), 314–321. https://doi.org/10.15294/jpii.v11i2.36384.
Sugiyono, P. (2017). *Statistik untuk Penelitian*. CV. ALFABETA.
Sulaiman, J., & Ismail, S. N. (2020). Teacher competence and 21st century skills in transformation schools 2025 (TS25). *Universal Journal of Educational Research*, *8*(8), 3536–3544. https://doi.org/10.13189/ujer.2020.080829.
Susiaty, U. D., & Oktaviana, D. (2019). Analisis Kebutuhan Instrumen Tes Berdasarkan Revisi Taksonomi Bloom Untuk Mengukur Higher Order Thinking Skills Siswa. *Proceedings of the 1st ICOLED – IKIP-PGRI Pontianak*, 171–178. https://pbing.org/wp-content/uploads/2020/04/201920completebook.pdf#page=181.
Widiyawati, Y., Nurwahidah, I., & Sari, D. S. (2019). Pengembangan Instrumen Integrated Science Test Tipe Pilihan Ganda Beralasan Untuk Mengukur HOTS Peserta Didik. *Saintifika*, *21*(2), 1–14. https://core.ac.uk/download/pdf/297204372.pdf.
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), 51–68. https://doi.org/10.29333/iji.2021.1444a.
Yan, Y. K., & Subramaniam, R. (2018). Using a multi-tier diagnostic test to explore the nature of students' alternative conceptions on reaction kinetics. *Chemistry Education Research and Practice*, *19*(1), 213–226. https://doi.org/10.1039/C7RP00143F.