ANALYSIS OF TEST INSTRUMENTS BASED ON HOTS CRITICAL THINKING ON PHYSICS IN THE SENIOR HIGH SCHOOL
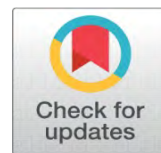
Sabani [1] , Wawan Bunawan [1] , Irham Ramadhani [1] , Maulana Tri Agung [1]

[1] Physics Education, Universitas Negeri Medan, Medan, Indonesia

ABSTRACT

Although learning has been implemented online because of the ongoing Covid-19 Pandemic, it is hoped that it does not eliminate the essence of the implementation of the revised K-13 curriculum, namely the scientific approach to learning process. The implementation of the scientific approach will structure the mind set of students to criticize a problem in their everyday life. This is what is required in the implementation of the revised K-13 curriculum, where critical thinking is a high-level thinking ability that students must familiarize with. HOTS questions are a measurement instrument used to measure higher order thinking skills, namely thinking skills that do not just remember, restate, or refer without processing. The purpose of this study is to: 1) determine the HOTS-based test instrument for high school critical thinking that has been used in the learning process and learning evaluation. 2). Know the level of difficulty and difference in power of the instrument and the effectiveness of the answer keys for high school tests. This research is descriptive research with qualitative and quantitative approaches to documentation study. The findings in the study indicate that the test instruments used at the high school level have not met the criteria for HOTS high-level thinking. As for the results of the calculation of the difficulty level of the test, 8% of the questions were in the easy category, 46% in the medium category and 46% in the difficult category. For the difference in power category, 42% of the questions were in the bad category, 33% were in the enough category, while 4% were in the good category and 21% of the questions were in the very good category. The limitation of this research is the HOTS category of critical thinking test instruments.

Keywords: HOTS Thinking Critical, Senior High School, Test Instrument

1. INTRODUCTION

The strategic plan of the State University of Medan for 2016-2020 was to launch and develop Unimed into a teaching and research institution that excels in producing scientific works. The type of programs that would be developed by Unimed was research, dedication, and science and technology that will be useful for solutions to problems in education, business, and the industrial worlds respectively. Unimed must also produce various learning developments, learning models and media, software, materials, and systems for solutions to educational problems at the center of learning innovation and research. Based on the Unimed Strategic Plan, the Physics Education Study Program tried to support and function through research activities that can be utilized by stakeholders (industry, business, and education).

The Minister of Education and Culture, Mr. Nadiem Makariem made a

breakthrough regarding the implementation of the Student Competency Test called the Minimum Competency Assessment and Character Survey, which replaced the National Examination that has been implemented so far, with impact of the implementation of the National Examination on students' thinking abilities, especially where the ability of students to engage in critical thinking is not significant. This can be seen from the results of the 2018 PISA which showed that Indonesia's scientific ability is still far below, namely with a score of 396. (https://edukasi.kompas.com). Since the Covid-19 pandemic in Indonesia, the Ministry of Education and Culture has issued circular No. 4 of 2020 concerning the Implementation of Education in the emergency period of Corona virus Disease (Covid-19) with the most important main thing, namely online / distance learning to provide meaningful experiences, without being burdened with demands to complete all curriculum achievements for class advancement or graduation. Thus, it required skills for teachers to really master information technology, so that its implementation can run smoothly.

Although the implementation of learning is carried out online, it is hoped that it does not eliminate the essence of the application of the revised K-13 curriculum, namely the existence of a scientific approach in the learning process which is known as 5M (Observing, Asking, Trying, Communicating, and Concluding). By implementing the scientific approach, it will structure the mindset of students on how they criticize a problem in their everyday life. This is what is required in the implementation of the revised K-13 curriculum: critical thinking, which is a high-level thinking ability that students must familiarize with. HOTS questions are a measurement instrument used to measure higher order thinking skills, namely thinking skills that do not just remember, restate, or refer without processing. HOTS questions in the assessment context measure the ability to: 1) transfer one concept to another, 2) process and apply information, 3) look for links from a variety of different information, 4) use information to solve problems, and 5) examine critical ideas and information. However, HOTS-based questions are not necessarily more difficult questions than memory problems. Brookhart (2010). High-level thinking skills really need to be developed and tested on students, especially at the high school level as a reference in the preparation of assessment tools.

Assessment is the process of gathering and processing information to measure the achievement of students' learning outcomes. Assessment of learning outcomes by education is used to: 1) measure and determine the achievement of student competencies, 2) improve the learning process, 3) compile progress reports on daily learning outcomes, midterm, end of semester, end of year, and / or class advancement (Permendikbud No. 23, 2016).

2. MATERIALS AND METHODS

This research is a descriptive type of research with qualitative and quantitative analysis approaches and documentation studies, namely by collecting data using written sources related to the research problem. The subjects in the study were students of class XII IPA at SMA Negeri 1 Per cut Sei Tuan. The data that has been collected in the form of Semester Final Examination items and students' answer sheets were then analysed qualitatively and quantitatively to determine the quality of the UAS odd questions in Physics subjects. Qualitative analysis is the validity of the contents of the questions, namely examining the items from the aspects of material, construction, language, and their relationship with Critical Thinking Indicators according to Fasi one, then analysing the distribution of the items based on the cognitive domain of Revised Bloom's Taxonomy, while analysing the

questions quantitatively by calculating the difficulty level index, index distinguishing power and the effectiveness of tricking questions.

3. RESULTS AND DISCUSSIONS
3.1. QUALITATIVE ANALYSIS RESULTS

Based on the results of the qualitative analysis carried out by examining the results of the content validity, analysis of the validity of the content was carried out by using a review sheet of the questions seen from the aspects of the material, construct, and language. This analysis was carried out by three reviewers in order to avoid the subjectivity of the study in identifying the items and connecting them with the indicators of critical thinking on the questions used during the final semester exams.

Table 1 Results of the validity study of the contents of the final exam questions in class XI		
Validation Aspects	Average expert percentage	Category
Theory	61%	Moderate
Construct	57%	Not good
Language	82%	Good

The quality of the test instrument can be seen from the construct of the test itself, which is based on material content, question construction and language. From the analysis, it was observed that the construction of the test instrument was about 61% of the material content and was in the medium category, the question construction content was about 57% in the poor category, and the language content was about 82% in the good category.

Furthermore, the test instrument is seen based on the cognitive level of Revised Bloom's Taxonomy to identify whether the test instrument used is low-order thinking or high-order thinking. From the data obtained, it shows that the test instruments used are 82% of the test instruments in the low-level thinking domain (C1 = 6%, C2 = 18%, and C3 = 58%). While the test instrument that is in the high order thinking domain is only 18% (C4 = 15%, and C5 = 3%) and of the 6 questions (18%), 5 questions are in the critical thinking category analyzing and 1 question in the critical thinking category evaluating.

Table 2 Results of the Revised Bloom Taxonomy Cognitive Domain Analysis			
Category	Number	Percentage (%)	Level
C1 (understanding)	1,15	6%	LOTS
C2 (knowing)	5, 6, 12, 19, 20, 22	18%	
C3 (application)	3, 4, 10, 11,13, 14, 16,17, 23, 24,25, 26, 27,28, 29, 30,31, 32, 33		
		58%	
C4 (analysis)	2, 7, 8, 9, 21	15%	HOTS (critical thinking skills only in analyzing and evaluating)
C5 (evaluation)	18	3%	
C6 (creation)		0%	

3.2. QUANTITATIVE ANALYSIS RESULTS

Table 3 Results of the Study of Differences, Level of Difficulty, **and effectiveness of answer choices**

Difference power		Level Of Difficulty		Effectiveness Of Answer Choices	
category	Percentage (%)	category	Percentage (%)	category	Percentage (%)
bad	42	easy	8	bad	17
enough	43	moderrately	46	enough	58
good	4	difficult	46	good	21
very good	21	Very difficult	0	very good	4

Analysis of the items carried out on the test instrument at SMA Negeri 1 Per cut Sei Tuan was the level of difficulty and difference power. In SMA Negeri 1 Per cut Sei Tuan, there are variations in the difficulty level of the items with a total of 24 items, namely 11 items that are categorized as difficult by 46%, 11 items that are categorized as moderate by 46% and 2 items that are categorized as easy by 8%. The difference power obtained from the results of the analysis of the test instruments at SMA Negeri 1 Per cut Sei Tuan was included in the category of distinguishing power of multiple-choice questions, very bad category 21% (5 items), 42% (10 items) was in the bad category, the questions were categorized as sufficient 33% (8 items), and 4% (1 item) in good category. From the analysis conducted, it shows that the test instruments used in SMA Negeri 1 Per cut Sei Tuan need to make fundamental changes related to the implementation of the Revised K-13 Curriculum.

4. DISCUSSION
4.1. DISCUSSION OF QUALITATIVE ANALYSIS RESULTS

The quality of the questions is reflected in the results of the implementation of tests carried out in an educational unit. From the data on the results of the material, construction and language study in the final exam questions for class XI at SMA Negeri 1 Per cut Sei Tuan, we observed that the questions from the material aspect are still in the medium category, which means that they are not comprehensively represented by all the material taught in class XI , this shows that in making the questions, there is still no reference to the curriculum which contains material content and basic competencies as a reference. Likewise, the construction aspects are in the bad category. This means that in the preparation of questions, they do not pay attention to the rules in question preparation. Furthermore, for the language aspect, it appears that the questions used in the category are good, this shows that the problems used in the language aspect do not give problems. From the results of the question identification review, whether the questions were included in HOTS or LOTS based on the cognitive level of the revised Bloom taxonomy, it turned out that only 18% of the questions used in the final semester examination were included in HOTS and the rest were still LOTS. This phenomenon shows that the implementation of the 2013 revised curriculum in schools has not really demanded the HOTS learning process. This has become a fundamental problem. However, there are two possibilities that are responsible for this occurrence, firstly because schools have not maximally provided training on the implementation of the 2013 revised curriculum, and secondly, the teachers' inability to develop HOTS-oriented

learning processes and HOTS-based test instruments. Using the critical thinking indicators developed by Facione, namely 1). Interpretation, 2) Analysis, 3) Evaluation, 4) inference, 5) Explanation, and 6) self-regulation, If you review the questions raised by teachers on the aspect of critical thinking skills, you will observe from the results of the analysis that the HOTS questions used in the implementation of the Final Semester Examination are only on the analysis and evaluation indicators, which means that the indicators of critical thinking skills have not been thoroughly given to students of SMA Negeri 1 Per cut Sei Tuan. The questions that are tested must require students to think critically, this is in accordance with the implementation of the 2013 curriculum which is expected to produce productive, creative, and innovative human resources, through competency measurement of attitudes, knowledge, skills, and instruments used to assess thinking skills, high level (HOTS) because it will encourage students to think broadly and deeply about the learning material. Students need to be trained in thinking skills by providing students with thinking skills and this has been carried out in UAS odd questions in Physics class XI IPA class of SMA Negeri in Deli Serdang regency, it's just that the HOTS type of questions are less in number than the LOTS type of questions that are available in the question manuscript (Utari, 2012). Assessment of learning outcomes is expected to help students to improve in higher order thinking skills (HOTS), because higher -order thinking can encourage students to think broadly and deeply about the subject matter. High thinking skills or HOTS are a solution to catching up. In order to catch up, one must survive, where one must be able to have high order thinking skills to solve the problems at hand. Hamzah and Masri's (2014), stated that someone who uses thinking skills will find it easier to complete a job when compared to someone who uses less thinking skills. These thinking skills can range from low-level thinking to high-level thinking. Higher order thinking skills can be achieved if low-level thinking skills have been mastered. Low-level thinking skills are thinking skills from the aspects of remembering to application, while higher-order thinking skills include aspects of analysing, evaluating and creating.

4.2. DISCUSSION OF THE RESULTS OF QUANTITATIVE ANALYSIS

To know whether the question has feasibility, several parameters must be tested, one of which is the item parameter that needs to be analyzed from the results of the item analysis carried out on the final exam questions for class XI SMA Negeri 1 Per cut Sei Tuan. It was observed that the diversity in the level of difficulty, difference power, and effectiveness of answer choices explains that a good question is one that is not too easy and not too difficult (Arikunto, 2015). Questions that are too easy do not stimulate students to maximize their efforts to solve the questions, preferably if the questions are too difficult it will cause students to become discouraged and not eager to try again because they will feel that it is beyond their capabilities. Questions C1 and C2 are categorized as questions with a scale that, questions C3 and C4 are categorized as medium scale questions and questions C5 and C6 are categorized as high-scale questions. Therefore, the proportion of good distribution is 30% easy questions, 40% medium questions and 30 difficult questions. Research on the analysis of the difficulty level of the items is also seen in the research conducted by Amalia and Widayati (2016), with the research title "Analysis of Quality Control Test Questions for Class XII Senior High School in Accounting Economics Subjects in Yogyakarta City, 2012". The various difficulty levels refer to problems with a difficult level of difficulty were 32.5% of the questions with a moderate difficulty level were 62.5%, while those with an easy difficulty level were 5%. The results of this study indicate that the proportion of the

difficulty level of the questions is not in accordance with the proportion that should be. The results of a similar study conducted by Kumusdawara (2016), was used to determine the level of difficulty of the multiple-choice items for the 2014/2015 academic year of the grade V Mathematics, which shows that the UAS items have varying levels of difficulty. The diversity of the difficulty level of the items is shown through the calculation of the difficulty level on each item, namely there are 6 items that can be categorized as easy by 20%, 20 items that are categorized as moderate 66.67% and 4 items that can be categorized as difficult for 13.33%. The results of this study are in accordance with the opinion of Sudjana (2009), which stated that the level of difficulty of the questions is determined by the criteria for questions that fall into the easy, medium, or difficult categories. From some of the research results above, the questions tested did not meet the proportion of the difficulty level of the questions that should have been 30% easy questions, 40% medium questions and 30% difficult questions. In addition to the difficulty level of the feasibility parameter is the difference in the power of the question. Based on the results of the analysis of the distinguishing power of the questions, the odd UAS questions in Physics class XI IPA SMA Negeri 1 Percut Sei Tuan which are included in the distinguishing power category of multiple-choice questions is very bad 21% items), 42% (10 items) were bad, 33% (8 items) had enough distinguishing power (8 items), and 4% (1 item). Similar research by Kumusdawara (2016) was used to determine the distinguishing power of multiple-choice items for the 2014/2015 academic year, which showed that the UAS items have 19 items which can be categorized very well with a percentage of 63.33%, 5 items that were categorized as quite good with a percentage of 16.67%, 4 items that were categorized as moderate with a percentage of 13.33%. In addition, there are 2 items that can be categorized as bad with a percentage of 6.67%. The results of this study are in accordance with the opinion of Kunandar (2014), which explains that the requirements for multiple choice tests are to have sufficient discriminating power to differentiate between students who have mastered the material (competence) and students who have not mastered the material. Sudjana (2009) explained that the analysis of distinguishing power aims to determine the ability of questions to distinguish students who are classified as having high achievement and students who have low achievement and is used to find out students who have or have not mastered the competence of lessons. Sarea and Hadi (2015) explained that there are several reasons for an item to have low distinguishing power, among others: questions that contain bias, questions that are too difficult and distractors that do not make sense. The existence of a distractor that does not make sense will make it easier for students to decide whether the distractor is right or wrong, so that the likelihood of students guessing correctly is very high and causes the item to be too easy, on the other hand a distractor that is too close to the truth value with the answer key causes the item to be too difficult. Arikunto (2013) also stated that good items are items that have a discrimination index of 0.4 to 0.7. Discriminatory power index that is negative (very poor criterion) should not be used, because it shows the student's best ability. Basuki and Hariyanto (2015) added that items that have distinctive power with bad criteria should also not be used, items that have sufficient criteria can be accepted but must be corrected, and items with good criteria can be used.

5. CONCLUSIONS AND RECOMMENDATIONS

Based on the findings in this study, it can be concluded that:

1) Out of the questions used by SMA Negeri 1 Per cut Sei Tuan in Physics subject for class XI, 18 percent met HOTS questions with critical thinking skills that are dominated by analyzing.

2) From the results of the calculation of the difficulty level of the test, 8% of the questions were in the easy category, 46% in the medium category and 46% in the difficult category. For the difference power, 42% of the questions were in the bad category, 33% of the questions were in the enough category, 4% of the questions were in the good category and 21% of the questions were in the very good category.

ACKNOWLEDGEMENTS

REFERENCES

Aiken, L. R. (1985) Three coefficients for analyzing the reliability and validity of ratings. Educational and psychological measurement, 45(1), 131-142. Retrieved from https://doi.org/10.1177/0013164485451012

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

Borg. W.R. & Gall, M.D. (1989) Educational Research: An Introduction. New York : Longman.

Brookhart, S. M. (2010) How to Assess Higher Order Thinking Skills in Your Classroom. Alexandria : ASCD.

Conklin, W. (2012) Higher-order thinking skills to develop 21st century learners. Huntington Beach : Shell Educational Publishing, Inc.

Ennis, R. H. (1993) Critical Thinking Assessment. Theory into Practice. Vol 32, No 3, page 179-186. Retrieved from https://doi.org/10.1080/00405849309543594