The Reflection of Cognitive Level in Test Item Constructed by English Teacher in Summative Assessment

Nisrina
Pendidikan Bahasa Inggris, Fakultas Keguruan dan Ilmu Kependidikan, IAIN Lhokseumawe
Email : nisrina.nina1909@iainlhokseumawe.ac.id

Abstract

This study aimed to analyze the six levels of cognitive domain based on Revised Blooms Taxonomy and the use of HOTS and LOTS in the Final School Exam questions for twelfth-grade students, at teachers' made test in summative assessment. This research was descriptive qualitative research that used content analysis as a method to analyze the specific character of each question. The researcher used Final School Exam questions from twelfth-grade students in the academic year 2021- 2022 at SMKN 8 Lhokseumawe as a data source and used a checklist tablein collecting and analyzing data to find out the research results. The resultsof data analysis showed that the most dominant cognitive domain is the level of understanding (C2), which is 60% or 24 of 40 questions. The second highest is the level of Remembering (C1), which is 30% or 12 of 40 questions. Next is the Analyzing level (C4), which is 10% or 4 of 40 questions. The cognitive domain is classified into two levels, namely Lower Order Thinking Skill (LOTS) and Higher Order Thinking Skill (HOTS). Of the three categories applied to the test, it can be shown that 90% or 36 of the 40 questions are included in the LOTS category, while the other 10% or 4 of 40 questions are included in the HOTS questions. LOTS questions are much more dominant than HOTS questions.

Keywords: *Cognitive Level, Test Item, Summative.*

INTRODUCTION

Assessment acts an integral role in the teaching-learning process. It is applied to identify how students are learning and running lessons taught at school (Ismail et al., 2022). Many kinds of assessment are used by teachers in the classroom as they must always be concerned with whether or not the students have reached the learning objectives or not. One of them is summative that are required to provide information about the learners' achievements, and are usually carried out at the end of a semester/course (Achadiyah et al., 2023). Specifically, they aim to determine whether students should be allowed to be in the next stage to deal with new, more advanced, objectives of learning. Without any tests, a teacher have difficulty in providing proof of the quality of his students. To construct a good test that is fair is not easy to do. A teacher needs to work hard. Brown (2004) sets out several stages of test construction which consist of determining test objectives, drawing up test specifications, devising test tasks, scoring, grading, and giving feedback. To produce a better one, a teacher must follow the available syllabus and deal with many references related to the rules on how the test items should be made. As a consequence, the teacher is not allowed to make a test based on his own desires without referring to the syllabus. A teacher has to consider the quality of a test while creating an effective test. In this case, a test must possess the instruction that insists students to reflect their behavior in achieving learning objectives. This means that a test has to be well constructed by considering the instructional leaning objectives. The test that is used and designed to determine what students know at specific point in time or summative assessments such as, mid test or final test. Assessment is a scientific method of the evaluation to acquire feedbacks related to the information of teaching and learning, make teachers and students see the achievements and shortcomings clearly, and improve teaching and learning efficiently (Zhao, 2017).

Laili et al., (2020) stated that the Indonesian government through the Minister of National Education increase the amount of item tests allocated HOTs items. It is suggested to allocate 25% of item test should be in HOTs item test. It means that the teachers and also educational unit institutions are required to help students to be more critical in thinking. One effort that needs to beimplemented is to design a test or assessment with a high level of thinking quality.There are two levels of thinking, they are Lower Order

Thinking Skill (LOTS) and Higher Order Thinking Skill (HOTS). Making assessment or test that based on LOTS and HOTS refers to the cognitive domain in Taxonomy Bloom to see how far the tests are constructed based on what government ask through the test, that is 25% of item test should be in the form of HOTs (Shalihah et al., 2022). It is crucial to make such kind study, since the teachers, at some points, do not realize that the tests they constructed did not accommodate the demanded achievement set.

One of the assessment that teachers familiar with is giving summative task. The study conducted by (Gordani, 2010) revealed that the pivotal role of assessment in the learning process is to give reflection whether the students have reached the learning objective or not. Moreover, the test designed must fit the requirement of learning objectives, one of those objectives are cognitive dimension. From cognitive perspective, the test must allocated LOTs and HOTs questions, HOTs question should be allocated 25% of total question (Laili et al., 2020). There should be six categories namely remembering (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), and creating (C6). The first three categories are included into Lower Order Thinking Skill (LOTS) level, while the other three categories are ca lled Higher Order Thinking Skill (HOTS) (Assaly & Smadi, 2015).

The writer chooses to analyze summative test because constructing good summative tests in the form of teachers-made test are quite hard to do. It makes interested to do study on that field. Teachers made tests usually are used for summative test since it is not launched until the learners take the test. Then the tendency to evaluate how the teachers construct the test by fulfilling demand from government set are crucial to do. It means that test objectives should be based directly on learning objectives and test content is derived from specific course content. Although the emphasis of summative test is on measuring learning outcomes, it should not be implied that testing is to be done only at the end of instruction. According to Heaton (1975) information concerning the performance of the students as a whole and of individual student is very important to teaching purposes, especially as many test result can show not only the types of error most frequently made but also the actual reasons for the error being made. The information got from test result can be as reflection both teachers and students (Soares, 2015), further, item testanalysis information can tell us about the level of questions elected.

Item analysis is a process which examines student responses to test items; to assess the quality of those items and of the test as a whole will show whether or not the test appropriate for the students. It is not only valuable in improving item which will be used again in later tests, but it can also be used to improve the test if the level of test cannot reflect the actual behavior (Semiun & Luruk, 2020). In addition, item analysis is valuable for increasing instructors skill in instruction, and identifying specific areas of course which need greater emphasis clarity regarding the level of instruction per se.

Considering that English summative test items should fulfill the requirement of good test and by looking at the benefits of analysis of test items above the researcher regards that it is very important to conduct research to report the quality of English summative test items related to the level of the question allocated based on Bloom taxonomy. It is expected can help the teacher to identify / finalize the most appropriate test items and make sure that new test is designed to meet the objective set by the government and truly reflect learning and fulfill the requirements of good test. Beside that, the researcher also expected that the information of the study can help the teacher to know the most cognitive level allocated.

METHOD

The researcher conducted descriptive qualitatively through content analysis since it can determine the characteristics of materials such as textbook, speeches, test and various other type. Current study took English summative test as the corpus data and it was analyzed based on Revised Bloom's Taxonomy to see the level of questions constructed. Regarding the instrument, content analysis checklists were built to seek out the data. The researcher constructed guidance that was

filled out with the six categories of cognitive levels of Bloom's taxonomy to analyze summative test items. The guidance was to calculate the level of cognitive dimension consisting of ten columns, those are, number, six level of cognitive domain, questioning stems, number of tests, and total. The vivid description of the instrument can be seen at the Tabel 1 below:

Table 1. content analysis checklist of test allocated

No	Number of Question	Questioning Stems	Cognitive Domain					
			C1	C2	C3	C4	C5	C6
1.								
dst								

This study uses a coding scheme, based on Bloom's taxonomy, to codify, classify, and analyze the summative test made by teachers. The purpose of developing the coding scheme was to detect the number of item tests that were accordingly included in HOTs, and LOTs. Classifying questioning stems into one of the six levels of Bloom's taxonomy was a crucial task since the broader lines among the six levels. Moreover, it was understandable that one item of the task might cover more than one cognitive level. For example, questions asked about students' comprehension and at the same time, asked them to apply their understanding to the new learning peripherals. In this case, it would be difficult to choose between one of the two levels.

To solve the problem, the researcher conducted a session to obtain a trustworthy of the data by administering reliability. For this sake, the researcher applies two kinds of reliability analysis, namely, intra and inter-rater reliability. To ensure intra-rater, the data was coded twice by the researcher in two weeks-span. In addition, inter-rater reliability was conducted by a colleague who was mastering in evaluating the item tests related to the level of cognitive, because the line to differentiate among cognitive dimensions could be very close, it was urgent to grab from different perspectives. The data of this study was the English summative test in the final school exam for the twelfth-grade students at SMKN 8 Lhokseumawe in the academic year 2022/2023. This final school exam consisted 40 multiple-choice questions.

RESULTS AND DISCUSSION

The objective of the current study was to identify the summative test items constructed based on the principle that the level of cognitive-based on Taxonomy Bloom, HOTs, MOTs, LOTs, would be constructed related to the government demand that HOTs tasks should be inserted 20% of total tasks in learning textbook.

There were 40 items of question in the final exam and the distribution of the item tests and the level of cognitive questions can be seen in Table 2 below:

Table. 2 The Percentage of Summative Test Based on Taxonomy Categories

Cognitive Domain	Cognitive Processes in Cognitive Domain	Data	Number of Cognitive Processes	Total	Percentage
Remembering C1	Recognizing	Q14, Q28	2	12	**30 %**
	Recalling	Q3, Q9, Q12, Q16, Q17, Q29, Q30, Q31, Q32, Q33	10		
Understanding C2	Interpreting	Q4, Q11, Q15, Q18, Q21, Q22, Q24, Q 27	8	24	**60 %**
	Exemplifying	Q5, Q23	2		
	Classifying	-			
	Summarizing	Q2, Q6, Q8, Q20,	4		
	Inferring	Q1, Q7, Q10, Q13, Q19, Q25, Q26, Q36, Q37, Q39	10		
	Comparing	-	-		
	Explaining	-	-		
Applying C3	Executing	-	-	-	**0%**
	Implementing	-	-		
Analyzing C4	Differentiating	-	-	4	**10 %**
	Organizing	Q34, Q35	2		
	Attributing	Q38, Q40	2		
Evaluating C5	Checking	-	-	-	**0%**
	Critiquing	-	-		
Creating C6	Generating	-	-	-	**0%**
	Planning	-			
	Producing	-			

Table 2 above showed that only 3 levels of taxonomy categories were applied in reading comprehension questions in the twelfth grade English final examination. The 3 categories were the level of remembering, understanding, and analyzing. *Remembering level* was found in 12 of 40 questions with a percentageof 30%. Of the 12 items, 2 of them is included in the thinking process of *Recognizing*, and 10 others were included in the thinking processes of *Recalling*. *Understanding level* was found in 24 of the 40 questions with a percentage of 60%. Of the 24 items, 8 items were included in the thinking processes of *Interpreting*, 2 item into the *Exemplifying*, 4 items into *Summarizing*, and 10 others into the thinking processes of *Inferring*. *Analyzing level* was found 4 of 40 questions with a percentage of 10%.Of the 4 items, 2 of them belongs to the thinking processes of *Organizing*, and the other 2 belongs to the thinking processes of *Attributing*. It can be concluded that the most

allocated taxonomy categories level at English final exam test was understanding (C2) level, there were 24 questions included in understanding level followed by Remembering (C1) and Analyzing (C4). The most frequent tasks observed at understanding level were Interpreting, Inferring, Exemplifying, and Summarizing. In addition, 4 item tests have been devoted to higher-order thinking skills. Overall, it was found that all of the tasks were convergent in the first three levels of Bloom's Taxonomy which are included in lower and middle levels of cognitive skills, and little contribution was allocated in the higher level of cognitive skills. From the task distribution, the results manifest that Bloom's levels of learning are not equally deposited in the English learning textbook.
Further the data was categorized based on the level of LOTs and HOTs level of the question, the data revealed as presented in Table 3 below:

Table 3 LOTS and HOTS Constructed in Summative Test

LOTS		HOTS	
C1	30%	C4	10%
C2	60%	C5	-
C3	-	C6	-
Total	90%	Total	10%

As we know that remembering, understanding, and applying were included in lower order thinking skill (LOTS) level, while analyzing, evaluating, and creating level is included in higher order thinking skill (HOTS). The total number of questions that relate to Lower Order Thinking Skill reach to 90% with the frequency 36 of 40 questions. The total number of questions that relate to Higher Order Thinking Skill were 10% with the frequency 4 of 40 questions. Applying,evaluating, and creating were included in higher order thinking skill (HOTS) level. It can be concluded that the most of item tests constructed belong to the LOTS level with a percentage of 90%, while the HOTS level was 10%. In otherwords, LOTS questions were more dominant than HOTS questions.

From the data displayed above it can be seen that the most distributed item tests of 40 were in Lower Order Thinking Skill (LOTs) that is 4 or 10%. The results of the current study are in line with Bernasela (2014), Rusma (2016) and Semiun and Luruk, (2020) who conducted a study on item test constructed in the summative test and the results showed that the tasks were not distributed equally among the levels of cognitive level based on Bloom's Taxonomy and it showed that there were hyperbole attention toward LOTs level of cognitive domain. Another study investigated by Pratiwi et al., (2019) revealed the unbalance of distributed item in summative test to a small number of HOTs item test had been devoted. Both studies found an exaggerated spread on the two lowest cognitive levels and a lack of encouragement of higher cognitive levels, MOTs, and significant insufficiency of higher-level thinking skills, HOTs.

This study are in line with Musliha et al., 2022 who conducted a study on teachers-made test at summative assessment and the results showed that the tasks were not distributed equally among the levels of cognitive level based on Bloom's Taxonomy. In addition, they revealed the unbalance of distributed item test and they found an exaggerated spread on the LOTs level of item test. In the other words there is a lack of encouragement of higher cognitive level.

The result of the current study also similar with the other studies conducted by Pratiwi et al., (2019). They found that there were not sufficient enough numbers of HOTs questions distributed in the guidance book. It was detected that there were some items in HOTs level. It is suggested teachers should analyze the item tests before applying it in the classroom to counterbalance the weakness if the teachers made the item test for summative test (Pradanti et al., 2018) or they must consider choosing the proper summative test set hence students can boost their HOTs level.

The consideration that item tests included in summative tests must accommodate the demand for life skills in the 21st century compels tasks abided by government regulation, that is it should reach 25% of total number of the test. The policy refers to the need for life skills in the 21st century that equip students with critical thinking, communication, and ICT literacy (Bakken & Andersson-Bakken, 2021). The implication of the policy was item tests should be constructed in the form of sufficient HOTs level.

CONCLUSION

Based on the first chapter there are two questions formulated. The first question was formulated to determine the taxonomic categories used in the summative assessment of English at final test at SMKN 8 Lhokseumawe. The second question was formulated to determine the level of questions used in the summative test of English. After analyzing item test at the final school for twelfth graders, the answers to the research questions were found.

Cognitive domain in Revised Bloom's Taxonomy is categorized into six levels,namely remembering (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), and creating (C6). Based on the findings, it was found that there were only 3 Taxonomy categories on item test that were applied based on Revised Bloom's Taxonomy theory, namely Remembering, Understanding, and Analyzing. Cognitive domain at the Understanding level (C2) is the most widely used, which is 24 out of 40 or 60%, the second most common level is Remembering (C1), which was 24 out of 40 or 30%, next is level Analyzing (C4), which is 4 out of 40 or 10%. While the other3 levels are not included in this reading comprehension question, they are Applying (C3), Evaluating (C5), and Creating (C6).

Based on Revised Bloom's Taxonomy theory, the cognitive domain is divided into 2 levels ranging from low, namely LOTS to high, namely HOTS. In this study, it was found that the questions categorized into the LOTS level contained 36 out of 40 questions or 90%, while the questions categorized into the HOTS level contained 4 out of 40 questions or 10%. In other words, questions with LOTS level are more dominant than questions with HOTS level. It can be concluded that low-level questions dominate summative test constructed by the teacher at the Final School Examination on Summative Assessment at SMKN 8 Lhokseumawe, it was for 90% of the total.

SUGGESTION

Based on the analysis of reading comprehension questions in the English final school exam, the researcher would like to give some suggestions to help increase the level of thinking in designing reading comprehension questions as follows:

Through the results of this study, the researcher suggest that English teachers are expected to design test questions related to educational goals that are included in the HOTS level category so that the quality of these questions can improve students' thinking ability. Moreover, in making final school exam questions, which are used as graduate requirements, the quality of the questions should also be high according to the objectives set by the government. In preparing questions, especially final school exam questions needed by students at the final stage Schools should pay more attention to the content of each question item. Where the English teachers should design final school exam test with more varied levels, not only at the level of remembering, understanding, and analyzing but also at three other levels, such as applying, evaluating, and creating. The teacher should also help students in facing the national exam by familiarizing students with high-level questions, so that students have no difficulty in answering the national exam questions later on.

REFERENCES

Achadiyah, R., Azmi, M. U., Kyai, U., & Faqih, A. (2023). *HOTS Manifestation on English Summative Assessment*. *8*(2).

Assaly, I. R., & Smadi, O. M. (2015). Using bloom's taxonomy to evaluate the cognitive levels of master class textbook's questions. *English Language Teaching*, *8*(5), 100–110. https://doi.org/10.5539/elt.v8n5p100

Bakken, J., & Andersson-Bakken, E. (2021). The textbook task as a genre. *Journal of Curriculum Studies*, *53*(6), 729–748. https://doi.org/10.1080/00220272.2021.1929499

Bernasela. (2014). *An Analysis on English Summative Test Items English Education Study Program Teacher Training And Education Faculty*.

Gordani, Y. (2010). An analysis of english textbooks used at iranian guidance schools in terms of Bloom's taxonomy. *Journal of Asia TEFL*, *7*(2), 249–278.

Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani,

E. (2022). Formative vs. summative assessment: impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia*, *12*(1). https://doi.org/10.1186/s40468-022-00191-4

Laili, M., Aini, N., & Christanti, A. (2020). High order thinking skills (HOTS) dalam penilaian bahasa Inggris siswa SMA. *Lintang Songo: Jurnal Pendidikan*, *3*(1), 18–25.

Musliha, S., Sudana, D., & Wirza, Y. (2022). The Analysis of Higher Order Thinking Skills (HOTs) in the Test Questions Constructed by English Teachers. *Proceedings of the Fifth International Conference on Language, Literature, Culture, and Education (ICOLLITE 2021)*, *595*(Icollite), 610–617. https://doi.org/10.2991/assehr.k.211119.095

Pradanti, S. I., Martono, M., & Sarosa, T. (2018). An Item Analysis of English Summative Test for The First Semester of The Third Grade Junior High School Students in Surakarta. *English Education*, *6*(3), 312. https://doi.org/10.20961/eed.v6i3.35891

Pratiwi, N. P. W., Dewi, N. L. P. E. S., & Paramartha, A. A. G. Y. (2019). The Reflection of HOTS in EFL Teachers' Summative Assessment. *Journal of Education Research and Evaluation*, *3*(3), 127. https://doi.org/10.23887/jere.v3i3.21853

Rusma, S. (2016). ANALYSIS OF SUMMATIVE TESTS FOR ENGLISH By Rusma Setiyana *. *English Education Journal*, 433–447.

Semiun, T. T., & Luruk, F. D. (2020). The quality of an English summative test of a public junior high school, Kupang-NTT. *English Language Teaching Educational Journal*, *3*(2), 133. https://doi.org/10.12928/eltej.v3i2.2311

Shalihah, M., Fikri, D., & Mustofa, M. (2022). Analyzing higher order thinking skills (HOTS) questions of reading essay tasks in senior high school English textbook. *English Education Journal*, *13*(1), 106–121. https://doi.org/10.24815/eej.v13i1.23956

Soares, S. (2015). No Title空間像再生型立体映像の 研究動向. *Nhk技研*, *151*(3), 10–17.

Zhao, Y. (2017). A Study on Problems of Minority Students in English Learning. *3rd International Conference on Economics, Social Science, Arts, Education and Management Engineering (ESSAEME)*, *119*(Essaeme), 1247–1250. https://doi.org/10.2991/essaeme-17.2017.257