

A revised bloom's taxonomy evaluation of formal written language test items

Yulis Setyowati ^{a1} Language and Literature Education, Postgraduate School, Universitas Negeri Surabaya, Universitas Wijaya Putra, Surabaya, Indonesia, <https://orcid.org/0000-0001-8368-0627>

Susanto Susanto ^b, Language and Literature Education, Postgraduate School, Universitas Negeri Surabaya, Indonesia, <https://orcid.org/0000-0002-0150-8952>

Ahmad Munir ^c, Language and Literature Education, Postgraduate School, Universitas Negeri Surabaya, Indonesia. <https://orcid.org/0000-0003-0606-3590>

Suggested Citation:

Setyowati, Y., Susanto, S. & Munir, A. (2022). A revised bloom's taxonomy evaluation of formal written language test items. *World Journal on Educational Technology: Current Issues*. 14(5), 1317-1331. <https://doi.org/10.18844/wjet.v14i5.7296>

Received from May 10, 2022; revised from July 11, 2022; accepted from September 15, 2022.

Selection and peer review under responsibility of Prof. Dr. Servet Bayram, Medipol University, Turkey

©2022 by the authors. Licensee Birlesik Dünya Yenilik Arastırma ve Yayıncılık Merkezi, North Nicosia, Cyprus.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract

This paper aims to portray the appropriateness of test items in language tests according to Bloom's Taxonomy. Thirty written language tests created by EFL lecturers were analyzed. Document analysis was applied, the data were categorized and examined. In the test for remembering, 'crucial questions was applied, finding specific examples or data, general concepts or ideas, and abstracting themes in comprehension test. Completing particular projects or solve issues in the applying test, whereas SWOT analysis conducted in analyzing test, and strategic plan should be demonstrated in evaluation test, and last, in creating test, new things or idea should be created, generalizing and make conclusion. The findings demonstrated test item using remembering mental level stood at 66%, understanding 16%, applying 2%. While analyzing level gets 9%, evaluating 2%, and creating group 5%. This addresses disparity between LOTs and HOTs usage. Hence, Bloom taxonomy was not distributed well in the language tests.

Keywords: test items, formal written language tests, Revised Bloom taxonomy

¹ ADDRESS OF CORRESPONDENCE: Yulis Setyowati, Language and Literature Education, Postgraduate School, Universitas Negeri Surabaya, Universitas Wijaya Putra, Surabaya, Indonesia
Email address: yulis.17070956007@mhs.unesa.ac.id

1. Introduction

Over the last few years, research on language tests has remarkably increased their presence. The language test is a report of the student's achievement often used by teachers to evaluate the students' competence. It is vital to support the students' learning to improve the quality of education and the progress of the students' pedagogical achievement (Cheng & Fox, 2017). Unfortunately, the fact that some EFL lecturers may find problems conducting tests creates a gap between the ideal language test and the practice. They may fail to create professional, comprehensive, and effective language tests. The design goals of language tests and how lecturers practice in everyday life (Morris, 2016), (Bierman et al. 2017), (Farhadi & Tavasoli, 2018), (Atiullah et al., 2018).

Moreover, EFL lecturers providing appropriate test items in composing the proper language test is one of the problematic issues. Besides, deciding on the article as the test is the most challenging part of designing a language test. (Bennett & Service, 2014).

A Written Language test (Final & Midterm) is an achievement test that seeks the students' cognitive ability decided through the test score. The test is referred to be essential in assessing the learners' performance. A good test accommodates various learners' capabilities (Soleimani, 2016)(Morales-Obod et al., 2020). One of the tools to adapt comprehensive skills in the test is the taxonomy from Bloom. The taxonomy is a broadly used device and is considered a suitable parameter to exemplify educational goals in academic records, from curriculum, and assessment, to evaluation of student's achievements (Omar et al., 2012);(Lee et al., 2017); (Köksal & Ulum, 2018); (Samira Baghaei et al., 2020); (Belarbi, Fatine & Bensafa, 2020);(S Baghaei et al., 2021). Thus, the revised Bloom's taxonomy was confirmed as an appropriate guideline to construct questions (Omara et al. I. In (Samira Baghaei et al., 2020), (Taras, 2010); (Chang & Chung, 2009); (Ebadi, 2016). The functional test tool measures learning and thinking skills based on the six criteria of Bloom's taxonomy. The six criteria consist of remembering, understanding, applying, analyzing, evaluating and creating.

To promote students' critical thinking in the 4.0 era, EFL lecturers are expected to have the ability to design both the learning process and its assessment. The lecturers are guided to develop the test by analyzing, evaluating, and creating refer to Bloom's taxonomy. In line with the research conducted by Köksal & Ulum (Köksal & Ulum, 2018), He said that among the questions collected from various universities in Turkey, Bloom's classification lacked a higher cognitive level.

Built on Bloom's classification, the purpose of the learning system in the learning process is that educators engage students. The Bloom's taxonomy in cognitive domain, fit to arrange pupils' mental stage (Omar et al., 2012), is the basic of categorizing statements based on the goals of teaching-learning activities (Krathwohl, 2002). Assessment practices according to the Bloom taxonomy are vital guidelines for the instructors. The taxonomy helps the teachers to increase contents' awareness and careful for the matter they teach and assess as well as the learning process. Furthermore, it can be used as a guide for the teacher to develop the students' cognition related to the test content. Due to the explanation above, this current study tries to answer, 'How do formal written language test items cover the Bloom taxonomy in revised version?'

Conceptual Framework

The Bloom Taxonomy

Fundamental and theoretical framework model of cognitive domains in Bloom's taxonomy (Anderson & Krathwohl, 2001), was used in this study to arrange the items test. The Revised Bloom's taxonomy classified upper steps and lower ones. (see figure 1). Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating are the six cognitive levels of increasing complexity. The first mental aspect, remembering, suggests to regaining concerned knowledge from long-term memory. Understanding is the following dimension, which "involves defining the implication of instructional communications, including oral, written, and graphic." The technique or learned matter usage in each condition is called applying. Analyzing entails breaking down further complex knowledge

into its fundamental segments, identifying the relationships among them, and recognizing the organizational principles at work. Evaluating refers to decisions making up according to specific principles and standards. As the final and highest dimension, creating entails assembling elements to form a novel and incorporating whole or developing an original product. (Krathwohl, 2002)

One of the uniqueness of the taxonomy is that the cognitive level may represent other levels. Namely, in analyzing phase people could also perform in remembering, understanding, and applying steps (Airasian & Miranda, 2014). The taxonomy was a part of the standard feature in the assessment methods in every item question of the test. Bloom's Taxonomy in 1956 (Bloom, 1956) and its revised version have been regarded as "a practical guideline in constructing good exam questions associated with generating enquiries (Lister & Leaney, 2003), appraising pupils' cognitive mastery (Oliver, Dobele, Greber, & Roberts, 2004), textbook appraisal (Parsaei, Alemokhtar, & Rahimi, 2017; Razmjoo & Kazempourfard, 2012; Sahragard & Zahed Alavi, 2016) and defining examination questions on the cognitive domain (Chang & Chung, 2009).

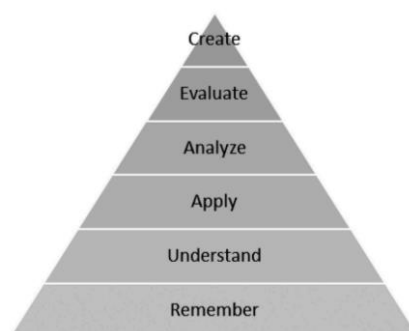


Figure 1: The revised Bloom's Taxonomy in cognitive domain

A. Test for remembering

When instruction aims to stimulate recollection of the given substance in considerably the similar form in which it was instructed, the related process type is remembering. Remembering involves retrieving related information as of long-term memory. Retaining ability is crucial for profound learning and problem-solving. Remembering such knowledge is critical for deep studying and problem solving when that knowledge is utilized in more complex assignments. When school teachers emphasize practical learning assessment, remembering expertise must be incorporated into the immense task of forming new intelligence or unravelling new challenges (Lorin and Krathwohl, 2001), (Airasian & Miranda, 2014). A remembering test demands scholars to remember, recognize or identify evidence to prove knowledge procurement.

B. Test for understanding

The largest class of transfer-based scholastic ideas emphasized in colleges is understanding. Pupils are expected to realize when they can assemble value from instructional communications, including oral, written, and graphic communications and material presented during teaching or in books. Pupils know when they build connections between new and prior knowledge. The incoming expertise is integrated with the remaining schemes and cognitive structures. (Lorin and Krathwohl, 2001), (Cullinane, 2015), (Eber & Parker, 1999). A comprehension test demands learners to give explanations, clarify, and assume theories or concepts of certain information to show knowledge comprehension.

C. Test for Applying

The application involves using techniques to work or solve problems. The implementation phase involves two behavioral approaches: performing the project (e.g., traditional to students) and practicing – if the workout is a concern (e.g., unused for students). Execution (recognized as execution) occurs when students practice the procedures for a familiar assignment. Implementation (known as use) occurs because students use certain techniques for unusual tasks. Unlike performing, which

focuses roughly on cognitive methods related to applications exclusively, management incorporates cognitive approaches to understanding and application (Lorin & Krathwohl, 2001).

A test to apply is a test that claims that scholars recruit or take advantage of information in a new general state of the knowledge usage.

D. Test for Analyzing

Analyzing encompasses splitting the material to each core parts, specifying how elements are associated and details in a general framework. The cognitive processes are allotted to differentiating, organizing, and attributing(analyzing)(Eber & Parker, 1999), (Cullinane, 2015).

Test analysis implies an exam that compels scholars to break, differentiate, and organize evidence into segments to explore Comprehension and association.

E. Test for Evaluating

The most used constraints are performance, stability, and efficiency. This category delivers the cognitive technique of evaluating (that deals with outcomes about inward strength) and criticizing (focused on beyond criteria). It could be identified that two primary skills; checking (mostly stated as monitoring, managing, remembering, or analyzing), would occur if pupils identify irregularities or miscalculations within a method and define a course or formation comprises efficiency of a process as being practised; critiquing (known as evaluating) occurs when pupils notice inconsistencies in creation or method and some external ideals amid development or design and outward standards, of a procedure for a subject, existing. Generally, criticizing is associated with open-minded opinion. Critiquing with pupil-individual proficiency could judge how decent an invention or design is. (Lorin and Krathwohl, 2001), (Cullinane, 2015), (Eber & Parker, 1999). A test for evaluation indicates a test which demands pupils to justify, verify, and judge a choice or way of action.

F. Test for Creating

Comprises arranging facets to formulate a sound logical, in another term, reshuffling pieces into a new arrangement or structure is creating. In this arrangement, making draws scholars' conception or gives a distinctive invention. For instance, a work of art and writing is mostly a cognitive process related to the product, but not always. In creating, pupils are offered a relentless portrayal of an empirical and deliver a design to fulfils the narrative. (Lorin and Krathwohl, 2001), (Wilson, 2016), (Cullinane, 2015). Overall, a test for create entails pupils to generate or deliver innovative thoughts, goods, or means of showing matters.

Related Research

Some studies have been conducted evaluating the Revised Bloom Taxonomy in formal language tests in a university context. For instance, Zareian et al. (2015) probed that questions in two ESP course books were mainly associated with recollection, comprehending, and utilizing as the three lower-level classifications; analyzing, evaluating, and creating where the three higher-level classifications had a minor frequency. A study by Koksai and Ulum (2018) applied the Bloom Taxonomy(revision) to inspect exam questions of General English courses and found that the higher-level cognitive skills in the exams only appeared in a few occurrences.

Moreover, Belari, Fatine & Bensafa, and Abdelkader (2020) acquired Low higher order skills questions that dominated Algeria's English foreign language (EFL) baccaulaureate exam. In a more recent study, Ginting et al. (2021) researched final semester exam questions and reported that the exams were in the tests almost entirely reliant on lower-order reasoning abilities. Only one of the five courses uses a question-writing system according to HOT skills, and none of the topics addresses mid-order thinking competencies. As a result, the cognitive degree on the instruments test has not increased the students' thinking competences and position on low level. Both studies suggested designers need to incorporate further higher-order questions

Purpose of the study

The study aims to figure out how the midterm and final test items cover the revised Bloom's taxonomy created by EFL lecturers. It investigates the performance of tests of remembering, an examination of understanding, an exam of applying, an exam of analyzing, a test of evaluating and an assessment of creating, which are written in test items.

2. Research Methodology

The current research is based on qualitative document text analysis since the study tries to analyze the item test in Mid-terms and Final terms created by English Foreign Language Lecturers. The study brings replicable and sound findings from textbooks to the perspectives they use (Morgan, 2022). The characteristics of a text are extensively categorized and evaluated in qualitative document text analysis. In reality, the primary objective of a qualitative document text analysis study is to investigate the implications and topics of messages to understand how they have been structured and portrayed. A qualitative document text analysis is the most general qualitative approach analysis, whereas significant leitmotifs were found, classified and investigated (Kutsyuruba, 2017).

Participant

This study involves four EFL lecturers who graduated not from English Language Teacher Training University. The only reason the subjects were selected was because of their educational background. The first and second subjects are EFL lecturers, former from universities in the United States and Europe, the third is an EFL lecturer who is a graduate from a prominent university outside Java, and the last one is a graduate from leading universities in Java. Gender and educational experience were neglected in selecting the subject of this study.

Data Collection Tools

The data was collected through observation and personal documents to obtain data. The researcher collected the files of the written language tests (Midterm and Final Terms) designed by the subjects. Fifteen Midterms and fifteen Final Terms in the form of written examinations consisted of one hundred sixty-six test items.

Data collection Process

The study uses a qualitative approach, whereas observation is a primary method for obtaining data. By observation, the researcher could receive the data immediately and interact in a natural situation (Ary et al., 2014). In the practice of qualitative research, in this study, two broad categories of data collection are used: observations and personal documents.

Data Analysis

The data analysis procedure is based on a design of document analysis that explains the occurrence of the steps covered by Bloom's Taxonomy at both high and low order cognitive levels. Put differently, the mental stages of Bloom's Taxonomy were utilized when classifying the test items used at the university level. This study investigates item trials presented in the written document test. The method covers data reduction, data display, and verification. The current research conducts a cross-check of interpretation to avoid performance subjectivity. Thus, the analysis is interpreted by at least two people (researcher and co-author). Furthermore, the study seeks the assistance of a university professor in justifying its findings.

Data Validation

Validation stage comprises the approval of the data analysis outcomes. The data shown in the table was double-checked in conjunction with data collected on or after the development to verify the discoveries. The validator was a professor who was proficient in language tests.

3. Results

The source of the data is in the form of written tests, particularly the students' midterm and final term examinations in the English department in four universities in Indonesia. The tests vary in terms of the name of the courses, the number of items, interval, and type of tests (subjective test and objective test form). The EFL lecturers who created the difficulties were chosen from their educational background but not their gender. The other reasons were neglected, such as the disciplines they educated, their expertise, interests, and others. The data of this research covers one hundred sixty-six test items.

The test entries collected and next were classified and then listed in a table following the six domain of Bloom's Taxonomy, namely: (1) the test for remembering, (2) the test for understanding, (3) the test for applying, (4) the test for analyzing (5) the test for evaluating and the test for creating. Every category entails examples of action verbs for each degree and the keywords that represent the cognitive domain. The table presents the data classification and makes it more manageable; the table is designed to collect categorization:

Table 1 Example of data categorization

Cognitive Level	Test items	Data Identity
Test item for remembering	Where is the nearest place to taste an excellent Italian restaurant and wine after getting tickets from Ticket's Booth in Leicester Square?	S2.2
Test item for understanding	Paraphrase the following text!	S3.1
Test item for applying	Write an essay employing any literary criticism we learned in class toward Franz Kafka's "Metamorphosis.	S1.4
Test item for analyzing	Analyze the usage of "components" by using Culture Approach and Discourse Pattern	S4.8
Test item for evaluating	Individually, write a 500 to 600-verses summary and comments on the matters of translation	S1.6
Test item for creating	Write an essay and Follow the guidelines below:	S1.4

This research has revealed some findings related to the appropriateness of test items in the language tests according to Bloom's Taxonomy. EFL lecturers are confirmed to apply Bloom's taxonomy in the formal language written tests. The diagram will explain in detail :

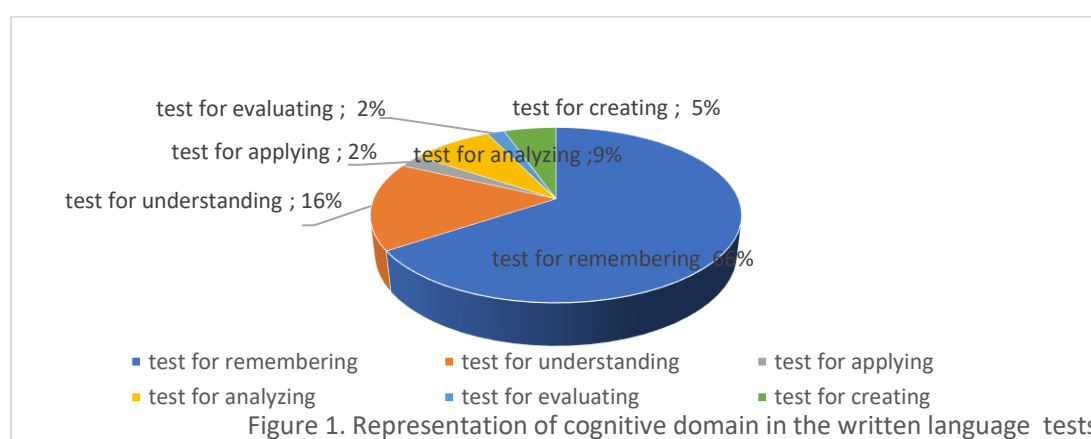


Figure 1 demonstrates the test for remembering (66%) is the most frequently employed in the tests, followed by a test for understanding (16%). This fact highlights that lower-order cognitive levels dominate in the test items but not for the test for applying. Even though the test for applying belongs to the lower order, this mental level only contributes 2%. Furthermore, the frequencies of the higher-order categories, i.e. analyzing, evaluating, and creating, is in a few %ages: the test of interpreting is 9%, while the test for assessing and test for creating are in the same number, 2%. It underlines that EFL lecturers highlighted lower-order thinking level in the language tests and set aside the use of the high-order thinking degree. The Revised Bloom Taxonomy did not well employ appropriately.

Test for remembering

As observed, the test for remembering primarily appeared in the trials; it is 66%. At this level, the EFL lecturers are requested to retain expressive memorizing of an issue or the learned materials. This test item was typically asked about intellectual materials saved in the brain's memory. Thus, the lecturers must recall the learned material they had previously memorized or remembered. Test takers had no difficulty answering this kind of test at this cognitive level. A question of placing is expressed in a test item asking the test takers to locate knowledge, recognize, find, or remove facts, concepts, or definitions.

In addition, it is found that test for remembering is acted in both objective and subjective type of written tests. In subjective tests, the item trials refer to identifying subcategories found in 16 test items, while inaccurate tests, 10 test items were placed. Twenty-six test items were categorized as a test for remembering.

In test items of remembering only performed in the meet and match exercise, multiple choices, true and false, were also performed in more completed and integrated tasks requesting new construction of knowledge or solving new problems. This kind of task focuses on meaningful learning and requires a lot of embedding into the border context

Test for understanding

Test for understanding which is categorized as a lower thinking order level is a test which demands the test taker to recall and modified learned materials. The cognitive activities in the category of understanding cover interpreting, exemplifying, classifying, summarising, inferring, comparing, and explaining. This level of mental performed in twenty-seven test items which means only 16% are 5 test items for interpreting, 3 test items for each classifying, summarising, inferring, and 13 test items for explaining. For test items, exemplifying can only be found in a single test item, and the test item for comparing is none. Even understanding level is lower order thinking order level. It didn't appear as many as the remembering level: 66%.

One example when the lecturer asked the test takers is to write a summary. He wrote, '*individually write a 500 to 600 words summary, ' indicating that the lecturer requested the test takers to document their understanding by concluding on text. In particular, test takers should give their summary by rearranging or re-building any expression using their own words or sentences. The test would be measured by how they could mix or re-build their own words or sentences to make one new sentence or concept depicting their Comprehension of the text.*

The lecturer directed the students to comprehend something behind the question asked. In this case, there would be an effort for the lecturer to head the students to grasp what was meant by the test item. Test of understanding asks the test takers to construct connotations from instructional communication, including oral, written, and graphic messages, by some actions such as interpreting by shifting from one representation to another.

Test for applying

In contrast, the test for applying the middle order thinking level only appeared in four test items: 2%; 3 test items for implementation and one execution constructed by two EFL lecturers. Test for

applying refers to tests that require students to execute and implement. In this test, takers must learn to apply for managing and implementing materials. This test item was typically distinguished by applying a procedure to a familiar or unfamiliar assignment. Thus, test takers are just required to use some courses to do chores or solve problems. This is an example:

Using the three literary criticisms, we have discussed in the class, analyze the short story (data S1.3)
Use AntConc as a Software Application. (data S4.8)

The test items above showed that the test for applying uses subcategories that requests students to choose and use a procedure to finish an unfamiliar assignment. Since students should decide on the correct method, they must understand the problems and provided alternative approaches.

From data S1.3, we can infer that the students have been given three literary criticisms discussed with the lecturers before. Then, they should apply those to the short story presented in the lectures. By analyzing the short story, the students are expected to match the suitable literary criticism to some particular lines in the short story. It is also possible to determine the suitability of the three literary criticisms of the short story. The students face a strange problem. This makes them unable to solve the problem directly. But, in this case, they have been given three literary criticisms. So, they have some alternative options to solve the problem. Since there is no absolute procedure to solve the problem, the students may modify procedures. Implementing is often associated with technic or methods using skills and algorithms. The first thing student should do reading each line of the short story. After that, they should compare the strings with the most suitable literary criticism. They will do it several times until they are sure they find one or more literary criticism which can be implemented to parts of the short story.

Test for analyzing

As explained earlier, the middle order thinking skills appeared in a few occurrences as in %age only 9%. There are three sub-categories, namely differentiating, organizing, and attributing. It was noticed that fourteen test items for analyzing cover 2 test items for determining, nine test items for organizing, and 3 test items for attributing. The test for diagnosing requires students to have an analysis that entails dissecting a subject into its parts and figuring out how the elements link to a larger structure. Test Items of Analysing aims to break down, determine the relationship between factors (to one another), and select the general construction. In formatting the test items, the EFL lecturers applied the same test format; response or selection task. Imposing a structure on material which covers a table, hierarchical diagram, matrix, or outline is part of organizing.

In constructed response, the lecturer probably asks the students to outline a passage in writing. In an assortment task, the lecturer gives the students instructions to select the alternative graphic hierarchy that best fits how a piece is organized. The student was given some written or spoken information as part of the attribution subcategory and then asked to create or choose a portrayal of the author's or speaker's opinions, goals, and other characteristics as attributing assessment. In constructed retort, the lecturer asks the students to analyze the reasons why the author or speaker wrote an essay. It involves the purpose, point of view, intentions, and the like. The lecturer probably asks the student to choose the correct answer to a multiple-choice test item in the selection task. Differentiating assesses the students by giving them some material and then asking them to find the crucial or pertinent portions and recognize which parts are most important or relevant. In constructed response, the lecturer probably asks to write a short answer. The lecturer probably asks the student to choose the correct answer to a multiple-choice test item in the selection task. The following is one example; *'Write a short "essay" about Othello. Explore what, how, and why Shakespeare's Othello is a tragedy,* these test items the lecturer has already given a specification. The assessment format of this test item will be a constructed response which makes the student produce a written outline of a passage as organizing lets the student impose a structure on material such as a table, hierarchical diagram, matrix,

or scheme. What, how, and why in the topic represent the outlines. Students are expected to write a constructed response based on the borders given.

Test for evaluating

The data analysis presented that the test item for evaluating the higher order thinking abilities was recognized in 2% in tripple test items formulated by EFL teacher within number the first and fourth sequentially. Test for assessing refers to test items which require students to judge. Judgments depend on the benchmarks and standards which students or others have determined. The examples involve steadiness, value, efficiency, and competence, while the standards involve quantitative (numerable) and qualitative (numerable). The cognitive process of evaluating entails some forms of conviction. Those are verifying and critiquing. Inspection discusses domestic consistency, while critiquing discusses conclusions influenced by external criteria. Here is the example: *write five hundred to six hundred-word comments on translation issues* (Data S1.6); this test item is categorized as critiquing subcategory. In making the judgment, two things are analyzed: determining the principles and standards. The criteria are consistency, quality, effectiveness, and efficiency. In this data, the four criteria are the issues of interpretation. Furthermore, there are two indicators for the measure: qualitative and quantitative. The data qualitative in this study is the number of words that must be written, namely five hundred – six hundred words of comments, while the qualitative is the issues of translation.

In the test for evaluation, students should be competent to critique assumptions or thoughts they or others have uttered. Thus, students should write critical thinking about the issues (translation) from their perceptions. Then, pupils must ascertain the criteria or basis for their criticism, whether positive, negative or both. In determining the cause for complaint, they should note the positive and negative characteristics of the process they value. Then, the result is that they can make judgments based on the traits they have made. The teacher sets a limitation (standard) for them to follow when composing their critical thinking.

Test for creating

Test for creating, as higher order thinking skill has three categories; generating, planning, and producing. Test items of creating refer to test items requiring students to compose elements to become coherent and functional. Nine test items were categorized as a test for creating and referring to producing subcategories formulated by three EFL lecturers.

Writing a 1,000 to 1,250-word paper detailing issues of literary translation is one of the test items for creating which students should be involved in executing the plan to accomplish problems that qualify certain specifications. The assignment is about composing an essay. The lecturer has also determined specifications. It should be a 1,000 to 1,250-word paper specifying issues of literary interpretation. In this assignment, the lecturer provided students with academic translation material. Lecturers also must introduce the types of works that other translators usually translate. The main objective of this assignment is that students can compose an essay detailing literary translation issues. In this assignment, the lecturer gives a limit; namely, the article cannot be less than 1,000 words and cannot be more than 1,250 words. Regarding the preparation process, students must re-read theories about literary translation, identify issues of literary translation from various sources such as books and the internet, and then paraphrase the points they have obtained from digging up the information in words.

4. Discussion

The current study was designed to illustrate the accuracy of test items in language tests created of the revised Bloom's classification. The analysis of 150 test items reveals the massive occurrence of test items that emerged from cognitive level domains (remembering, understanding, applying, analyzing, evaluating, and creating). It standpoints out that test for recognizing level is more prevalent in all sections of test items in the language tests (66.0%), whereas the test for understanding is 16%,

the applying 2%, the analyzing 9%, the evaluating 2%, and test for creating domain occupies 5%. In particular, as depicted in figure 1, the test items in language tests formulated by the EFL lecturers were primarily drawn on the LOTS rather than HOTS domains.

Moreover, the EFL language tests are appropriate to the "remembering" or "understanding" level of Bloom's taxonomy. It is linear to the finding of (Muhayimana et al., 2022) 's, (Ginting et al., 2021) 's, (Abahussain et al., 2020) 's, (Al-Khayyat, 2020) 's, (Belarbi, Fatine & Bensafa, 2020) 's, (Atiullah et al., 2019) 's, and (Köksal & Ulum, 2018) 's, (Rezaee & Golshan, 2016) 's, (Assaly & Smadi, 2015) 's, research. These studies also discovered the examination questions lack of higher-order thinking level. The phenomenon, a tendency to place a premium and pay special attention to remembering, understanding, and applying levels, seems to be a universal occurrence over the past years.

In general, the tendency of putting remembering, understanding, and applying levels into a prior place is driven by the belief that these three domains are the most significant. (Tikhonova & Kudinova, (2015) & Tsaparlis, (2020) in (Muhayimana et al., 2022). Thus, (Krathwohl, 2002) stated that knowledge and Comprehension are repeatedly deliberated as the groundwork for all the other educational purposes (Chmielewka & Gilanyi, 2018 in (Muhayimana et al., 2022). Knowledge and grasp, which is remembering and understanding level in the Revised Bloom taxonomy, are required to understand more complex thinking exercises comprehensively. It seems that remembering, learning, and applying levels (LOT levels) are the most common type of intellectual capacity to be concentrated in higher education. The more learners know, the better they can understand, interpret, and judge the world around them. (S Baghaei et al., 2021) mentioned that the higher levels of analyzing, evaluating, and creating require the condition that the learners should remember what they have learned in the past. In this case, learners and teachers could increase the use of more LOTs than the higher levels tasks. It is believed that learners must construct a solid knowledge base and strengthen lower-order thinking processes early to move to more complicated tasks later on. (Tikhonova & Kudinova (2015), in (Muhayimana et al., 2022).

In line with this study, three published scientific papers reveal a high incidence of lower-order cognitive processes in EFL texts and evaluation methods at university levels since students' proficiency is not high. The sample research papers at illustrated in the following table:

no	Country, author, year	Materials analyzed	level	Lot	Hot
1	Iran, Zareian et al., 2015	ESP course book	universities	90%	10%
2	Turkey, Koksal and Ulum., 2018	Exam questions of general English courses	universities	90%	10%
3	Algeria, Belari, Fatine & Bensafa, Abdelkader, 2020	A university entrance exam	universities	100%	0%
4	Indonesia, Ginting et al., 2021	Final Semester Examination	universities	mostly	A few

As the table displays, there is a significant difference between LOTS and HOTS; the more substantial number goes to low order thinking skills, and HOTS only get a few of the occurrence. Such an imbalance questions on the six cognitive domains may contribute to negative impacts on instructional quality and student learning due to the washback effect.

The present study's findings regarding the portrayal of the appropriateness of test items in the language tests taken from The Revised Bloom's Taxonomy turns out to be unexpected. The 'unexpected finding' confirms the misalignment between the lecturer's objectives and designing assessment and how lecturers practice in the education field (Bearman et al., 2017), (Zareian et al., 2015). Test items frequently covered LOTS, remembering, understanding, and applying levels. In the 4.0 era, English foreign language lecturers are opposed to layout both a learning process and an

assessment that will encourage critical thinking among students. Within the revised Bloom's classification, lecturers are instructed to plan and develop an assessment that helps to promote their skills by analyzing, evaluating, and creating important thinkers. However, this current study and the previous research mentioned above did not result in suitable conditions.

5. Conclusion

The current research findings demonstrated that the written language tests were more engaged with lower-order thinking levels, which showed 79%. In other words, the tests did not lead to an increase in students' critical thinking skills. The story of thinking (cognitive level) on the questions in the language tests is not based on HOTS as a whole. Referring to these findings, further socialization and training related to curriculum updating is necessary, especially concerning the design of learning evaluation quality. Lecturers must design learning evaluations that foster students' critical thinking and creativity. The learning evaluation also needs to be adjusted to the advancement of language learning and the realities of life. Besides, the stakeholders should be engaged to develop lecturers' ability to design language learning and its evaluation.

Finally, it should be emphasized that Bloom's taxonomy is to assist educators in gaining clarity while planning and developing the teaching process. There is no connotation that instructors must follow a predetermined, controlled progression from one level of objectives to the next. The taxonomy also enables a teacher to be aware of the cognitive levels of each pedagogical activity, as well as needs and outcome evaluation. The instructional design supports a cohesive and progressively organized learning process that allows mastery of the topic while keeping Bloom's taxonomy in mind. It encourages students to develop their personality evaluation skills.

Recommendation.

Regarding the finding which indicated the unequal distribution of use of Bloom taxonomy, it is recommended that EFL lecturers design language tests that maximize the HOTS level and minimize the LOTs level to promote critical thinking. Hence the stakeholders should create a workshop to develop the skills of EFL in designing the language tests. Decision makers need to put more effort into integrating Bloom's Taxonomy into the EFL teaching and learning process.

References

- Abahussain, M., Iqbal, M., & Khan, I. (2020). Standardization of EFL Undergraduate Skill Exam Papers: A Case Study at Majmaah University. *Arab World English Journal (AWEJ) Volume*, 4(December), 508–518. <https://doi.org/10.24093/awej/vol11no4.24> Pp.363-381 Standardization
- Airasian, P. W., & Miranda, H. (2014). The Role of Assessment in the Revised Taxonomy. *Theory into Practice*, 41(4), 249–254.
- Al-Khayyat, A. S. J. (2020). Exploring the Educational Digital Learning Systems Utilized by EFL Postgraduate Students During the Covid19 Pandemic Quarantine. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(6), 11481–11488.
- Ary, D., Jacobs, L. C., Sorensen, C. K., & Walker, D. A. (2014). *Introduction to Research in Education 9th Edition* (M. Kerr (ed.); Ninth Edit). Cengage Learning.
- 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>
- Atiullah, K., Fitriati, S. W., & Rukmini, D. (2019). Using revised Bloom's taxonomy to evaluate Higher Order Thinking Skills (Hots) in reading comprehension questions of English textbook for year X of High School. *English Education Journal*, 9(4), 428–436.
- Baghaei, S, Bagheri, M. S., & Yamini, M. (2021). Learning Objectives of IELTS Listening and Reading Tests:

- Focusing on Revised Bloom's Taxonomy. *Research in English Language Pedagogy*, 9(1), 182–199. <https://doi.org/10.30486/relp.2021.1916940.1244>
- Baghaei, Samira, Bagheri, M. S., & Yamini, M. (2020). Analysis of IELTS and TOEFL reading and listening tests in terms of Revised Bloom's Taxonomy. *Cogent Education*, 7(1). <https://doi.org/10.1080/2331186X.2020.1720939>
- Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D., & Joughin, G. (2017). How university teachers design assessments: a cross-disciplinary study. *Higher Education*, 74, 49–64. <https://doi.org/10.1007/s10734-016-0027-7>
- Belarbi, Fatine & Bensafa, A. (2020). An Evaluation of the Algerian EFL Baccaalaureate Exam under the Cognitive Domains of Bloom's Taxonomy. *Arab World English Journal (AWEJ)*, 11(4), 508–518.
- Bennett, R. E., & Service, E. T. (2014). *Formative assessment : A critical review* This article was downloaded by : [Karlstads Universitetsbibliotek] *Assessment in Education : Principles , Policy & Practice Formative assessment : a critical review*. April, 3–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Bloom, B. (1956). *Taxonomy of Educational Goals*. David McKay Company.
- Chang, W., & Chung, M. (2009). *Automatic Applying Bloom' s Taxonomy to Classify and Analysis the Cognition level of English Question Items*. 727–734. <https://doi.org/10.1109/JCPC.2009.5420087>
- Cheng, L., & Fox, J. (2017). *Assessment in the Language Classroom* (the first). Palgrave.
- Cullinane, A. (2015). Bloom' s Taxonomy and its Use in Classroom Assessment. *NCE-MSTL*, 1(13).
- Ebadi, S. (2016). Exploring Bloom's Revised Taxonomy of Educational Objectives in TPSOL Textbooks. *Journal of Teaching Persian to Speakers of Other Languages*, 5(1), 65–93.
- Eber, P. A., & Parker, T. S. (1999). *Assessing Student Learning : Applying Bloom' s Taxonomy* Bloom' s Taxonomy. *National Organization of Human Service, 1956*, 45–54.
- Farhady, H., & Tavassoli, K. (2018). Developing a Language Assessment Knowledge Test for EFL Teachers: A Data-Driven Approach. *Iranian Journal of Language Teaching Research*, 6(3), 79–94. <https://doi.org/10.30466/IJLTR.2018.120602>
- Ginting, P., Hasnah, Y., Hasibuan, S. H., & Batubara, I. H. (2021). Evaluating Cognitive Level of Final Semester Examination Questions Based on Bloom's Revised Taxonomy. *AL-ISHLAH: Jurnal Pendidikan*, 13(1), 186–195. <https://doi.org/10.35445/alishlah.v13i1.385>
- Köksal, D., & Ulum, Ö. G. (2018). Language a sssessment through Bloom' s Taxonomy. *Journal of Language and Linguistic Studies*, 14(2), 76–88.
- Krathwohl, D. R. (2002). *A Revision of Bloom' s Taxonomy : An Overview*. 41(4).
- Kutsyuruba, B. (2017). Using Document Analysis Methodology to Explore Educational Reforms and Policy Changes in Post-Soviet Ukraine. In *Reimagining Utopias*. Sense. https://doi.org/10.1007/978-94-6351-011-0_12
- Lee, Y.-J., Mijung, K., Jin, Q., Yoon, H.-G., & Kenji, M. (2017). *East-Asian Primary Science Curricula An overview Using Revised Bloom's Taxonomy* (SpringerBr). Springer Nature. <https://doi.org/10.1007/978-981-10-2690-4>
- Lorin and Krathwohl, D. R. A. (2001). *Taxonomy for Learning, Teaching and Assessing* (L. Anderson, D. Krathwohl, P. Airasian, K. Cruikshank, R. Mayer, P. Pintrich, J. Raths, & M. Wittrock (eds.); A Bridged). Addison Wesley Longman, Inc.
- Morales-Obod, M., Valdez Ramirez, M. ., Satria, E., & Indriani, D. . (2020). Effectiveness on the use of mother tongue in teaching the concepts of fraction among second grade of elementary school pupils. *Journal for the Education of Gifted Young Scientists*, 8(1), 291–304.
- Morgan, H. (2022). Conducting a Qualitative Document Analysis. In *Qualitative Report* (Vol. 27, Issue 1). <https://doi.org/10.46743/2160-3715/2022.5044>
- Morris, M. M. (2016). *Evaluating university teaching and learning in an outcome-based model : replanting Bloom*.

- Muhayimana, T., Kwizera, L., & Nyirahabimana, M. R. (2022). Using Bloom's taxonomy to evaluate the cognitive levels of Primary Leaving English Exam questions in Rwandan schools. *Curriculum Perspectives*. <https://doi.org/10.1007/s41297-021-00156-2>
- Omar, N., Sufi, S., Hassan, R., & Arshad, H. (2012). *Automated analysis of exam questions according to bloom's taxonomy*. 59(1956), 297–303. <https://doi.org/10.1016/j.sbspro.2012.09.278>
- Rachmawati, D. L., & Purwati, O. (2021). EFL Teachers' Attitudes and Competence in Developing HOTS-Based Formative Assessment. *JEEES (Journal of English Educators Society)*, 6(2), 184–196. <https://doi.org/10.21070/jees.v6i2.1060>
- Rezaee, M., & Golshan, M. (2016). *Investigating the Cognitive Levels of English Final Exams Based on Bloom's Taxonomy*. 3(4), 57–68.
- Soleimani, H. (2016). *An Evaluation of TEFL Postgraduates' Testing Classroom Activities and Assignments Based on Bloom's Revised Taxonomy*. 6(4), 861–869. <https://doi.org/http://dx.doi.org/10.17507/tpls.0604.26>
- Taras, M. (2010). *Assessment for learning: assessing the theory and evidence*. 2(2), 3015–3022. <https://doi.org/10.1016/j.sbspro.2010.03.457>
- Wilson, L. O. (2016). Anderson and Krathwohl Bloom's Taxonomy Revised Understanding the New Version of Bloom's Taxonomy. *The Second Principle*, 1–8.
- Zareian, G., Davoudi, M., Heshmatifar, Z., & Rahimi, J. (2015). An Evaluation of Questions in Two ESP Coursebooks Based on Bloom's New Taxonomy of Cognitive Learning Domain. *International Journal of Education and Research*, 3(8), 313–326.
- Abahussain, M., Iqbal, M., & Khan, I. (2020). Standardization of EFL Undergraduate Skill Exam Papers: A Case Study at Majmaah University. *Arab World English Journal (AWEJ) Volume*, 4(December), 508–518. [https://doi.org/DOI: https://dx.doi.org/10.24093/awej/vol11no4.24](https://doi.org/DOI:https://dx.doi.org/10.24093/awej/vol11no4.24) Pp.363-381 Standardization
- Airasian, P. W., & Miranda, H. (2014). The Role of Assessment in the Revised Taxonomy. *Theory into Practice*, 41(4), 249–254.
- Al-Khayyat, A. S. J. (2020). Exploring the Educational Digital Learning Systems Utilized by EFL Postgraduate Students During the Covid19 Pandemic Quarantine. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(6), 11481–11488.
- Ary, D., Jacobs, L. C., Sorensen, C. K., & Walker, D. A. (2014). *Introduction to Research in Education 9th Edition* (M. Kerr (ed.); Ninth Edit). Cengage Learning.
- 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>
- Atiullah, K., Fitriati, S. W., & Rukmini, D. (2019). Using revised Bloom's taxonomy to evaluate Higher Order Thinking Skills (Hots) in reading comprehension questions of English textbook for year X of High School. *English Education Journal*, 9(4), 428–436.
- Baghaei, S, Bagheri, M. S., & Yamini, M. (2021). Learning Objectives of IELTS Listening and Reading Tests: Focusing on Revised Bloom's Taxonomy. *Research in English Language Pedagogy*, 9(1), 182–199. <https://doi.org/10.30486/relp.2021.1916940.1244>
- Baghaei, Samira, Bagheri, M. S., & Yamini, M. (2020). Analysis of IELTS and TOEFL reading and listening tests in terms of Revised Bloom's Taxonomy. *Cogent Education*, 7(1). <https://doi.org/10.1080/2331186X.2020.1720939>
- Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D., & Joughin, G. (2017). How university teachers design assessments: a cross-disciplinary study. *Higher Education*, 74, 49–64. <https://doi.org/10.1007/s10734-016-0027-7>
- Belarbi, Fatine & Bensafa, A. (2020). An Evaluation of the Algerian EFL Bacallaureate Exam under the Cognitive Domains of Bloom's Taxonomy. *Arab World English Journal (AWEJ)*, 11(4), 508–518.
- Bennett, R. E., & Service, E. T. (2014). *Formative assessment : A critical review This article was downloaded by : [Karlstads Universitetsbibliotek] Assessment in Education : Principles , Policy & Practice Formative assessment : a critical review*. April, 3–25. <https://doi.org/10.1080/0969594X.2010.513678>

- Bloom, B. (1956). *Taxonomy of Educational Goals*. David McKay Company.
- Chang, W., & Chung, M. (2009). *Automatic Applying Bloom's Taxonomy to Classify and Analysis the Cognition level of English Question Items*. 727–734. <https://doi.org/10.1109/JCPC.2009.5420087>
- Cheng, L., & Fox, J. (2017). *Assessment in the Language Classroom* (the first). Palgrave.
- Cullinane, A. (2015). Bloom's Taxonomy and its Use in Classroom Assessment. *NCE-MSTL*, 1(13).
- Ebadi, S. (2016). Exploring Bloom's Revised Taxonomy of Educational Objectives in TPSOL Textbooks. *Journal of Teaching Persian to Speakers of Other Languages*, 5(1), 65–93.
- Eber, P. A., & Parker, T. S. (1999). Assessing Student Learning : Applying Bloom's Taxonomy Bloom's Taxonomy. *National Organization of Human Service, 1956*, 45–54.
- Farhady, H., & Tavassoli, K. (2018). Developing a Language Assessment Knowledge Test for EFL Teachers: A Data-Driven Approach. *Iranian Journal of Language Teaching Research*, 6(3), 79–94. <https://doi.org/10.30466/IJLTR.2018.120602>
- Ginting, P., Hasnah, Y., Hasibuan, S. H., & Batubara, I. H. (2021). Evaluating Cognitive Level of Final Semester Examination Questions Based on Bloom's Revised Taxonomy. *AL-ISHLAH: Jurnal Pendidikan*, 13(1), 186–195. <https://doi.org/10.35445/alishlah.v13i1.385>
- Köksal, D., & Ulum, Ö. G. (2018). Language a sssessment through Bloom's Taxonomy. *Journal of Language and Linguistic Studies*, 14(2), 76–88.
- Krathwohl, D. R. (2002). *A Revision of Bloom's Taxonomy : An Overview*. 41(4).
- Kutsyuruba, B. (2017). Using Document Analysis Methodology to Explore Educational Reforms and Policy Changes in Post-Soviet Ukraine. In *Reimagining Utopias*. Sense. https://doi.org/10.1007/978-94-6351-011-0_12
- Lee, Y.-J., Mijung, K., Jin, Q., Yoon, H.-G., & Kenji, M. (2017). *East-Asian Primary Science Curricula An overview Using Revised Bloom's Taxonomy* (SpringerBr). Springer Nature. <https://doi.org/10.1007/978-981-10-2690-4>
- Lorin and Krathwohl, D. R. A. (2001). *Taxonomy for Learning, Teaching and Assessing* (L. Anderson, D. Krathwohl, P. Airasian, K. Cruikshank, R. Mayer, P. Pintrich, J. Raths, & M. Wittrock (eds.); A Bridged). Addison Wesley Longman, Inc.
- Morales-Obod, M., Valdez Ramirez, M. ., Satria, E., & Indriani, D. . (2020). Effectiveness on the use of mother tongue in teaching the concepts of fraction among second grade of elementary school pupils. *Journal for the Education of Gifted Young Scientists*, 8(1), 291–304.
- Morgan, H. (2022). Conducting a Qualitative Document Analysis. In *Qualitative Report* (Vol. 27, Issue 1). <https://doi.org/10.46743/2160-3715/2022.5044>
- Morris, M. M. (2016). *Evaluating university teaching and learning in an outcome-based model : replanting Bloom*.
- Muhayimana, T., Kwizera, L., & Nyirahabimana, M. R. (2022). Using Bloom's taxonomy to evaluate the cognitive levels of Primary Leaving English Exam questions in Rwandan schools. *Curriculum Perspectives*. <https://doi.org/10.1007/s41297-021-00156-2>
- Omar, N., Sufi, S., Hassan, R., & Arshad, H. (2012). *Automated analysis of exam questions according to bloom's taxonomy*. 59(1956), 297–303. <https://doi.org/10.1016/j.sbspro.2012.09.278>
- Rachmawati, D. L., & Purwati, O. (2021). EFL Teachers' Attitudes and Competence in Developing HOTS-Based Formative Assessment. *JEES (Journal of English Educators Society)*, 6(2), 184–196. <https://doi.org/10.21070/jees.v6i2.1060>
- Rezaee, M., & Golshan, M. (2016). *Investigating the Cognitive Levels of English Final Exams Based on Bloom's Taxonomy*. 3(4), 57–68.
- Soleimani, H. (2016). *An Evaluation of TEFL Postgraduates' Testing Classroom Activities and Assignments Based on Bloom's Revised Taxonomy*. 6(4), 861–869. <https://doi.org/http://dx.doi.org/10.17507/tpis.0604.26>
- Taras, M. (2010). *Assessment for learning : assessing the theory and evidence*. 2(2), 3015–3022.

Setyowati, Y., Susanto, S. & Munir, A. (2022). A revised bloom's taxonomy evaluation of formal written language test items. *World Journal on Educational Technology: Current Issues*. 14(5), 1317-1331. <https://doi.org/10.18844/wjet.v14i5.7296>

<https://doi.org/10.1016/j.sbspro.2010.03.457>

Wilson, L. O. (2016). Anderson and Krathwohl Bloom's Taxonomy Revised Understanding the New Version of Bloom's Taxonomy. *The Second Principle*, 1–8.

Zareian, G., Davoudi, M., Heshmatifar, Z., & Rahimi, J. (2015). An Evaluation of Questions in Two ESP Coursebooks Based on Bloom's New Taxonomy of Cognitive Learning Domain. *International Journal of Education and Research*, 3(8), 313–326.