

Integrating Bloom's Taxonomy into the Design of the Final English Test for Eighth Graders

Luky Dwi Ratnasari^{1*}, Syafi'ul Anam², Ali Mustofa³

Faculty of Languages and Arts, State University of Surabaya^{1,2,3}

luky.22013@mhs.unesa.ac.id¹, Syafiul.anam@unesa.ac.id², alimustofa@unesa.ac.id³

* is Corresponding Author

Abstract

Testing is essential in education because it helps teachers measure how much students have learned. Teachers typically use multiple-choice and essay questions for this purpose. This research examined what types of thinking skills English teachers focus on when creating final exams for eighth-grade students, using Bloom's educational framework as a guide. The study also looked at the difficulties teachers encounter while developing these tests. The researchers used interviews and systematic observation methods to gather information. They analyzed a 50-question test that included 45 multiple-choice items and 5 essay questions. The results revealed that the test mainly assessed three types of thinking: memorizing information, comprehending concepts, and analyzing material, with comprehension questions being the most common. Teachers identified two main challenges: limited time for test development and difficulty categorizing questions according to Bloom's taxonomy. To address these issues, they suggested scheduling test creation during holidays and offering clearer institutional guidelines. They also recommended enhancing teachers' understanding of Bloom's taxonomy to improve test design. This study contributes to English language assessment by revealing a strong focus on lower-order thinking skills. It highlights the need for professional development and institutional support to help teachers design more balanced assessments that include higher-order thinking skills. Such improvements can lead to more effective language assessment practices aligned with broader educational goals.

Submitted:
April 30, 2025
Accepted:
June 5, 2025
Revised:
June 21, 2025
Published:
June 29, 2025

Keywords: *Test; Assessment; Bloom's Revised Taxonomy; Cognitive Process Dimension.*

INTRODUCTION

Evaluation is a widely used term in education, yet often misunderstood—even by educators responsible for assessing student progress. Many struggle to grasp its broader purpose, despite its crucial role in shaping effective learning programs. Arifin and Suryanto (2011) stress that evaluation is essential for measuring learning outcomes and guiding instructional decisions. Bloom describes evaluation as a structured process of collecting evidence to determine student progress and assess learning changes. In the same way, Stufflebeam describes it as the organized collection and examination of data to help with making decisions (Ahmad, 2019).

In this study, evaluation specifically refers to assessing students' learning achievements at the end of a semester. As a central component of instruction, effective assessment involves testing, measurement, data analysis, and feedback—tools that inform teaching practices and foster student

growth. John (2013) emphasizes that meaningful feedback is vital, as it identifies areas requiring reinforcement and informs future instruction.

Good assessments need to match what students are supposed to learn in order to give accurate results. Lee (2019) explains that a task is any organized language activity that gets students to use, understand, or create the language they're studying. Teachers often choose multiple-choice and fill-in-the-blank questions because they're quick to grade and give consistent results. Other testing methods include short and long written responses, matching exercises, true/false questions, and descriptive activities. The 2013 curriculum divides assessments into three main types: written exams, spoken tests, and homework assignments (Permendikbud, 2016).

Assessment is not merely about assigning grades—it must enhance learning. For this to happen, it should be intentionally designed, aligned with objectives, and capable of offering actionable insights. Testing is a crucial component of the assessment process. Brown (2004) identifies five main types of language tests: 1) Language Aptitude Tests: Predict a learner's potential to acquire a new language; 2) Proficiency Tests: Evaluate overall language ability, independent of instruction; 3) Placement Tests: Assign students to appropriate levels within a program; 4) Diagnostic Tests: Identify specific areas of weakness for targeted support; 5) Achievement Tests: Measure student mastery of specific instructional content. Tests are formal instruments used to assess performance against set benchmarks (Anas, 2015). However, standardized tests may not fully reflect students' abilities, especially in communicative contexts. A more holistic approach—integrating multiple forms of assessment—can provide a comprehensive view of student progress.

High-quality assessments require deliberate construction. Brown (2004) outlines five key principles for classroom assessment: reliability, practicality, validity, authenticity, and washback. Brookhart (2010) complements these with three foundational guidelines: (1) define clear learning goals, (2) design tasks that reveal what students know and can do, and (3) establish transparent evaluation criteria. The choice of principles should be context-dependent and tailored to the assessment's purpose.

Furthermore, tests should measure the right kinds of thinking skills. Bloom's Updated Taxonomy organizes thinking into six stages: Remember, Understand, Apply, Analyze, Evaluate, and Create (Anderson & Krathwohl, 2001). These stages help teachers judge how deep and complex their students' thinking is. Anastasi and Urbina (2002) suggest that analyzing test questions before giving the test is important for finding poorly written items and making the test better overall.

This study looks at the *Penilaian Akhir Semester* (PAS), which is a final test given at the end of each semester to measure students' overall performance (Suryanto, 2015). The research took place at MTs YPM 1 Wonoayu, where the PAS follows the national education standards. Initial observations showed that while the Ma'arif Educational and Social Foundation supervises how these tests are created, the teachers had never formally analyzed their test questions to check if they were well-made.

Previous studies have examined test item analysis from various angles. For example, Amaliyah (2020) found a dominant use of lower-order cognitive levels—remembering, understanding, and applying—in high school English tests. Similarly, Ayaturrochim (2020) discovered that English textbooks emphasized remembering, limiting students' critical thinking development. Nisa (2021) assessed the validity of final English tests and identified inconsistencies despite alignment with test blueprints.

While previous studies have looked at textbooks, teacher-made tests, or high school evaluations, this research specifically examines how Bloom's Updated Taxonomy is used in the Final English Test for 8th-grade students at MTs YPM 1 Wonoayu. The study also investigates what difficulties teachers encounter when trying to create tests that match different levels of thinking skills.

This study's goal are to: Examine what types of thinking skills are included in the Final English Test for 8th grade students by using Bloom's Updated Taxonomy. Besides that, this study identifies the challenges teachers encounter in constructing test items that align with appropriate cognitive levels.

By addressing these objectives, the study contributes to improving English language assessment practices, ensuring they support meaningful learning and cognitive development.

METHOD

This research used a qualitative descriptive approach. As emphasized by Utari (2019), a research design functions as a structured framework that guides researchers through each phase—from question formulation to data analysis—ensuring clarity and methodological coherence. While this structure promotes organization, it may also limit adaptability in response to emerging insights during the research process.

The research took place at MTs YPM 1 Wonoayu, a private Islamic middle school in Sidoarjo, East Java. The focus was on the eighth-grade English teacher, as they were responsible for designing the semester final test analyzed in this study. The eighth-grade level was specifically selected because it represents a transitional academic stage where students are expected to consolidate prior learning in preparation for the more rigorous demands of ninth grade.

The participant—the teacher—was chosen through purposive sampling, based on their direct involvement in test development. This selection ensured relevance and depth in analyzing the alignment of test items with Bloom's Revised Taxonomy. Before data collection, the researcher obtained informed consent from the participant and ensured that all ethical considerations, including confidentiality and voluntary participation, were upheld.

Two instruments were used to collect data. To address the first research question, the researcher examined the Final English Test questions using a checklist based on the Cognitive Process Dimension Table from Anderson and Krathwohl's (2001) updated version of Bloom's Taxonomy. The checklist placed each test question into one of six thinking levels: Remember, Understand, Apply, Analyze, Evaluate, and Create. The checklist was custom-developed for this study, drawing on the theoretical framework provided by Anderson and Krathwohl (2001). To enhance its validity, the checklist was reviewed by two experts in English education and assessment. Inter-rater reliability was also addressed by involving a second coder to independently categorize a sample of test items, with discrepancies discussed and resolved to reach consensus.

To explore the teacher's experiences and challenges in constructing the test, a semi-structured interview was conducted. This format allowed for open-ended responses while still addressing specific themes, such as familiarity with Bloom's Revised Taxonomy, difficulties in aligning test items with cognitive levels, and institutional constraints. The interview was conducted in Indonesian, recorded with permission, and later transcribed for analysis. While interviews provide rich insights into participant perspectives, the responses may be influenced by personal bias or contextual pressures, and thus should be interpreted with caution. Data from the test item analysis were tabulated and categorized according to the cognitive dimensions defined by Anderson and Krathwohl (2001). Frequency counts were used to determine the distribution of cognitive levels within the test. Interview information was examined using thematic analysis, finding common themes related to the difficulties teachers face when creating tests.

RESULT

Level of Cognitive Process Dimension

The findings related to the first research question *“what level of cognitive process dimension in Bloom’s Revised Taxonomy that the teacher use in the construction of Final English Test for eight graders?”*

The result shows in this table:

Table 1. Cognitive Process Dimension Distribution in the Final English Test

| No. | Cognitive Process Dimension | Distribution Number |
|-----|-----------------------------|---|
| 1 | Understand | 4, 5, 8, 11, 12, 18, 19, 24, 25, 26, 27, 29, 30, 31, 32, 34, 35, 36, 37, 38, 40, 42, 43, 44, 46, 48, 49 |
| 2 | Analyze | 9, 10, 13, 14, 17, 20, 21, 22, 23, 28, 41, 47 |
| 3 | Remember | 1, 2, 3, 7, 15, 16, 50 |
| 4 | Evaluate | 39, 45 |
| 5 | Apply | 33 |
| 6 | Create | 6 |

The analysis of the Final English Test items for eighth-grade students revealed that the cognitive process dimension most frequently targeted was "Understand", with a total of 27 items (e.g., items 4, 5, 8, 11–12, 18–19, etc.). This indicates that the test heavily emphasizes comprehension-level thinking, such as interpreting, summarizing, and explaining information.

The second most frequent category was "Analyze", comprising 12 items (e.g., items 9, 10, 13–14, 17, 20–23, etc.), suggesting that the test moderately engaged students in higher-order thinking processes like differentiating and organizing information.

The "Remember" category included 7 items (e.g., items 1–3, 7, 15–16, 50), reflecting a smaller focus on recalling facts or basic information.

In contrast, higher-order categories such as "Evaluate" and "Create" were minimally represented. Only 2 items (items 39 and 45) tested evaluation skills, and just 1 item (item 6) required students to create original responses. Similarly, the "Apply" dimension was addressed in only 1 item (item 33).

Overall, the test focused mainly on basic to moderate thinking skills (Remember, Understand, Analyze), with very few questions testing advanced thinking skills (Apply, Evaluate, Create). This indicates that teachers need to create more balanced tests that include a broader variety of thinking skills to encourage deeper student learning and critical thinking abilities.

Challenges Faced by the Teacher in Designing the Final English Test

The information from the second research question, "What challenges do teachers encounter when creating the Final English Test for eighth graders using Bloom's Revised Taxonomy?"

Based on Bloom's Revised Taxonomy

Allocating Time for Test Creation

Creating a well-structured test presents a significant time management challenge for the teacher. The assessment consists of 50 items, including 45 multiple-choice questions and 5 essay questions, all of which must comply with the foundation's guidelines. The teacher highlights that developing a comprehensive test requires considerable effort, as each question must be thoughtfully designed, reviewed, and arranged. This underscores the difficulty of balancing test preparation with other teaching duties.

Classifying items according to the Levels of Cognitive Process Dimension

Another major difficulty is properly classifying test questions according to Bloom's Updated Taxonomy, which has six thinking levels: Remember, Understand, Apply, Analyze, Evaluate, and Create. The Ma'arif Educational and Social Foundation requires a specific breakdown: 25% for basic thinking skills, 50% for moderate skills, and 25% for advanced skills. However, even with these clear instructions, the teacher has trouble categorizing test questions correctly, which suggests there may be a lack of professional training or not enough access to helpful resources for proper implementation.

A critical perspective highlights that while such structured allocations aim to ensure balanced cognitive engagement, they may inadvertently constrain test design by imposing rigid numerical targets rather than fostering organic, meaningful assessment practices. Additionally, if teachers lack sufficient support in mastering Bloom's framework, these classification challenges could lead to misalignment between test content and students' actual cognitive development. Addressing these issues requires ongoing teacher training, better instructional resources, and greater flexibility in applying Bloom's model to accommodate real-world classroom dynamics. Overall, these challenges indicate that the test development process demands better time management strategies and improved support for cognitive level categorization to ensure a balanced and effective assessment.

According to Susan M. Brookhart's Theory

The teacher applies the basic rules of test creation from Susan M. Brookhart's theory. Based on the interview, the teacher stresses careful planning before making the test, making sure each assessment follows the established rules. This shows a dedication to creating a well-organized evaluation; however, it doesn't show whether the teacher thoughtfully adjusts these rules to fit the specific needs of students or the curriculum. While following guidelines is important, strictly applying them without flexibility might reduce how well the assessment can measure different types of learning results.

DISCUSSION

Level of Cognitive Process Dimension

The findings related to the first research question *"what level of cognitive process dimension in Bloom's Revised Taxonomy that the teacher use in the construction of Final English Test for eight graders?"*

Remember Level

Remembering happens when students recall information from their memory, whether they just learned it or knew it before. An example of a Remember level question is question number 1.

Teacher: Are you listening to me Tono?

Tono: ...

Tono will answer "..."

- A. Yes, I do.
- B. Yes, you do.
- C. Yes, I'm.
- D. Yes, you are.

In this question, students need to choose the right answer to finish the dialogue. This question falls under the Remember level. So, the correct answer is C.

Understand Level

Understanding happens when students get meaning from various sources, like messages, texts, and conversations. An example of an Understand level question is question number 4.

“One of the students is taking the ball”.

The negative form of the sentence above is ...

- A. One of the students not is taking the ball.
- B. Not one of the students is taking the ball.
- C. One of the students not taking the ball.
- D. One of the students is not taking the ball.

In this question, students need to explain how to make a given sentence negative. It belongs to the *Understand* level. Therefore, the correct answer is D.

Apply Level

Application happens when students use a particular method to do experiments or solve problems. An example of an Apply level question is question number 33.

A Short Message

When will the examination last?

- | | |
|----------------|---------------------|
| A. On 3rd June | C. On the 23rd June |
| B. On 7th June | D. On the 30th June |

In this question, students need to do a calculation using the information given in the text. It belongs to the *Apply* level. Therefore, the correct answer is B.

Analyze Level

Analyzing occurs when students break down a problem into its main parts, examine how these parts relate to each other, and evaluate how these connections might lead to possible difficulties. An example of a question at the Analyze level is question number 9.

The text below is for questions 8 through 10! I have two brothers named Jamal and Arif. I am the eldest child and Arif is the youngest. Jamal is overweight and weighs 76 kg. My weight is 60 kg. Arif is the skinniest child, but he is 170 cm tall. I am 5 cm shorter than Arif. Jamal is 160 cm tall.

The writer is ... than Jamal

The best comparative degree used to complete the statement is....

- | | |
|------------|-----------|
| A. younger | C. fatter |
| B. older | D. taller |

In this question, students need to select the best comparative form to finish the sentence. This question falls under the Analyze level. So, the correct answer is B.

Evaluate Level

Evaluation happens when students examine a situation or idea using set criteria and standards, including quality, effectiveness, efficiency, and consistency. An example of an Evaluate level question is question number 39.

The following song is for question no 39 and 40!

The Lyric of song: History by One Direction

The song mainly tells us about

- A. love and affection
- B. struggle and sacrifice
- C. happiness and sadness
- D. friendship and memories

In this question, students need to organize the answer based on the song's words. It belongs to the Evaluate level. Therefore, the correct answer is A.

Create Level

Creation means putting together different elements to make a complete whole, encouraging students to produce new results by rearranging and reshaping various parts into new structures or patterns. An example of a Create level question is question number 6.

Is - Lisa – sweeping - watering - the floor – her - and - are – parents – the flowers.

1 2 3 4 5 6 7 8 9 10

Please arrange the jumbled words above into a good sentence!

- A. 2-1-3-5-7-6-9-8-4-10
- B. 2-8-3-5-7-6-9-1-4-10
- C. 2-1-4-5-7-6-9-8-3-10
- D. 2-8-4-5-7-6-9-1-3-10

In this question, students need to put mixed-up words in order to make a meaningful sentence. It belongs to the *Create* level. Therefore, the best answer is B.

Challenges Faced by the Teacher in Designing the Final English Test

This section addresses the second research question:

"What challenges do teachers encounter when creating the Final English Test for eighth graders using Bloom's Revised Taxonomy?"

Information was gathered through a semi-structured interview with the teacher who was responsible for creating the assessment.

Based on Bloom's Revised Taxonomy

Time Constraints in Test Construction

One of the most significant challenges reported by the teacher was the limited time available for test preparation. The Final English Test comprises 50 items—45 multiple-choice and 5 essay questions. Crafting items that accurately reflect Bloom's cognitive levels requires thoughtful design, alignment with curriculum standards, and multiple review cycles. However, the teacher indicated that such thorough planning is often hindered by competing responsibilities:

"One major issue is the limited time provided by the institution for test development. Furthermore, teachers have multiple responsibilities, such as conducting lessons, grading assignments, and managing other academic tasks."

This time limitation can directly affect test quality, as insufficient planning may lead to an over-reliance on lower-order thinking questions (e.g., remembering or understanding), rather than promoting critical thinking and creativity. As a result, students may lose chances to participate in

deeper learning experiences that match advanced thinking skills like *Analyzing, Evaluating, and Creating*.

To address this issue, the teacher proposed a practical recommendation:

"I recommend that the foundation schedule test development sessions during school holidays. This would prevent test preparation from overlapping with other academic duties, thereby easing the workload on teachers."

This suggestion underscores a systemic challenge that could be improved through institutional policy. Schools and educational foundations should consider allocating dedicated, non-instructional periods for assessment design—particularly when teachers are expected to align tests with pedagogically rigorous frameworks such as Bloom's Revised Taxonomy. Such policy adjustments would not only support teacher well-being but also enhance the validity and depth of student assessments.

Classifying items according to the Levels of Cognitive Process Dimension

One of the main challenges the teacher faces when creating the Final English Test is correctly classifying test questions based on the Cognitive Process Dimension in Bloom's Updated Taxonomy. This system has six levels: Remember, Understand, Apply, Analyze, Evaluate, and Create, and having a balanced mix across these levels is crucial for good assessment.

The Ma'arif Educational and Social Foundation requires a specific breakdown: 25% for basic thinking skills, 50% for moderate thinking skills, and 25% for advanced thinking skills. However, the teacher has difficulty properly placing each test question into these thinking categories. The problem comes from action words that look similar but actually belong to different thinking levels. The teacher explains:

"One major challenge in classification is that many operational verbs seem interchangeable but actually belong to different domains. For example, the verb 'comparing' is associated with the 'apply' level but also appears in other levels. This overlap creates confusion because test designers may not fully grasp these distinctions."

Furthermore, the teacher raises concerns about how clear the provided guidelines are: "The directions given are not as specific as I hoped. They just say that 25% of the test should target basic thinking skills, 50% on moderate skills, and 25% on advanced skills without giving more details."

To address these challenges, the teacher stresses the importance of better understanding each thinking level. This would improve their skill in categorizing test questions more correctly. The teacher recommends:

"To address these challenges, especially regarding ambiguous operational verbs, we should seek guidance from experienced educators and participate in regular training sessions. This would help clarify the distinctions between different cognitive levels and ensure that test items are categorized correctly. With proper training, test designers can eliminate confusion and improve the accuracy of classification."

According to Susan M. Brookhart Theory

To clarify the teacher's approach to test design, the researcher asked:

"Does the teacher apply the three fundamental rules of test creation according to Susan M. Brookhart's theory?"

In response, the teacher confidently affirmed:

"Absolutely. Yes, I do."

The teacher then elaborated on the fundamental principles of assessment design, as outlined by Susan M. Brookhart:

- 1) Preparation Phase: *"I begin by clearly defining the type of thinking required, identifying the content to be assessed, and reviewing the learning objectives I aim to evaluate."*
- 2) Assessment Design: *"Next, I ensure that the test effectively prompts students to demonstrate both the necessary knowledge and cognitive skills."*
- 3) Interpreting Student Responses: *"After students complete the assessment, I analyze their responses to determine whether they provide valid evidence of the intended learning outcomes."*

The interview findings suggest that the teacher follows a structured and deliberate approach to test creation, ensuring alignment with Brookhart's principles to design meaningful and effective assessments.

CONCLUSION

Testing is an essential part of teaching and learning, helping educators measure how well students are performing and guide their teaching choices. This research examined how Bloom's Updated Taxonomy was used in the Final English Test for eighth-grade students and looked into the difficulties teachers face when matching test questions with the right thinking skill levels.

Using a qualitative descriptive approach, data were collected through an observation checklist and a teacher interview. The checklist was used to analyze a 50-item Final English Test, while the interview provided contextual insights into the teacher's test-construction experience.

The findings revealed a predominance of lower-order thinking skills, with most items categorized under the Remember, Understand, and Analyze levels. Higher-order dimensions—such as Evaluate and Create—were minimally represented. Two primary challenges emerged: (1) insufficient time allocated for test development due to overlapping academic duties, and (2) difficulty in correctly classifying test items according to Bloom's taxonomy.

To address these issues, the teacher suggested scheduling dedicated test preparation periods—preferably during school holidays—to reduce workload pressure. Additionally, professional development opportunities focused on assessment design and taxonomy-based question writing were identified as essential for improving test quality.

In sum, aligning assessments with Bloom's Revised Taxonomy not only enhances test validity but also promotes deeper student learning. Institutional support through time allocation and targeted training can significantly improve teachers' capacity to design cognitively balanced assessments.

REFERENCES

- Ahmad, S. (2019). *Educational assessment and evaluation*. Alfabeta.
- Amaliyah, A. (2020). An analysis of cognitive levels in English summative test items. *Journal of Language Teaching and Research*, 11(2), 145–152.
- Anas, I. (2015). *Assessment for learning*. Pustaka Pelajar.
- Anastasi, A., & Urbina, S. (2002). *Psychological testing* (7th ed.). Pearson Education.
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Arifin, Z., & Suryanto, H. (2011). *Evaluasi pembelajaran*. Direktorat Jenderal Pendidikan Islam.
- Ayaturrochim, A. (2020). Cognitive domain analysis in reading tasks of English in Focus textbook. *Journal of English and Education*, 8(1), 25–34.
- Brookhart, S. M. (2010). *How to create and use rubrics for formative assessment and grading*. ASCD.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education.

- John, R. (2013). *Assessment and student learning*. Routledge.
- Lee, J. F. (2019). *Tasks and communicating in language classrooms*. Heinle & Heinle.
- Ministry of National Education. (2007). *Guidelines for test item analysis*. BSNP.
- Nisa, N. K. (2021). An evaluation of final test items developed by MGMP. *Indonesian Journal of Applied Linguistics*, 11(1), 95–102.
- Permendikbud. (2016). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2016 tentang Standar Penilaian Pendidikan*. Kementerian Pendidikan dan Kebudayaan.
- Suryanto, H. (2015). *Evaluasi pembelajaran: Teori dan praktik*. UM Press.
- Utari, S. (2019). *Metodologi penelitian pendidikan*. Surya Pena Gemilang.