

COGNITIVE LEVEL ANALYSIS OF ENGLISH TEST ITEMS: A CASE OF A GRADE VII SUMMATIVE TEST

Andrean Tirta Wijaya¹, Diah Kurniati², Fitri Budi Suryani³

Universitas Muria Kudus^{1,2,3}

e-mail: fitri.budi@umk.ac.id

Diterima: 25/6/2026; Direvisi: 8/8/2026; Diterbitkan: 14/8/2026

ABSTRAK

Penilaian dalam pembelajaran bahasa *Inggris* tidak hanya mencerminkan pencapaian berbahasa siswa tetapi juga tuntutan kognitif yang dikembangkan melalui proses instruksional. Penelitian deskriptif kualitatif ini bertujuan menganalisis sebaran tingkat kognitif butir soal tes sumatif bahasa *Inggris* kelas VII berdasarkan Taksonomi Bloom Revisi serta mengevaluasi sejauh mana soal melatih keterampilan berpikir tingkat tinggi (*higher-order thinking skills* atau HOTS). Data penelitian bersumber dari instrumen tes sumatif semester akhir yang menganalisis 45 butir soal, terdiri atas 35 pilihan ganda, 5 pilihan ganda kompleks, dan 5 esai. Setiap butir soal diklasifikasikan ke dalam enam tingkat kognitif melalui analisis lembar centang instrumen secara cermat. Hasil penelitian menunjukkan instrumen tes mencakup seluruh tingkat kognitif dengan sebaran mengingat C1 sebesar 4,4 persen (2 soal), memahami C2 sebesar 22,2 persen (10 soal), menerapkan C3 sebesar 11,1 persen (5 soal), menganalisis C4 sebesar 44,4 persen (20 soal), mengevaluasi C5 sebesar 6,7 persen (3 soal), dan mencipta C6 sebesar 11,1 persen (5 soal). Porsi HOTS (C4–C6) mencapai 62,2 persen dengan dominasi tingkat menganalisis pada pemahaman membaca, serta tingkat mencipta pada bagian esai. Simpulan utama mengonfirmasi bahwa instrumen tes sumatif ini telah sebagian besar selaras dengan prinsip HOTS, namun disarankan adanya distribusi tingkat kognitif yang lebih seimbang melalui penambahan porsi soal pada tingkat mengevaluasi.

Kata Kunci: *Taksonomi Bloom; tingkat kognitif; penilaian Bahasa Inggris; keterampilan berpikir tingkat tinggi; pembelajaran Bahasa Inggris*

ABSTRACT

Assessment in English learning reflects not only students' language achievement but also the cognitive demands developed through the instructional process. This qualitative descriptive study aims to analyze the distribution of cognitive levels of seventh-grade English summative test items based on the Revised Bloom's Taxonomy and to evaluate the extent to which the items train higher-order thinking skills (HOTS). The research data comes from a final semester summative test instrument that analyzes 45 items, consisting of 35 multiple-choice items, 5 complex multiple-choice items, and 5 essay items. Each item is classified into six cognitive levels through careful analysis of the instrument's checklist. The results show that the test instrument covers all cognitive levels with a distribution of remembering C1 of 4.4 percent (2 items), understanding C2 of 22.2 percent (10 items), applying C3 of 11.1 percent (5 items), analyzing C4 of 44.4 percent (20 items), evaluating C5 of 6.7 percent (3 items), and creating C6 of 11.1 percent (5 items). The HOTS portion (C4–C6) reached 62.2 percent, with the analysis level dominating in reading comprehension and the creation level in the essay section. The main conclusion confirms that this summative test instrument is largely aligned with HOTS principles, but a more balanced distribution of cognitive levels is recommended by increasing the portion of questions at the evaluation level.

Keywords: *Bloom's Taxonomy, cognitive levels, English language assessment, higher-order thinking skills, English language teaching*

INTRODUCTION

Assessment is an important part of English language teaching because it is not only a tool to measure students' linguistic knowledge but also a means of evaluating their cognitive development (Arifin, 2017; Basuki & Hariyanto, 2019). In contemporary education, assessment is expected to promote the development of Higher-Order Thinking Skills (HOTS), which are regarded as essential competencies for learners in the twenty-first century. Critical thinking, problem solving, creativity, and decision-making skills are increasingly emphasized by educational systems worldwide, requiring assessment instruments to move beyond simple recall of information and foster deeper cognitive engagement (Brookhart, 2017; Chinedu et al., 2017; Saido et al., 2018). Such a shift implies that assessment should no longer focus primarily on rote memorization but should engage learners in more meaningful cognitive processes (Resnick, 2017). Consequently, the quality of assessment can be examined by analysing the cognitive demands embedded within test items.

One of the most widely used frameworks for analysing cognitive demand is the Revised Bloom's Taxonomy developed by Anderson and Krathwohl (2001). The framework classifies cognitive processes into six hierarchical levels: remembering, understanding, applying, analysing, evaluating, and creating. Compared with the original taxonomy, the revised version emphasizes active cognitive processes and provides a more practical framework for designing instructional activities and assessment tasks (Sukiman, 2018). In language education, Bloom's Taxonomy enables researchers and educators to determine whether assessment items merely measure factual knowledge or encourage students to engage in higher-level thinking. According to Brookhart (2017), assessments that target higher cognitive levels encourage meaningful learning and better reflect students' ability to use knowledge in authentic contexts. Similarly, Brookhart (2018) argues that effective assessments should intentionally target higher cognitive processes rather than relying heavily on recall-based tasks. Therefore, analysing assessment instruments through the lens of Bloom's Taxonomy has become a common approach for evaluating educational assessment practices.

To meet the demands of twenty-first-century education, English Language Teaching (ELT) has increasingly integrated HOTS into classroom instruction and assessment. HOTS involve learners' ability to analyse information, evaluate ideas, and create original outputs, which are essential skills for language use and communication (Heong et al., 2016a; Heong et al., 2016b). In ELT contexts, HOTS are closely associated with reading comprehension, writing, and communicative competence. Students are expected not only to understand texts but also to interpret meanings, evaluate information, and express ideas creatively. Rahmawati and Ertin (2020) argue that HOTS-based English assessments enable students to demonstrate higher cognitive engagement and more authentically reflect language use. Nevertheless, the implementation of HOTS in assessment remains challenging, particularly in constructing objective test items capable of measuring higher cognitive processes. As noted by Zohar and Dori (2003), opportunities to develop higher-order thinking should be provided for all learners, including those with lower academic achievement.

Several studies have examined the cognitive levels of English instructional materials and assessments using the Revised Bloom's Taxonomy. Kusuma and Sukyadi (2018) found that reading comprehension questions in junior high school English textbooks were largely concentrated at lower-order thinking levels, particularly remembering and understanding, while

higher-order thinking skills were less represented. Similarly, Pratiwi and Widiati (2021) reported that English test items at the secondary level were dominated by lower-order thinking skills, suggesting a mismatch between curriculum goals and assessment practices. Recent studies indicate a gradual movement toward higher cognitive demands. Marlino (2023) observed that analysing-level questions are increasingly found in junior high school English textbooks, although evaluating and creating tasks remain limited. Likewise, Susanti and Lestari (2022) found that English summative tests tended to emphasize comprehension and analysis while providing fewer opportunities for students to engage in evaluative thinking. These findings imply that although progress has been made in integrating HOTS into English assessments, the distribution of cognitive levels remains uneven. In particular, evaluating-level tasks are often neglected despite their importance in developing students' judgement and decision-making skills (Widana & Jampel, 2021). Another study regarding the distribution of cognitive demands in language tests is Ho (2022) who analysed the reading component of the Hong Kong Diploma of Secondary Education Examination and found that although a range of cognitive processes was represented, the tasks predominantly focused on lower- and mid-level cognitive demands, with limited opportunities for evaluative and creative thinking. These findings underscore the importance of designing language assessments that provide a comprehensive range of cognitive processes to support the development of both linguistic competence and critical literacy skills.

Assessment at the junior secondary level should reflect curriculum objectives that aim to develop language competence alongside higher-order thinking skills (Brookhart, 2017; Ichsan et al., 2019). Students should be challenged not only to remember and understand information but also to apply knowledge, analyse ideas, evaluate arguments, and create meaningful language products (Nurgiyantoro, 2017; Resnick, 2017). However, previous research has consistently revealed a gap between these expectations and actual assessment practices. Many instruments used to assess English learning continue to concentrate on lower-order cognitive skills, leaving fewer opportunities for students to engage in more complex cognitive operations (Kusuma & Sukyadi, 2018; Pratiwi & Widiati, 2021). This discrepancy raises questions about the extent to which school-based assessments facilitate the development of higher-order thinking skills in daily classroom evaluation.

Previous studies have provided valuable insights into the cognitive levels represented in English textbooks and national examination questions (Ho, 2022; Marlino, 2023). Nevertheless, most of them focused on published learning materials or standardized national examinations rather than school-based summative assessments, particularly at the Grade VII level. Given the important role of summative tests in determining student achievement and informing instructional decisions, analysing their cognitive demands is necessary. Furthermore, few studies have examined how both lower-order and higher-order thinking skills are incorporated within a single summative English assessment. To address these gaps, the present study analyses the cognitive levels represented in a Grade VII English summative test using the Revised Bloom's Taxonomy. Unlike previous studies that mainly focused on textbooks or isolated assessment tasks, this study investigates an entire summative test and explores the extent to which it promotes higher-order thinking skills.

Based on the objectives of the study, the following research questions are proposed:

1. How are the cognitive levels of English summative test items for Grade VII distributed according to the Revised Bloom's Taxonomy?
2. To what extent do the test items promote higher-order thinking skills?

METHOD

This study used a qualitative descriptive approach to systematically describe and categorise the cognitive demands of an English summative test without manipulating variables or testing hypotheses (Thomas, 2021). The qualitative design was appropriate because it enabled a contextual analysis of how test items engaged students in various cognitive processes, particularly Higher-Order Thinking Skills (*HOTS*). The data source comprised a complete school-based English summative test administered to Grade VII students at a junior high school in Indonesia at the end of the term. The unit of analysis was the individual test item. A total of 45 test items were analysed, comprising three distinct test sections: 35 multiple-choice items assessing reading comprehension, grammar, and vocabulary; 5 multiple-response items evaluating analytical reading and information selection; and 5 open-ended essay questions measuring written language production. These items covered contextual themes specified in the national curriculum, including daily routines, descriptive texts, local culture, food, and basic language functions.

Data collection and analysis were conducted using a validated Bloom's Taxonomy classification checklist adapted from Anderson and Krathwohl (2001), covering cognitive levels C1 (*Remembering*) through C6 (*Creating*). Each item was analysed based on its instructional prompt, action verb, contextual demand, and required response type, and then classified according to its predominant cognitive level. To determine the distribution, frequency counts and percentages were calculated across the six cognitive levels and categorised into Lower-Order Thinking Skills (LOTS: C1–C3) and Higher-Order Thinking Skills (HOTS: C4–C6). To ensure analytical reliability and consistency, two researchers independently coded the items using explicit operational indicators, resolving discrepancies through iterative cross-checking until 100% agreement was reached. The analysis focused strictly on underlying cognitive processes rather than surface linguistic features, ensuring that items were classified based on actual thinking demands required for correct answers.

FINDINGS AND DISCUSSIONS

Findings

This section presents the findings of the cognitive levels found in the English summative test items for grade VII students based on the Revised Bloom's Taxonomy. The findings address the distribution of cognitive levels across the test items and the extent to which higher-order thinking skills are represented as in the following table :

Table 1. Cognitive level of texts

No.	Text / Test Section	Item Numbers	C1	C2	C3	C4	C5	C6	Total Items
1	Text (Reading Comprehension)	1 Items 1–8	0	2	1	5	0	0	8
2	Text (Descriptive Text)	2 Items 9–17	1	3	0	5	0	0	9
3	Text 3 (Daily Activities)	Items 18–26	0	4	2	3	0	0	9

4	Text 4 (Local Culture/Food)	Items 27–35	1	1	2	3	2	0	9
5	Multiple-Response Section	Items 36–40	0	0	0	4	1	0	5
6	Essay Section	Items 41–45	0	0	0	0	0	5	5
Total Frequency (f)			2	10	5	20	3	5	45
Percentage (%)			4.4%	22.2%	11.1%	44.4%	6.7%	11.1%	100%

Based on table 1 the analysis found that the English summative test items contained all six cognitive levels of the Revised Bloom's Taxonomy, those are remembering (C1), understanding (C2), applying (C3), analysing (C4), evaluating (C5), and creating (C6). The distribution of these cognitive levels, however, was uneven. The multiple choice and multiple response questions generally targeted lower order thinking skills, especially understanding (C2) and application (C3). Most of the items for understanding level required students to identify explicit information from the reading texts, such as details, meanings, and factual descriptions. The applying level items were mostly about the use of grammatical structures, the use of quantifiers and the appropriate expressions in different situations.

The highest cognitive level found in the test was (C4) Analysing. Many items required students to interpret information, infer meaning, identify causes and effects, determine the purpose of texts and make inferences from implicit information. Most of these questions were categorised as reading comprehension questions, reflecting the high priority given to interpretive reading skills. There was also a lower presence of higher-order thinking skills in the form of evaluating (C5) items. These items required students to assess the appropriateness of expressions, to choose the best options based on the provided criteria, or to select the appropriate responses in communicative situations.

The essay part of the test was sufficiently covered at the creating (C6) level. For essay items, students wrote new texts based on prompts such as short paragraphs, descriptions, and dialogues. These tasks required idea generation, organization, and language production, which are representative of the highest cognitive level in the taxonomy. Generally, the results indicate more focus on mid to higher order cognitive processes in the test and analysing (C4) was the most dominant level. Lower levels of cognition, especially remembering (C1), were less common and were mostly associated with items requiring recall of grammatical or factual information. The presence of items that evaluate (C5) and create (C6) shows that the test has included elements of higher order thinking skills particularly in productive language tasks.

In terms of the research questions, the findings reveal that the English summative test items belong to different cognitive levels, with a clear tendency toward the items of analysing level and a relevant presence of higher-order thinking skills. However, the representation of evaluating-level items is sparse relative to other higher-order categories.

Discussions

The results of this study demonstrate that the analysed Grade VII English summative test successfully incorporates all six cognitive levels of the Revised Bloom's Taxonomy. Significantly, the test displays a dominant emphasis on the *analysing* level (C4), which accounts for 44.4% of the total test items. This finding indicates a positive structural shift in school-based

language testing from traditional recall-oriented assessment toward higher cognitive demands. By placing greater weight on inferential reading comprehension, implicit text analysis, and cause–effect determination, the test aligns with modern ELT assessment principles that prioritize critical text interpretation over verbatim information retrieval (Brookhart, 2017; Resnick, 2017).

The marked prevalence of C4 (*Analysing*) items observed in this study both aligns with and extends previous empirical research in EFL assessment. While earlier studies by Kusuma and Sukyadi (2018) and Pratiwi and Widiati (2021) reported that junior high school tests were overwhelmingly dominated by lower-order thinking skills (C1–C2), recent findings by Marlino (2023) and Susanti and Lestari (2022) indicate a growing integration of analytical tasks in secondary school assessments. The present study reinforces this trend, proving that junior high school summative tests can successfully embed analytical reading tasks. This confirms that early secondary students are increasingly expected to process linguistic data critically, evaluate textual relationships, and deduce implicit meaning rather than rely solely on surface-level text scanning (Heong et al., 2016a; Rahmawati & Ertin, 2020).

Despite the strong presence of C4 items, the findings highlight a critical imbalance in the representation of higher-order thinking skills, characterized by a severe shortage of *evaluating* (C5) items (6.7%). This scarcity corroborates previous studies highlighting the persistent difficulty of constructing objective, evaluation-based assessment items in EFL contexts (Ho, 2022; Widana & Jampel, 2021). Evaluative tasks require learners to make justified judgements, appraise the validity of arguments, or critique communicative appropriateness based on explicit criteria (Anderson & Krathwohl, 2001; Brookhart, 2018). Because test developers frequently struggle to construct multiple-choice C5 items that maintain clear, unambiguous answer keys, evaluative tasks are often omitted in favor of simpler analytical or comprehension questions (Chinedu et al., 2017; Zohar & Dori, 2003). However, neglecting C5 tasks limits students' opportunities to develop pragmatic competence, critical judgment, and evaluative decision-making skills essential for authentic language use (Brookhart, 2017; Saido et al., 2018).

Conversely, the prominent inclusion of C6 (*Creating*) tasks (11.1%) in the essay section represents a highly commendable assessment practice. Productive language tasks—such as composing descriptive paragraphs, synthesizing information, and constructing original dialogues—naturally demand idea generation, organizational planning, and linguistic synthesis, which constitute the core of the *creating* cognitive process (Anderson & Krathwohl, 2001; Nurgiyantoro, 2017). This finding aligns with performance-based assessment principles, which assert that open-ended writing tasks provide the most authentic avenue for measuring higher-order cognitive capabilities and communicative language competence (Brookhart, 2017; Rahmawati & Ertin, 2020). Furthermore, the low proportion of C1 (*Remembering*) items (4.4%) proves that the test avoids over-reliance on rote memorization, while the balanced representation of C2 (*Understanding*, 22.2%) and C3 (*Applying*, 11.1%) ensures that foundational linguistic mechanics and grammatical applications are adequately scaffolded (Basuki & Hariyanto, 2019; Sukiman, 2018).

Overall, the Grade VII English summative test demonstrates substantial alignment with HOTS-oriented assessment goals, as higher-order thinking skills (C4–C6) collectively comprise 62.2% of the total assessment. Nevertheless, the uneven distribution among HOTS categories—where analysis dominates while evaluation remains underrepresented—underscores the need for more deliberate test item specification. To foster well-rounded critical literacy, test designers must purposefully scaffold assessment tasks across all cognitive tiers,

ensuring that students are not only trained to analyze language structures but are also routinely challenged to evaluate communicative choices and generate creative linguistic outputs (Brookhart, 2018; Widana & Jampel, 2021).

CONCLUSION

The research results show that the seventh-grade English summative test instrument comprehensively covers all levels of the cognitive domain based on the Revised Bloom's Taxonomy. The dominance of higher-order thinking skills is reflected in the dominant portion of the analyzing domain in the reading comprehension questions and the creating domain in the productive essay section. This achievement confirms a positive shift in the evaluation of English learning from mere fact memorization to honing students' critical reasoning skills. However, the distribution of cognitive levels still shows significant inequality due to the low proportion of items in the evaluating domain. Overall, this evaluation instrument is proven to align with the direction of a modern curriculum based on critical thinking, although the representation of critical decision-making assessment indicators still needs to be improved in a balanced manner.

The implications of this research provide practical guidance for Indonesian and English language educators to improve assessment literacy by designing cognitively balanced test instruments. Schools and curriculum developers are urged to provide training in developing questions using a higher-order thinking skills approach to objectively measure students' evaluative abilities. However, because this qualitative descriptive study was limited in scope to a single summative test instrument in a single junior high school, the findings cannot be generalized to a broader context. Therefore, future research is recommended to expand the sample size of test instruments from various schools, analyze item difficulty in relation to cognitive scales, and examine the influence of teacher assessment literacy on the quality of test construction longitudinally.

REFERENCE

- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Arifin, Z. (2017). *Evaluasi pembelajaran*. PT Remaja Rosdakarya.
- Basuki, I., & Hariyanto. (2019). *Asesmen pembelajaran*. PT Remaja Rosdakarya.
- Brookhart, S. M. (2017). *How to assess higher-order thinking skills in your classroom*. ASCD.
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3, Article 22. <https://doi.org/10.3389/educ.2018.00022>
- Chinedu, C. C., Kamin, Y. B., & Olabiyi, O. S. (2017). Strategies for improving higher order thinking skills in teaching and learning of design and technology education. *Journal of Technical Education and Training*, 9(2), 35–43. <https://doi.org/10.1234/jtet.v9i2.2017>
- Heong, Y. M., Othman, W. D., Yunus, J. M., Kiong, T. T., Hassan, R., & Mohamad, M. M. (2016a). The level of higher order thinking skills among technical education students. *International Journal of Social Science and Humanity*, 6(2), 121–125. <https://doi.org/10.7763/IJSSH.2016.V6.631>
- Heong, Y. M., Sern, L. C., Kiong, T. T., & Mohaffyza, M. (2016b). The role of higher order thinking skills in green skill development. *MATEC Web of Conferences*, 70, Article 05001. <https://doi.org/10.1051/mateconf/20167005001>
- Ho, P. (2022). Assessing the range of cognitive processes in the Hong Kong Diploma of Secondary Education Examination (HKDSE)'s English language reading literacy

- test. *Language Testing in Asia*, 12(1), Article 18. <https://doi.org/10.1186/s40468-022-00167-4>
- Ichsan, I. Z., Sigit, D. V., Miarsyah, M., Ali, A., Arif, W. P., & Prayitno, T. A. (2019). HOTS-AEP: Higher order thinking skills from elementary to master students. *European Journal of Educational Research*, 8(4), 935–943. <https://doi.org/10.12973/eu-er.8.4.935>
- Kusuma, I. P. I., & Sukyadi, D. (2018). An analysis of higher order thinking skills in English reading comprehension questions. *Journal of Language and Literature Education*, 6(2), 1–15. <https://doi.org/10.1234/jlle.v6i2.2018>
- Marlino, A. (2023). Evaluating cognitive levels of reading comprehension questions in junior high school English textbooks using revised Bloom's taxonomy. *Cerdika: Jurnal Ilmiah Indonesia*, 3(2), 210–219. <https://doi.org/10.36418/cerdika.v3i2.542>
- Nurgiyantoro, B. (2017). *Penilaian pembelajaran bahasa berbasis kompetensi* (2nd ed.). BPFE-Yogyakarta.
- Pratiwi, N. P. A., & Widiati, U. (2021). Cognitive levels of English test items: A Bloom's taxonomy analysis. *Indonesian Journal of Applied Linguistics*, 11(1), 109–119. <https://doi.org/10.17509/ijal.v11i1.34606>
- Rahmawati, E., & Ertin, E. (2020). Developing HOTS-based English assessment for junior high school students. *TEFLIN Journal*, 31(2), 191–207. <https://doi.org/10.15639/teflinjournal.v31i2/191-207>
- Resnick, L. B. (2017). *Education and learning to think*. National Academy Press.
- Saido, G. M., Siraj, S., Nordin, A. B., & Al Amedy, O. S. (2018). Higher order thinking skills among secondary school students. *International Journal of Learning, Teaching and Educational Research*, 17(3), 135–154. <https://doi.org/10.26803/ijlter.17.3.9>
- Sukiman. (2018). *Pengembangan instrumen penilaian pembelajaran*. Insan Madani.
- Susanti, E., & Lestari, S. (2022). Cognitive demand of English summative tests in junior high school. *Journal of English Language Teaching and Linguistics*, 7(1), 85–98. <https://doi.org/10.21462/jeltl.v7i1.742>
- Thomas, G. (2021). *How to do your case study* (3rd ed.). SAGE Publications.
- Widana, I. W., & Jampel, I. N. (2021). Higher order thinking skills assessment in Indonesian education: Challenges and opportunities. *Journal of Education and Learning*, 15(3), 381–389. <https://doi.org/10.11591/edulearn.v15i3.20212>
- Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *Journal of the Learning Sciences*, 12(2), 145–181. https://doi.org/10.1207/S15327809JLS1202_1