MIKAILALSYS
Journal of Multidisciplinary Sciences

e-ISSN : 2987-2286
p-ISSN : 2987-3924

Index: Harvard, Boston, Sydney, Copernicus, Sinta 4, Scilit, Semantic, Google, etc

https://doi.org/10.58578/mikailalsys.v4i2.10984

21st-Century EFL Assessment Innovation: HOTS-Based Evaluation Instruments in an Indonesian Government-Published English Textbook

Yeli Zastika, Syahrul Ramadhan, Yenni Rozimela
Universitas Negeri Padang, Indonesia
yelizastika@gmail.com; syahrul_r@fbs.unp.ac.id

Article Info:

Submitted:	Revised:	Accepted:	Published:
May 19, 2026	Jun 16, 2026	Jun 28, 2026	Jul 3, 2026

Abstract

Although the distribution of Higher Order Thinking Skills (HOTS) in English language textbooks has received increasing scholarly attention, research specifically examining the cognitive profile of evaluation instruments across six language skills and Genre-Based Approach (GBA) stages in Merdeka Curriculum senior high school textbooks remains limited. This study aims to analyze the distribution of HOTS and Lower Order Thinking Skills (LOTS) in the Grade XI English textbook *English for Change* and identify gaps in its alignment with 21st-century assessment needs. A mixed-methods sequential explanatory design was employed, involving 211 evaluation instruments selected through total sampling across five units and six language skills. Data were collected using a structured coding sheet and analyzed through descriptive quantitative frequency analysis, followed by descriptive interpretive content analysis. The findings indicate that HOTS accounts for 58.8% of the total evaluation instruments, with Creating (C6) as the most dominant cognitive level (25.6%) and Understanding (C2) as the least represented (6.6%). Writing (77.1%) and Presenting (76.7%) show the highest HOTS proportions, whereas Listening

remains LOTS-dominant (52.8%), and all End-of-Unit Reflection instruments are entirely LOTS-based (100%). These findings suggest a productively asymmetric distribution of HOTS that supports learning quality while revealing the need for stronger cognitive scaffolding in critical listening, analytical viewing, and evaluative reflection. Theoretically, this study contributes to the application of Revised Bloom's Taxonomy in cross-skill textbook analysis and expands understanding of cognitive scaffolding within GBA-based instructional design. Practically, it provides a replicable needs-analysis framework and recommends that textbook developers, teachers, and policymakers strengthen evaluation instruments to better support 21st-century assessment in the Merdeka Curriculum. This study also opens opportunities for comparative research across Merdeka Curriculum English textbooks in Indonesia.

Keywords: Higher Order Thinking Skills; Lower Order Thinking Skills; Evaluation Instruments; *English for Change*; Merdeka Curriculum

INTRODUCTION

Assessment is a strategic component that shapes the overall quality of education. In 21st-century learning, evaluation instruments must go beyond measuring recall and comprehension to stimulate higher-order cognitive abilities actively (Kemendikbudristek, 2022; Marzuki et al., 2023; Wardarita et al., 2024). This imperative is further underscored by Indonesia's PISA 2022 results, which showed an average student reading score of 371, well below the OECD average of 476, with deficiencies in critical inference identified as a primary cause (OECD, 2023). Textbooks, as the main learning resource in schools, directly shape students' cognitive orientation through the tasks and questions embedded in them (Ariawan et al., 2023; Ekalia et al., 2025; Wibowo et al., 2024). Government-issued English textbooks under the Merdeka Curriculum, including Work in Progress (Grade X), English for Change (Grade XI), and English for Nusantara (Junior High), are used by millions of students nationwide, positioning them as instruments of considerable curricular reach (Indar et al., 2023; Kemendikbudristek, 2022). The Merdeka Curriculum explicitly demands HOTS-oriented learning and 21st-century competency development, yet these goals can only be realized if the evaluation instruments embedded in textbooks are genuinely designed to achieve them (Barlian et al., 2022; Pratiwi, 2025; Sujinem, 2023). Systematic needs analysis of evaluation instruments is therefore essential to audit what exists and to map what remains to be developed for truly innovative and responsive learning assessment.

Government-issued English textbooks have been the subject of ongoing feasibility evaluations since the introduction of the Merdeka Curriculum. Most studies apply BSNP criteria alongside Cunningsworth's or Tomlinson's frameworks to assess content, language, and presentation quality. Studies of (Handayani & Amelia, 2023; Indar et al., 2023; Tobing et al., 2021) generally found these textbooks to be feasible and aligned with the goals of the Merdeka Curriculum. However, studies employing the CEFR framework reveal a more critical picture: (Pratycia & Zahro, 2025) found only moderate alignment with CEFR B1, particularly in listening and speaking tasks that lack authentic, interactive communicative contexts. Meanwhile, (Permatasari & Arsyad, 2023; Ripalga & Fitrawati, 2023; Rohimajaya et al., 2022) further highlight that texts in senior high school textbooks fall short in terms of textual coherence and level-appropriate difficulty. Regarding genre distribution, (Indar et al., 2023) and (Handayani & Amelia, 2023) report an uneven genre spread that does not adequately reflect the genre competency demands of the curriculum. While general feasibility can be affirmed, the cognitive alignment of embedded evaluation instruments with curriculum demands warrants deeper scrutiny.

Studies examining the cognitive levels of evaluation instruments in Kemendikbud English textbooks through the Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001) reveal inconsistent patterns across books and language skills. For Work in Progress Grade X, (Husni & Ginting, 2023; Manalu & Harahap, 2025; Ripalga & Fitrawati, 2023) found that reading tasks tend to be HOTS-dominant, particularly at the Analyzing level (C4). Conversely, studies of English for Nusantara yield contrasting results: (Ardiyansari & Jemadi, 2024; Ekalia et al., 2025; Ramadhani, 2023) found LOTS dominance, with Remembering (C1) and Understanding (C2) most frequent and Creating (C6) nearly absent. (Fitriani & Kirana, 2021; Wibowo et al., 2024) confirm that HOTS distribution is uneven across units and language skills, establishing the need for more comprehensive, multi-dimensional examination. Taken together, these findings suggest that single-skill analyses are insufficient to capture the full cognitive profile of any given textbook and that a cross-skill, multi-dimensional analytical framework is warranted.

Cultural content and value representation have received growing attention since the Merdeka Curriculum's emphasis on the Pancasila Student Profile. (Deswila et al., 2021; Faya et al., 2023; Lestari et al., 2023) found that target and international cultures dominate representation in Kemendikbud English textbooks, while Indonesian source culture remains underrepresented. Studies on intercultural communicative competence (ICC) reveal a

comparable pattern: (Astrid & others, 2024; Bahri & Agustina, 2023; Ramadhani & others, 2025) found that ICC content is uneven and skewed toward global cultural perspectives. Regarding Pancasila values, (Pratycia & Zahro, 2025; Ravyansah & Abdillah, 2021) found that critical reasoning was the most represented dimension, while global diversity and self-reliance require stronger integration. Collectively, these studies confirm that achieving a meaningful balance among local, global, and national values in textbook content remains a persistent challenge for curriculum developers and textbook authors, one that has yet to be satisfactorily resolved across successive curriculum reforms.

The linguistic characteristics of Kemendikbud English textbooks, including text readability, genre distribution, and language skill proportions, have similarly attracted scholarly attention. Readability analyses using the Dale-Chall Formula and the Cloze Procedure (Permatasari & Arsyad, 2023) and comprehensive content analysis. Regarding skill proportions, (Husni & Ginting, 2023; Paramitha et al., 2023; Rohimajaya et al., 2022) found that Writing consistently dominates the distribution, while Listening holds the lowest proportion. Genre distribution studies by (Febraningrum & Suroso, 2023; Handayani & Amelia, 2023; Pratycia & Zahro, 2025) show an uneven distribution that does not adequately reflect the genre-competency demands of the Merdeka Curriculum. These structural imbalances in the proportions of skills and genre representation have direct implications for the range and quality of evaluation instruments available for each language skill, thereby constraining the textbook's capacity to support the holistic development of language competence.

The integration of 21st-century skills and digital literacy in Kemendikbud English textbooks constitutes a nascent yet rapidly expanding area of scholarly inquiry. Studies of (Anwar et al., 2024; Rohimajaya et al., 2022; Sujinem, 2023) found that the 4C skills, namely Critical Thinking, Collaboration, Communication, and Creativity, have been integrated into Merdeka Curriculum textbooks, though unevenly across units and skills. On digital literacy, (Astrid & others, 2024; Palupi & Subianto, 2024; Pratycia & Zahro, 2025) found that functional skills dominate while critical evaluation of digital information remains severely limited. (Ariawan et al., 2023; Pratiwi, 2025; Wibowo et al., 2024) further reveal a persistent gap between stated 21st-century skill demands and the actual quality of available evaluation instruments. Despite evident progress toward 21st-century curricular alignment, the evaluation instruments embedded in these textbooks remain insufficiently examined for their

capacity to genuinely measure and stimulate the competencies they are intended to develop, leaving a substantive analytical gap in the existing literature.

In light of these gaps, a review of the five research clusters delineated above reveals significant unaddressed lacunae. First, no prior study has examined the cognitive levels of evaluation instruments specifically in English for Change Grade XI, as HOTS-LOTS studies to date have focused exclusively on Work in Progress Grade X and English for Nusantara (Husni & Ginting, 2023; Ekalia et al., 2025; Ripalga & Fitrawati, 2023). Second, all existing HOTS-LOTS analyses cover only one language skill, typically Reading, which precludes a comprehensive cross-skill cognitive picture (Ardiyansari & Jemadi, 2024; Fitriani & Kirana, 2021). Third, no study has employed GBA stages as the unit of analysis for mapping HOTS-LOTS distribution, nor has any examined End-of-Unit Reflection instruments as formal assessment components. Fourth, all prior studies are descriptive-evaluative in orientation, without situating cognitive findings within the broader framework of 21st-century learning demands (Anwar et al., 2024; Pratiwi, 2025; Sujinem, 2023). Against this backdrop, this study addresses three research questions: (1) How are HOTS and LOTS distributed across six language skills, GBA stages, and units in English for Change Grade XI? (2) Which Bloom's Taxonomy level is most dominant and least represented? (3) How do existing evaluation instruments compare against the demands of a responsive 21st-century and Merdeka Curriculum-aligned assessment?

METHODS

This study employs a mixed methods design with a sequential explanatory approach, in which quantitative data are collected and analyzed first, followed by a qualitative phase that deepens and contextualizes the quantitative findings (Creswell & Creswell, 2018; Ivankova et al., 2006; Samsu, 2017). The quantitative phase serves as the primary strand that produces frequency distribution data, while the qualitative phase provides an interpretive account of the numerical findings. The data consist of all evaluation instruments embedded in Bahasa Inggris: English for Change for Senior High School Grade XI, published by Indonesia's Ministry of Education, Culture, Research, and Technology in 2022 (ISBN 978-602-427-944-8). These instruments encompass all task instructions and questions across six language skills, namely Listening, Speaking, Reading, Viewing, Writing, and Presenting, drawn from all five units, as well as What Have You Learned So Far?, Mid-Lesson Reflection,

and End-of-Unit Reflection tasks. The use of a single nationally standardized source ensures data consistency and replicability for future researchers (Febraningrum & Suroso, 2023; Krippendorff, 2019; Pratycia & Zahro, 2025).

Data were collected through document study using content analysis as the primary method, enabling systematic identification, classification, and enumeration of textual characteristics (Ivankova et al., 2006; Krippendorff, 2019; Samsu, 2017). Six sequential and interdependent stages were followed. In the first stage, all activities across Units 1 to 5 were inventoried, and any item containing a task instruction or question was recorded as a unit of analysis, yielding 211 evaluation instruments. In the second stage, a coding sheet was constructed with columns for instrument number, unit, language skill, GBA stage, task instruction excerpt, identified operational verb, and cognitive level code (C1–C6), referencing Anderson and Krathwohl (2001) and Kemendikbudristek (2022). In the third stage, the researcher independently coded all 211 instruments. In the fourth stage, an independent learning assessment expert coded the same instruments without knowledge of the first rater's results. In the fifth stage, inter-rater reliability was calculated using Cohen's Kappa ($\varkappa \geq 0.70$ as the minimum threshold), and disputed items were resolved through consensus discussion. In the sixth stage, each instrument was qualitatively annotated for unit theme, text type, expected response mode, and task structure, generating the interpretive material for the qualitative analysis phase.

Data analysis proceeded in two sequential phases. In the quantitative phase, coded data were analyzed descriptively to generate frequency distributions and percentages for each Revised Bloom's Taxonomy level (C1–C6), then grouped into LOTS (C1+C2+C3) and HOTS (C4+C5+C6) (Anderson et al., 2001; Husni & Ginting, 2023; Wibowo et al., 2024). Distribution was calculated across four dimensions simultaneously: overall across all 211 instruments, per unit, per language skill, and per GBA stage. In the qualitative phase, quantitative findings provided the basis for descriptive-interpretive content analysis (Ivankova et al., 2006; Krippendorff, 2019; Samsu, 2017). This phase involved identifying dominant operational verb patterns, analyzing alignment between instrument cognitive demands and Merdeka Curriculum learning outcomes and CEFR B1 targets, identifying gaps between observed HOTS-LOTS distribution and the distribution required by 21st-century learning, and formulating development recommendations (Kemendikbudristek, 2022; Pratycia & Zahro, 2025; Yolanda & Maisarah, 2026). Integration of the two phases was achieved through a connecting procedure, in which quantitative results guided the qualitative

focus, ultimately producing a mutually reinforcing account of the textbook's cognitive evaluation profile.

RESULTS

HOTS and LOTS Distribution Across Language Skills, GBA Stages, and Units

Analysis of 211 evaluation instruments across all five units and six language skills established that HOTS-categorized instruments numbered 124 (58.8%), while LOTS-categorized instruments numbered 87 (41.2%). The textbook is therefore overall HOTS-dominant, though the 41.2% LOTS proportion indicates considerable scope for further development. Per-unit distribution shows a broadly progressive pattern: Unit 1 (Digital Literacies and My Identities) records the lowest HOTS proportion at 51.1% (23 of 45); Unit 2 (Love Your Environment) at 51.4% (19 of 37); Unit 3 (Healthy Life for a Healthy Future) rises to 61.0% (25 of 41); Unit 4 (Indonesian Environmental Figures) reaches the highest proportion at 70.7% (29 of 41); and Unit 5 (Personal Money Management) registers 59.6% (28 of 47). Notably, all five units surpass the 50% HOTS threshold, suggesting a curriculum-level orientation toward higher-order cognitive demands. This unit-level distribution, however, masks substantial variation at the skill and GBA-stage levels, as subsequent analyses reveal. Table 1 presents the complete unit-level distribution.

Table 1. HOTS-LOTS Distribution per Unit

Unit	Theme	Total	HOTS	%	LOTS	%
1	Digital Literacies and My Identities	45	23	51.1%	22	48.9%
2	Love Your Environment	37	19	51.4%	18	48.6%
3	Healthy Life for a Healthy Future	41	25	61.0%	16	39.0%
4	Indonesian Environmental Figures	41	29	70.7%	12	29.3%
5	Personal Money Management	47	28	59.6%	19	40.4%
Total		211	124	58.8%	87	41.2%

Per-skill distribution reveals considerably greater variation than the unit-level data. Writing leads with the highest HOTS proportion at 77.1% (27 of 35 instruments), followed by Presenting at 76.7% (23 of 30), Speaking at 58.8% (20 of 34), Reading at 52.6% (20 of 38), and Viewing at 51.5% (17 of 33). Listening is the only skill dominated by LOTS at 52.8% (19 of 36), a finding consistent with prior observations that receptive, input-dependent skills tend to be assigned fewer higher-order cognitive demands. The most striking finding at this

level is that all five End-of-Unit Reflection instruments (100%) are categorized as LOTS, suggesting a systematic misalignment between the reflective function these tasks are designed to serve and the cognitive demands they actually impose on students. The 25.6-percentage-point gap between Writing HOTS (77.1%) and Listening HOTS (47.2%) within the same textbook underscores the depth of this cross-skill asymmetry. Table 2 presents the full per-skill distribution.

Table 2. HOTS-LOTS Distribution per Language Skill

Language Skill	Total	HOTS	%	LOTS	%	Dominant
Listening	36	17	47.2%	19	52.8%	LOTS
Speaking	34	20	58.8%	14	41.2%	HOTS
Reading	38	20	52.6%	18	47.4%	HOTS
Viewing	33	17	51.5%	16	48.5%	HOTS
Writing	35	27	77.1%	8	22.9%	HOTS
Presenting	30	23	76.7%	7	23.3%	HOTS
End-of-Unit Reflection	5	0	0.0%	5	100.0%	LOTS

Distribution across GBA stages yields the study's most significant finding regarding the textbook's internal cognitive architecture. At the Building Knowledge of the Field (BKoF) stage, LOTS dominates at 73.5% (25 of 34 instruments), reflecting the stage's function in activating prior knowledge and introducing contextual content. At Modeling of Text (MoT), LOTS maintains a slight lead at 53.6% (37 of 69). The balance shifts substantially at Joint Construction of Text (JCoT), where HOTS reaches 70.7%, and continues to rise through Independent Construction of Text (ICoT) at 83.9% and Linking Related Texts (LRT) at 90.3%. This linear progression confirms that cognitive demand in English for Change escalates systematically across pedagogical stages, consistent with GBA design principles. However, End-of-Unit Reflection (EoU) reverts entirely to LOTS (100%; 5 of 5), representing a notable and pedagogically significant discontinuity at the close of each unit. Table 3 presents the full GBA-stage distribution.

Table 3. HOTS-LOTS Distribution per GBA Stage

GBA Stage	Code	Total	HOTS	%	LOTS	%
Building Knowledge of the Field	BKoF	34	9	26.5%	25	73.5%
Modeling of Text	MoT	69	32	46.4%	37	53.6%
Joint Construction of Text	JcoT	41	29	70.7%	12	29.3%
Independent Construction of Text	IcoT	31	26	83.9%	5	16.1%
Linking Related Texts	LRT	31	28	90.3%	3	9.7%
End-of-Unit Reflection	EoU	5	0	0.0%	5	100.0%

The Most Dominant and Least Represented Bloom's Taxonomy Levels

Analysis of the six Revised Bloom's Taxonomy levels reveals a distribution pattern that is markedly uneven and substantively different from findings in prior textbook studies. Rather than clustering at lower levels as commonly observed in comparable analyses, the instruments in English for Change are concentrated at the top and middle of the taxonomy, with relatively little representation at the foundational comprehension level. Creating (C6) accounts for 54 instruments (25.6%), making it the most frequently demanded level in the textbook. Evaluating (C5) and Applying (C3) are jointly second at 41 instruments each (19.4%), followed by Remembering (C1) at 32 instruments (15.2%) and Analyzing (C4) at 29 instruments (13.7%). Understanding (C2) is the least represented level with only 14 instruments (6.6%), a finding that raises questions about the degree to which the textbook scaffolds basic comprehension prior to demanding higher-order performance. The resulting rank order, C6 > C5 = C3 > C1 > C4 > C2, constitutes what this study terms an inverted Bloom's pyramid, in which the highest taxonomic level is also the most frequently demanded. Table 4 presents the complete distribution across all six levels.

Table 4. Distribution per Revised Bloom's Taxonomy Level

No.	Level	Code	n	%	Category
1	Remembering	C1	32	15.2%	LOTS
2	Understanding	C2	14	6.6%	LOTS
3	Applying	C3	41	19.4%	LOTS
4	Analyzing	C4	29	13.7%	HOTS
5	Evaluating	C5	41	19.4%	HOTS
6	Creating	C6	54	25.6%	HOTS

Creating (C6) is the most dominant level, with 54 instruments (25.6%), indicating that more than a quarter of all instruments require students to produce new language products such as written texts, oral performances, presentations, or posters. C6 is most concentrated in Writing (18 instruments, 51.4%) and Presenting (11 instruments, 36.7%), both of which inherently demand independent production. Across GBA stages, C6 peaks at ICoT and LRT at 13 instruments each (41.9%), confirming that independent production stages are consistently structured to elicit the highest cognitive demand. Evaluating (C5) and Applying (C3) are jointly second at 41 instruments each (19.4%). C5 is strongest at LRT (14 instruments, 45.2%) and ICoT (11 instruments, 35.5%), while C3 predominates in Listening (13 instruments, 36.1%) and BKoF (14 instruments, 41.2%), reflecting the guided-practice

character of early learning stages. Understanding (C2) is the least represented level with only 14 instruments (6.6%), a distribution that raises questions about the extent to which foundational comprehension is scaffolded before higher-order performance is required. The overall rank order, C6 (25.6%) > C5 (19.4%) = C3 (19.4%) > C1 (15.2%) > C4 (13.7%) > C2 (6.6%), constitutes an inverted Bloom's pyramid in which the highest level is also the most frequently demanded.

Characteristics of Existing Instruments Against 21st-Century Assessment Needs

Qualitative analysis identified five key characteristics that define the cognitive profile of English for Change and assessed each in relation to 21st-century assessment demands. First, productive-creative dominance characterizes the two output-oriented skills: Writing and Presenting are overwhelmingly C6-led (51.4% and 36.7% respectively), with dominant operational verbs including create, design, write, produce, and develop. Representative C6 instruments include "Think of a job that you dream of. Create a social media caption about it using subject questions and transitional words" (Unit 1, LRT) and "Work individually. Create a digital or non-digital poster of a healthy lifestyle" (Unit 3, LRT). This pattern aligns well with the Merdeka Curriculum's production-based learning goals. Second, Listening tasks remain oriented toward guided application: Listening is dominated by C3 Applying (36.1%), with operational verbs such as fill in, complete, use the expression, write down, and identify prevailing. While C5 (25.0%) and C6 (13.9%) do appear in Listening, most tasks direct students to respond to surface information rather than evaluate source credibility, detect bias, or synthesize multiple audio perspectives, which represent the critical listening competencies most essential in contemporary digital information environments.

Third, the progressive GBA cognitive scaffolding pattern is empirically confirmed: HOTS increases linearly across every stage from BKoF (26.5%) through MoT (46.4%), JCoT (70.7%), ICoT (83.9%), and LRT (90.3%), with no stage showing a decline. At LRT, C5 and C6 together account for 87.1% of instruments, indicating that the text-linking stage is structured to elicit critical evaluation and creative synthesis. Fourth, End-of-Unit Reflection constitutes a systematic gap: all five reflection instruments are LOTS-categorized, four at C1 and one at C3. The format is uniform across all units, prompting students to recall content they have learned, identify what they most understand, note what they want to learn more about, and offer their general opinion of the unit. This design requires recall rather than the evaluation of relevance or synthesis of new understanding that instruments at C5 or C6

would require. Fifth, C4 Analyzing is underrepresented among HOTS levels, appearing in only 3 Listening instruments (8.3%) and 2 Speaking instruments (5.9%), leaving tasks such as comparing arguments or identifying causal relationships virtually absent in the two interactive skills. These five characteristics collectively confirm a productively asymmetric cognitive profile, with critical listening, analytical viewing, and evaluative reflection as the three dimensions most urgently requiring systematic development.

DISCUSSION

Three major findings collectively yield a comprehensive picture of the cognitive profile of English for Change. First, HOTS dominates overall at 58.8%, a finding of considerable significance given that prior studies on comparable textbooks consistently documented LOTS dominance. Second, the six Bloom's levels follow an inverted pyramid in which C6 Creating is the most frequent (25.6%) while C2 Understanding is the least (6.6%), a distributional pattern not previously documented in Kemendikbud English textbooks. Third, a systematic asymmetry exists: Writing (77.1% HOTS) and Presenting (76.7%) far exceed Listening, which is LOTS-dominant (52.8%), while all End-of-Unit Reflection instruments are entirely LOTS (100%). Together, these findings demonstrate that HOTS orientation in English for Change is concentrated in productive skills and later GBA stages rather than evenly distributed across the textbook. The claim that this textbook is fully HOTS-oriented therefore requires methodological qualification: it is structurally HOTS-oriented through its GBA design, but not evenly HOTS-distributed across language skills and reflective components (Anderson & Krathwohl, 2001; Kemendikbudristek, 2022; Pratycia & Zahro, 2025).

The overall HOTS dominance (58.8%) aligns with prior work on Merdeka Curriculum textbooks, particularly Work in Progress Grade X. Studies of (Husni & Ginting, 2023; Manalu & Harahap, 2025; Ripalga & Fitrawati, 2023) consistently found HOTS-dominant reading tasks in Work in Progress, with C4 Analyzing most frequent, indicating that senior high school Merdeka Curriculum textbook writers have broadly incorporated higher-order cognitive demands. The prominence of C6 Creating in the present study also resonates with (Anwar et al., 2024; Rohimajaya et al., 2022; Sujinem, 2023), who found that creativity and independent production, core 21st-century competencies, have been integrated into Merdeka Curriculum English textbooks, albeit unevenly. These convergent findings lend

support to the argument that the cognitive shift from Curriculum 2013 to the Merdeka Curriculum is real and measurable, and that English for Change embodies that curricular trajectory. The evidence further suggests that HOTS orientation in Merdeka Curriculum textbooks is the product of a broader, curriculum-consistent publishing policy rather than an isolated textbook-level decision (Barlian et al., 2022; Kemendikbudristek, 2022; Pratiwi, 2025).

The skill-level imbalance, particularly LOTS dominance in Listening, echoes patterns observed in language-skills proportion studies of Kemendikbud textbooks. (Paramitha et al., 2023; Rohimajaya et al., 2022) consistently found that Listening had the lowest proportion of instruments and the lowest instrument quality, with tasks oriented toward mechanical reception, including table completion, surface-level question answering, and phrase identification. That this pattern persists in the latest Merdeka Curriculum textbook suggests the Listening imbalance is a cross-curriculum structural problem rather than an idiosyncratic feature of any individual textbook. The minimal critical listening prompts in English for Change similarly align with findings by (Pratycia & Zahro, 2025), who found that listening and speaking tasks in Merdeka Curriculum textbooks do not consistently provide authentic, interactive communicative contexts for evaluative development. The 100% LOTS classification of End-of-Unit Reflection instruments further corroborates (Ardiyansari & Jemadi, 2024; Wibowo et al., 2024; Yolanda & Maisarah, 2026), who found that reflective and summative instruments in Indonesian textbooks remain anchored to recall and comprehension rather than synthesis and critical evaluation.

The most fundamental divergence from prior studies lies in the distribution of Bloom's levels. All available HOTS-LOTS studies on Kemendikbud English textbooks, whether on English for Nusantara (Ramadhani, 2023; Yolanda & Maisarah, 2026) or Curriculum 2013 senior high school books (Fitriani & Kirana, 2021), consistently report C1 or C2 as the most dominant level with C6 the least frequent or absent. Ekalia et al. (2025) found C1 at 56.0% in English for Nusantara Grade VIII with C6 at only 2.67%, while (Wibowo et al., 2024) found no C6 instances at all in a Kemendikbud Indonesian language textbook. This stands in marked contrast to the present study's findings, in which C6 Creating is the most dominant (25.6%) while C2 Understanding is the least (6.6%). This divergence reflects fundamentally distinct pedagogical architectures: the Genre-Based Approach in English for Change, particularly through ICoT and LRT stages that explicitly demand independent production, structurally generates far more C6 instruments than

receptive-cognitive instructional models do (Anderson & Krathwohl, 2001; Ariawan et al., 2023; Husni & Ginting, 2023).

A second methodological divergence is that this study simultaneously analyzes HOTS-LOTS distribution across all six language skills and all GBA stages, an analytical scope not achieved by any prior study. Earlier work, including (Husni & Ginting, 2023; Ramadhani, 2023; Yolanda & Maisarah, 2026), confined analysis to Reading, while (Ardiyansari & Jemadi, 2024) examined Writing alone. This scope limitation produced partial pictures that could not capture the real cognitive dynamics of a textbook as a whole. By analyzing all six skills simultaneously, this study reveals a 29.9-percentage-point gap between Writing HOTS (77.1%) and Listening HOTS (47.2%) within the same book, a disparity that would be entirely obscured by single-skill analysis. The linear, progressive increase in HOTS across GBA stages, from 26.5% at BKoF to 90.3% at LRT, is likewise a genuinely novel empirical contribution, as no prior study employed GBA stages as the unit of analysis. This finding extends and deepens the work of (Handayani & Amelia, 2023; Indar et al., 2023; Pratycia & Zahro, 2025), who evaluated Merdeka Curriculum textbooks from a general feasibility angle without examining internal cognitive distribution across pedagogical stages.

Theoretically, this study makes three contributions. First, it extends and validates the Revised Bloom's Taxonomy (Anderson et al., 2001) as a comprehensive analytical framework applicable across all six language skills simultaneously, complementing a body of literature that has applied it only partially (Husni & Ginting, 2023; Wibowo et al., 2024; Yolanda & Maisarah, 2026). Second, it introduces and empirically validates the concept of GBA-stage-based cognitive distribution as a new unit of analysis in English textbook studies. The finding that HOTS proportions increase consistently and linearly from BKoF to LRT provides the first empirical evidence that GBA design in textbooks carries a measurable, verifiable cognitive structure, a claim that had previously been posited on theoretical-normative rather than empirical grounds (Handayani & Amelia, 2023; Indar et al., 2023; Pratycia & Zahro, 2025). Third, this study contributes a replicable evaluation instrument needs analysis framework that integrates quantitative distribution mapping with qualitative characteristic analysis to produce an actionable gap map, a model transferable to textbook analyses across other subjects and grade levels (Creswell & Creswell, 2018; Krippendorff, 2019; Samsu, 2017).

In practice, the findings carry direct implications for four stakeholders. For textbook developers, the distribution map delineates precisely where the gaps reside: C4- and C5-based critical listening instruments require significant addition; End-of-Unit Reflection tasks need redesigning toward at least C5, for instance through a 4Rs framework (Return, Relate, Reason, Reconstruct); and C4 Analyzing must be proportionally strengthened in Listening and Speaking (Anwar et al., 2024; Ariawan et al., 2023; Kemendikbudristek, 2022). For senior high school English teachers, the cognitive map provides a practical basis for developing supplementary evaluation instruments that address the textbook's existing imbalances. For curriculum and assessment policymakers, the data constitute empirical evidence that the Merdeka Curriculum's HOTS demands have been addressed unevenly across language skills within the same textbook, an asymmetry that demands systematic attention in national assessment policy and skill-specific question development guidelines (Barlian et al., 2022; Pratiwi, 2025; Pratycia & Zahro, 2025). For future researchers, the cross-skill GBA-stage methodology developed here can be replicated across other Merdeka Curriculum English textbooks to build a comprehensive comparative cognitive map (Creswell & Creswell, 2018; Krippendorff, 2019; Samsu, 2017).

CONCLUSION

This study empirically demonstrates that the evaluation instruments in English for Change Grade XI are overall HOTS-dominant, with 124 of 211 analyzed instruments (58.8%) classified at the higher-order level, and that this dominance follows a productively asymmetric distribution. With respect to the first research question, per-skill analysis reveals a significant imbalance: Writing (77.1%) and Presenting (76.7%) lead in HOTS proportion, while Listening is the sole LOTS-dominant skill (52.8%) and all End-of-Unit Reflection instruments are entirely LOTS (100%). The GBA-stage distribution confirms linear, progressive cognitive scaffolding from BKoF (26.5%) to LRT (90.3%), validating the textbook's pedagogical design. Concerning the second research question, Creating (C6) is the most dominant level (25.6%) while Understanding (C2) is the least (6.6%), constituting an inverted Bloom's pyramid not previously documented in any Kemendikbud English textbook. As regards the third research question, three principal gaps are identified: minimal C4- and C5-based critical listening instruments, low proportions of C4 analysis in receptive-interactive skills, and the complete absence of evaluative reflection in End-of-Unit tasks.

Overall, English for Change has responded structurally to the Merdeka Curriculum's HOTS mandate through its GBA design but necessitates systematic strengthening in critical listening, analytical viewing, and evaluative reflection to achieve genuinely balanced 21st-century-responsive evaluation across all learning dimensions.

Several limitations of this study merit consideration. First, this study draws data from a single textbook, so findings cannot be directly generalized to other Merdeka Curriculum English textbooks. Second, cognitive-level classification was based solely on operational verbs in task instructions, without fully accounting for the cognitive complexity that may be demanded by accompanying texts, a dimension that could yield different classifications if incorporated. Third, no primary data were collected from teachers or students who used the textbook directly, leaving the actual impact of the observed HOTS-LOTS distribution on classroom learning outcomes unmeasured. These limitations simultaneously illuminate three productive avenues for future research: comparative studies across all Merdeka Curriculum English textbooks to produce a comprehensive cross-level cognitive map; research integrating teacher and student perspectives through surveys or interviews to validate document analysis findings against classroom realities; and developmental research specifically designing, trialing, and evaluating critical listening and evaluative reflection instruments grounded in this study's findings. Taken together, these directions represent a meaningful pathway toward more innovative, equitable, and impactful English language assessment in Indonesia.

REFERENCES

AJR, A. E. N., Al-Baekani, A. K., & Hakim, P. K. (2024). An analysis of writing task in English textbook for seventh grade of junior high school based on cognitive domain of revised Bloom's taxonomy. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 12*(1), 258–277. https://doi.org/10.24256/ideas.v12i1.4123

Alfaya, A. F., Rochsantiningsih, D., & Sumardi. (2023). Cultural representation in the ELT textbook "Work in Progress." *DIKSI, 31*(2), 233–243. https://doi.org/10.21831/diksi.v31i2.65026

Ameliana, A., & Zahro, S. K. (2025). Assessing the alignment of "English for Change" textbook within the Merdeka Curriculum framework. *JOLLT Journal of Languages and Language Teaching, 13*(4), 2091–2102. https://doi.org/10.33394/jollt.v13i4.15433

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.

Anwar, R., Pujiawati, N., & Mobit. (2024). The 21st-century skills in Indonesia senior high school English textbook. *Jurnal Ilmiah Wahana Pendidikan, 10*(3), 755–767. https://doi.org/10.5281/zenodo.10645934

Ariawan, S., Kholidi, M. A., & Putra, M. (2023). The level of thinking skills in the reading section of EFL textbook in Indonesia. *VELES: Voices of English Language Education Society, 7*(1), 117–125. https://doi.org/10.29408/veles.v7i1.7672

Arlansyah, A., Puspita, H., & Saputra, E. (2023). Reading questions in "English for Nusantara" textbook by using revised Bloom's taxonomy. *Journal of English Education and Teaching, 7*(2), 361–375. https://doi.org/10.33369/jeet.7.2.361-375

Badan Standar, Kurikulum, dan Asesmen Pendidikan. (2022). *Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Menengah*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. https://kurikulum.kemdikbud.go.id/wp-content/uploads/2022/06/Panduan-Pembelajarn-dan-Asesmen.pdf

Bahri, K., & Agustina, N. (2023). Cultural content and intercultural communicative competence in Indonesian EFL textbooks. *Tell-Us Journal, 9*(3), 746–759. https://doi.org/10.22202/tus.2023.v9i3.7321

Barlian, U. C., Solekah, S., & Rahayu, P. (2022). Implementasi Kurikulum Merdeka dalam Meningkatkan Mutu Pendidikan. *JOEL: Journal of Educational and Language Research, 1*(12), 2105–2118. https://doi.org/10.53625/joel.v1i12.3015

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Deswila, N., Kustati, M., Yusuf, Y. Q., Raja Harun, R. N. S., & Rusman. (2021). Cultural contents in the ELT textbook prepared by the Indonesian Ministry of Education for secondary schools. *Journal of Nusantara Studies, 6*(1), 222–241. https://doi.org/10.24200/jonus.vol6iss1pp222-241

Febraningrum, M. I. N., & Suroso, B. (2023). A content analysis of English textbook "English for Nusantara" for Grade 7 junior high school in Kurikulum Merdeka. *Proceedings Series on Social Sciences & Humanities, 13*, 145–152. https://doi.org/10.30595/pssh.v13i.897

Fitriani, D. H., & Kirana, D. P. (2021). Higher order thinking skill (HOTS) in English language textbook in senior high school. *IJRETAL: International Journal of Research on English Teaching and Applied Linguistics, 2*(2), 55–66. https://doi.org/10.30863/ijretal.v2i2.2455

Hakim, A. N. M., & Wahyuni, D. S. (2024). Cultural content in English textbook: Analyzing the representation of culture in the English textbook "English for Change." *Journal of English Language Teaching and Learning (JETLE), 6*(1), 13–22. https://doi.org/10.18860/jetle.v6i1.27768

Handayani, N., & Amelia, R. (2023). Evaluating "English for Nusantara" for Grade 8: A qualitative content analysis based on Tomlinson's framework. *Journal of Foreign Language Learning and Teaching, 4*(2).

Husni, S. S., & Ginting, S. A. (2023). Higher order thinking skills of reading exercises in English textbook "Bahasa Inggris: Work in Progress" based on revised Bloom's taxonomy. *GENRE: Journal of Applied Linguistics of FBS Unimed, 12*(3), 106–114. https://doi.org/10.24114/gj.v12i3.53918

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods, 18*(1), 3–20. https://doi.org/10.1177/1525822X05282260

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications.

Lestari, D. D., Surono, & Ika, S. (2023). Cultural contents of senior high school English textbook entitled "Bahasa Inggris." *Teaching English as a Foreign Language Journal, 1*(2), 103–113. https://doi.org/10.12928/tefl.v1i2.255

Manalu, S., & Harahap, N. A. S. (2025). Higher order thinking skills in reading questions of English textbook for 10th grade high school students. *Invention: Journal Research and Education Studies, 6*(1), 139–154. https://doi.org/10.51178/invention.v6i1.2422

Marzuki, I., Sholihah, T., & Imansyah, F. A. (2023). Urgensi Aspek Penilaian dalam Evaluasi Pembelajaran. *Tadarus Tarbawy: Jurnal Kajian Islam dan Pendidikan, 5*(1), 1–6. https://doi.org/10.31000/jkip.v5i1.8634

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Palupi, T. M., & Subianto, K. A. (2024). The incorporation of digital literacy in EFL learning materials on junior high school. *Indonesian EFL Journal, 10*(1), 87–96. https://doi.org/10.25134/ieflj.v10i1.9293

Pangestuti, A. F. N., Ana, S., Syahra, E. F. A., & Sari, S. N. (2024). Analisis Distribusi Soal Higher Order Thinking Skills (HOTS) dan Lower Order Thinking Skills (LOTS) dalam Buku Ajar Bahasa Indonesia. *Indonesian Journal of Action Research, 3*(1), 25–31. https://doi.org/10.14421/ijar.2024.31-03

Paramitha, P. A., Sari, S. S. N., Syafitri, W., & Sundari, H. (2023). An analysis of gender visibility in the EFL textbook for Grade 7 in Indonesia. *Prosiding Konferensi Berbahasa Indonesia Universitas Indraprasta PGRI*, 211–216. https://doi.org/10.30998/kibar.27-10-2022.6315

Permatasari, C. Y., & Arsyad, S. (2023). Coherence analysis of reading texts in English textbooks for class X SMA "Work in Progress" published by Ministry of Education and Culture of the Republic of Indonesia. *Journal of English for Specific Purposes in Indonesia, 2*(1), 59–66. https://doi.org/10.33369/espindonesia.v1i2.26130

Pratiwi, W. R. (2025). Transforming education with the Merdeka Belajar curriculum: Digging into challenges and crafting effective strategies in rural Indonesia. *ENGLISH FRANCA: Academic Journal of English Language and Education, 9*(1), 111–128. https://doi.org/10.29240/ef.v9i1May.12737

Ravyansah, & Abdillah, F. (2021). Tracing "Profil Pelajar Pancasila" within the civic education textbook: Mapping values for adequacy. *Jurnal Moral Kemasyarakatan, 6*(2), 96–105. https://doi.org/10.21067/jmk.v6i2.5906

Ripalga, & Fitrawati. (2023). An analysis of questions on reading tasks in the English textbook "Work in Progress" in Merdeka Curriculum. *Journal of English Language Teaching, 12*(2), 570–581. https://doi.org/10.24036/jelt.v12i2.123128

Rohimajaya, N. A., Hartono, R., Yuliasri, I., & Fitriati, S. W. (2022). Kurikulum 2013 dan Kurikulum Merdeka dalam Pembelajaran Bahasa Inggris untuk SMA di Era Digital. *Prosiding Seminar Nasional Pascasarjana, 5*(1), 825–829.

Samsu. (2017). *Metode Penelitian: Teori dan Aplikasi Penelitian Kualitatif, Kuantitatif, Mixed Methods, serta Research & Development*. Pusaka Jambi.

Silviana, Z. J., & Triastuti, A. (2025). The representation of multicultural values in Indonesian ELT textbook "English for Nusantara." *Formosa Journal of Sustainable Research, 4*(3), 477–494. https://doi.org/10.55927/fjsr.v4i3.104

Sujinem. (2023). Analysis of implementation of communicative approach in teaching English under "Kurikulum Merdeka." *Revorma: Jurnal Pendidikan dan Pemikiran, 3*(1), 56–68. https://doi.org/10.62825/revorma.v3i1.46

Tobing, H. E. L., Somakim, & Susanti, E. (2022). Development of e-module based on HOTS questions on distance material for high school students. *Jurnal Pendidikan Matematika, 16*(1), 1–14. https://doi.org/10.22342/jpm.16.1.14694.1-14

Wardarita, R., Batubara, H. H., Wahyuni, N., & Astina, T. (2024). Urgensi Evaluasi Pembelajaran dalam Menciptakan Pendidikan Berkualitas di Era Merdeka Belajar. *Innovative: Journal of Social Science Research, 4*(1), 1583–1590. https://doi.org/10.31004/innovative.v4i1.8023

Yolanda, C., & Maisarah, I. (2026). An analysis of cognitive dimensions in reading assessment tasks of the English for Nusantara grade 8 textbook. *Indonesian Journal of English Language Studies (IJELS), 12*(1), 60–68. https://doi.org/10.24071/ijels.v12i1.213