

EVALUATING THE GRADE 10 EFL COURSEBOOK UNDER KURIKULUM MERDEKA: A MULTIDIMENSIONAL ANALYSIS

Heldy Hasfid Hariri^{1)*}, Nazril Azmi Huzein²⁾, Dwi Wara Wahyuningrum³⁾,
Dinar Dipta⁴⁾

^{1), 2), 3), 4)}, English Language Teaching, Tarbiyah Faculty,
Universitas Darussalam Gontor, Indonesia

Heldyhasfidhariri12@student.tbi.unida.gontor.ac.id^{1)*},
nazrilazmihuzein76@student.tbi.unida.gontor.ac.id²⁾, dwiwahyuningrum@unida.gontor.ac.id³⁾
dinardipta@unida.gontor.ac.id⁴⁾

Abstract

This study evaluates the Grade 10 English coursebook *Bahasa Inggris untuk SMA/MA Kelas X* (Kemdikbudristek, 2021) within Indonesia's Kurikulum Merdeka framework. A qualitative evaluative design guided the analysis, with document analysis conducted using a validated instrument adapted from Cunningsworth (1995), McGrath (2002), Tomlinson (2011), and Miekley (2005) across five dimensions: quality, variety, effectiveness, alignment, and appropriateness. The coursebook scored 104 out of 125 points (83.2%), placing it in the Good category. Alignment earned the highest rating (88.0%), supported by coherent skill integration and communicative language teaching task design. Three systematic weaknesses emerged: uneven HOTS task distribution across units, repetitive inter-unit activity formats, and insufficient inclusivity scaffolding for Indonesia's sociocultural diversity. The coursebook is suitable as a primary instructional resource, provided teachers supplement it in three targeted areas: HOTS task design, format variation, and differentiated instruction.

Keywords: *Coursebook evaluation, Kurikulum Merdeka, Indonesian EFL, higher-order thinking, communicative language teaching.*

©Pendidikan Bahasa Inggris FKIP UM Palembang

DOI: <https://doi.org/10.32502/ecj.v10i2.11553>

Introduction

Textbooks are the primary vehicle for language input, task design, and syllabus organization in most EFL classrooms (Cunningsworth, 1995; Harmer, 2007; Tomlinson, 2011). In Indonesia, this dependence runs deep: the Ministry of Education has issued standardized English textbooks since the 1970s, and their pedagogical quality shapes the learning outcomes of millions of students nationwide (Musthafa, 2010). Kurikulum Merdeka (2021–present) extended this pattern. The new curriculum replaced K-13 by reorienting instruction around Learning Outcomes (Capaian Pembelajaran/CP), student agency, and the Pancasila Student Profile Strengthening Project (P5). The Ministry published a new Grade 10 coursebook, *Bahasa Inggris untuk SMA/MA Kelas X* (Kemdikbudristek, 2021), now the primary instructional resource for English at Phase E level

across Indonesian public secondary schools.

Systematic ELT textbook evaluation draws on four major frameworks. Cunningsworth (1995) assesses learner needs, language content, skills, methodology, and teacher support. In 2002 McGrath (2002) distinguishes impressionistic, checklist-based, and in-depth evaluation. Tomlinson (2011) grounds evaluation in second language acquisition theory. Miekley (2005) provides a quantifiable rubric adaptable to diverse EFL contexts. International studies applying these frameworks reveal a recurring finding: coursebooks score acceptably on general pedagogical criteria but underperform on higher-order thinking task design, authentic communicative practice, and sociocultural representation.

Khodabakhshi (2014) documented this for the Skyline series in Iran. Chan and Cheuk (2020) used corpus analysis

to show that pedagogical ideals built into textbook design can diverge substantially from the tasks students complete. Lavrenteva and Orland-Barak (2023) showed that multimodal analysis at both micro and macro levels is needed to trace how cultural representation operates across units. Demir and Töngür (2026) applied corpus methods to assess lexical pedagogical value, adding a methodological tool now available to textbook evaluators.

Indonesian textbook evaluation has documented recurring weaknesses across successive curriculum reforms. Azizah (2020), Rahimah (2021), and Yuliani and Aziza (2022) identified gaps in pronunciation pedagogy, critical thinking task design, and learner autonomy scaffolding across K-13 English textbooks. Herliana, Djajanegara, and Suendarti (2020) found that teachers and students flagged vocabulary loading, audio support, and task sequencing as problems even in textbooks that met basic ELT standards. Whether these gaps persist under Kurikulum Merdeka is an open question. Berlin and Wirza (2026) evaluated the Grade X government-issued EFL textbook and found that the curriculum's learner-agency goals are not consistently realized in textbook tasks, reinforcing the case for evaluation calibrated to Kurikulum Merdeka's competency targets. Phase E Learning Outcomes set communicative competence targets in five skill areas at CEFR A2–B1 level (Kemdikbudristek, 2021), and any evaluation of this coursebook must measure performance against those targets.

Classroom conditions shape what teaching materials can deliver. Indonesian public secondary schools present concrete constraints: class sizes of 30 - 40 students, uneven technology access in rural areas, and significant teacher competency variation (Nation &

Macalister, 2010a; Tomlinson, 2011). A textbook that scores well on pedagogical criteria but cannot function within these constraints will not serve Indonesia's student population.

No published study has evaluated Bahasa Inggris untuk SMA/MA Kelas X as a whole. Existing Indonesian EFL evaluations cover earlier K-13 editions (Azizah, 2020; Herliana et al., 2020; Rahimah, 2021) or narrower constructs such as learner autonomy in the Kurikulum Merdeka textbook (Berlin & Wirza, 2026). No study has applied an integrated multi-framework instrument covering quality, variety, effectiveness, alignment, and contextual appropriateness in one analysis. Without this evidence, teachers have no principled basis for deciding when and how to supplement the coursebook they use every day (Hamied & Lengkanawati, 2018; Litz, 2005).

This study evaluates Bahasa Inggris untuk SMA/MA Kelas X using an instrument adapted from Cunningsworth (1995), McGrath (2002), Tomlinson (2011), and Miekley (2005) across five dimensions: quality, variety, effectiveness, alignment, and appropriateness. Combining four frameworks produces a more internally consistent assessment than any single framework allows. The study asks: How does Bahasa Inggris untuk SMA/MA Kelas X perform across these five dimensions, and what do the results mean for teachers working within Kurikulum Merdeka? The findings address teachers, school administrators, material developers, and Ministry of Education policymakers who need evidence-based grounds for evaluating and supplementing this coursebook.

Two delimitations apply. First, this study covers the Grade X student book and its companion Teacher's Book; digital materials on the Merdeka Mengajar platform provided contextual

background only. Second, a single evaluator conducted the analysis, which precludes inter-rater reliability. Future research should involve multiple raters.

Research Method

This study uses a qualitative evaluative design. Patton (2015) defines this as a scientific effort that integrates theory and practice to assess an object's quality here, an English coursebook rather than a purely descriptive exercise. Document analysis was chosen as the primary method because textbook content requires interpretive reading that numerical scoring alone cannot capture (Erlingsson & Brysiewicz, 2017).

The main data sources are the student book Bahasa Inggris untuk SMA/MA Kelas X and its accompanying Teacher's Book, both published by the Ministry of Education, Culture, Research, and Technology (2021). Each unit serves as the unit of analysis, examined against a validated instrument covering five dimensions: quality, variety, effectiveness, alignment, and appropriateness. Digital materials on the Merdeka Mengajar platform were referenced for contextual background only.

The coding scheme follows the category-based approach modeled by Egbert et al. (2016). Analysis operated at two levels: micro-evaluation of individual texts, tasks, and images, and macro-evaluation of cross-unit patterns, following Lavrenteva and Orland-Barak (2023). Each unit was coded for the presence, adequacy, and quality of features in the checklist, then recorded in a structured evaluation sheet. Thematic analysis organized the coded data through Braun and Clarke's six phases: familiarization, initial coding, theme grouping, and final naming before interpretation against Kurikulum Merdeka's competency targets.

To maintain credibility, the instrument was pilot-tested on one unit before full analysis, and coding decisions were reviewed periodically against the original checklist. Because a single evaluator conducted the analysis, inter-rater reliability could not be established as a limitation this study acknowledges openly. Future research should involve multiple raters to strengthen validity.

Results and Discussion

The evaluation of English for SMA/MA Class X (Maman et al., 2021) resulted in an aggregate score of 104 out of 125 points that could be achieved in 26 assessment items (80.0%), placing this book in the good category based on the Cunningsworth–McGrath adaptation rubric. Table 1 below breaks down the composite scores into five evaluative dimensions.

Table 1. Summary of Evaluation Result by Dimension

N o.	Evaluation Dimensions	Num ber of Items	Ra w Sco re	Max Scor e.	Percent age	Ratings / Catego ries
1	Quality	5	21	25	84.0%	Good
2	Variety	5	21	25	84.0%	Good
3	Effectiveness	5	20	25	80.0%	Good
4	Alignment	5	22	25	88.0%	Good – Very Good
5	Appropriateness	5	20	25	80.0%	Good
TOTAL		25	104	125	83.2%	Good

In this table, it can be seen that the evaluation carried out by the researcher reached an aggregate of 80.0%, placing this textbook in the performance range that was also reported by a number of evaluations of EFL books published by the Indonesian government previously. Herliana et al., (2020), who evaluated the previous edition of the same series, found that teachers and students expressed objections to vocabulary loading, audio support, and assignment

sequencing, issues that in this 2023 edition have been partially addressed, though not completely, as outlined in the following sections.

Quality of Activity Design

The Quality Dimension obtained a score of 21/25 (84.0%), reflecting a generally solid pedagogical construction with one substantive gap. Criteria 1.1 to 1.4, which included clarity of instruction, structural regularity, age appropriateness, and linguistic accuracy, were rated Good (4) respectively. These findings are in line with the observation of Hanifa (2018) that textbooks produced by the Ministry of Education and Culture generally meet the standards of the basic ELT methodology: easy-to-read layout, clear instructions, and maintained accuracy of content. The 2023 edition retains these features, and the evaluative evidence (p. 3, 15, 40 and p. 5, 22, 55) confirms that the task instructions are easy to understand as well as the units progress logically from receptive input to productive output.

Criterion 1.5, which is an activity that encourages deep learning rather than just memorization, is the only item in this dimension that is rated Sufficient (3). Comments on the evaluation sheet identified the presence of reflection tasks and high-level thinking items on a given page (pp. 18, 60), but the overall score signaled that the tasks were episodic, not systematic. Febriyani et al., (2020). analyzing the title of the parallel Ministry of Education and Culture (English Class XII), it was documented that the cognitive dimension that dominated all language skill tasks was remembering (C1) by 41%, while the analysis and evaluation tasks were only a minority of the task population. A similar tendency toward low-level cognitive work is the most plausible explanation for the Enough rating here.

This pattern has a well-documented equivalent in the broader

EFL literature. Mrah (2017) and Mu et al. (2026) who evaluated high school EFL textbooks in China, both concluded that the majority of reading comprehension tasks target the three lowest categories of Bloom's Taxonomy, so critical and creative thinking does not get adequate scaffolding. Tyas et al, (2020), in a case study in East Java, found that Indonesian English teachers had an adequate conceptual understanding of HOTS, but remained substantially dependent on textbook assignments that did not institutionalize them systematically. The 2023 edition of this Class X book reproduces a structural deficit endemic in the ecology of EFL Indonesian textbooks: the task of reflection is present, but its distribution across units is not enough to build cumulatively high-level thinking competencies.

Pedagogical Implications for the Independent Curriculum: The Pancasila Student Profile Framework explicitly makes critical, creative, and collaborative thinking as graduate competencies. Textbooks that selectively embed HOTS assignments create a misalignment between the stated curriculum philosophy and the embodied classroom experience. Teachers should treat things. 18 and 60 are not as the final objectives of the HOTS, but rather as a starting point for an additional sequence of tasks, such as replacing closed-ended comprehension questions with Socratic question protocols or comparative text analysis (Bağ & Gürsoy, 2021).

Variations of Activity Formats

With a score of 21/25 (84.0%), the Variation dimension emerges as one of the most obvious strengths of this textbook. Criteria 2.1 to 2.3 were obtained Very Good (5) or Good (4) respectively: individual, pair, and group configurations were well represented (pp. 10, 35, 70); task types include

matching, filling, discussion, and role-playing (pp. 14, 50); and the balance between the exercise of receptive and productive skills is considered "quite good" (pp. 20, 65).

Criterion 2.4, non-repetitive or monotonous activities, which obtained a rating of Sufficient (3), contains structural significance. The evaluator noted that "some units have a similar pattern." This finding resonates with the weaknesses identified, Hanifa, (2018) explains that in KTSP and K-13 English textbooks: the grammar of the basic task format, once recognized by students, reduces novelty and can decrease motivational engagement over time. Nation and Macalister (2010a) argue that repetition of tasks without cognitive escalation results in superficial processing; learners complete familiar task formats automatically instead of engaging with the target language in full. In textbooks structured across multiple units with repetitive genre cycles, a design feature of the text-based approach embedded in K-13 and passed on to the Independent Curriculum, the risk of formatting monotony is structurally increased.

Criterion 2.5, the activity provided different levels of challenge, obtained a score of Good (4), with evaluative comments confirming the presence of "gradual scaffolding." These findings are consistent with the Vygotskian scaffolding principle: tasks that begin with guided practice and gradually draw guidance allow learners to operate within the upper limit of their zone of proximal development (ZPD). This textbook partially realizes this design logic, but scaffolding architecture is not always explicitly presented to students, a problem that can be addressed in future editions by annotating the progress of the assignment using explicit guidance language.

The pedagogical implications for the Independent Curriculum in this section are that teachers should utilize a variety of documented interaction formats as a foundation for differentiated teaching. The unit-level flexibility provided by the Independent Curriculum opens up room for deliberate rotation of formats across lessons. Breaking down the similarities between the patterns between units identified in Criterion 2.4 can be achieved through teacher-generated enrichment tasks that introduce a problem-based or inquiry-based format Pardede (2020).

Effectiveness of Activities in Achieving Learning Outcomes

The Effectiveness Dimension obtained a score of 20/25 (80.0%), with a distributed rating between Excellent (5) and Adequate (3). Criteria 3.1, 3.2, and 3.5 scored Good (4) and the unit's objectives were explicit (pp. 2, 42, 78), application-oriented practices were "adequately available" (pp. 16, 54), and the review section in each unit supported retention (pp. 44, 95). These features are in line with Harmer's (2015) principle that textbooks should provide transparent learning trajectories and cyclical review mechanisms that consolidate declarative knowledge into procedural competencies.

Criterion 3.3, the activity of encouraging meaningful communication and use of language, achieved Very Good (5) (pp. 25, 68), the only peak ranking in this dimension. These results deserve special attention. Communicative task design is the cornerstone of the post-structural ELT methodology, and it is a dimension in which many Asian government-produced EFL textbooks have historically underperformed. The highest score on the provision of communicative assignments indicates that the 2023 revision internalizes the principles of communicative language teaching (CLT)

more thoroughly than its predecessor editions, a finding that echoes with the acknowledgment of Herliana et al., (2020) stating that even previous editions embedded "*opportunities that encourage students' creativity, interest, and curiosity.*"

Criterion 3.4, the activity of developing critical thinking and problem-solving, obtained a score of Enough (3) (pp. 38, 90), with the comment "critical thinking exists but can be strengthened." This assessment meets precisely with Qasrawi & Talafhah's (2020) analysis of the Unlock textbook series and the study of Mu et al., (2026) on the Chinese EFL textbooks showing the structural tendency across EFL textbooks around the world to include high-level thinking tasks as a supplement, rather than as the main scaffolding architecture. When the task of critical thinking is additive, teachers who face time pressure, which is an almost universal condition in Indonesian public secondary schools, will systematically deprioritize it.

The Pancasila Student Profile Strengthening Project (P5) provides a curricular mechanism to fill this gap. Project-based tasks embedded in the P5 schedule can operationalize the high-level thinking hinted at by textbooks but has not yet been fully realized. Rehman et al., (2024) and Saad & Zainuddin (2022), who examined project-based learning in the context of EFL Indonesia, documented consistent gains in critical thinking, collaboration, communication, and creativity when PBL was implemented with adequate teacher facilitation, right in pedagogical areas where the Criterion 3.4 gap demanded supplementation.

Fit for Student Level, Age, and Context

The Conformity dimension obtained a score of 20/25 (80.0%). Criteria 5.1 and 5.2 obtained Excellent (5) and Good (4) respectively: the

calibration of textbook proficiency was assessed as appropriate for CEFR A2-B1 (pp. 5, 30, 65), and the selection of topics relevant to the adolescent's interests (pp. 15, 50). Criterion 5.3 targeting local cultural contexts obtained Good (4): the Indonesian context is integrated into the assignments, a design feature that distinguishes this edition from global ELT textbooks that inherently refer to Anglophone cultural norms and risk alienating learners from their own linguistic and cultural identities (Javadi & Tahmasbi, 2020).

Criterion 5.4, inclusive and diversity-sensitive activities, was rated Adequate (3) (pp. 40, 88), with the comment "inclusivity can be improved." This evaluation is significant given Indonesia's sociocultural heterogeneity: more than 700 regional languages, various religious traditions, and populations spread across more than 17,000 islands. Dörnyei & Ushioda (2011) argue that centralized and standardized materials are inherently at risk of exclusion because their content is calibrated to dominant cultural norms, and may fail to serve students from rural, ethnic minority, or economically marginalized contexts. Motschenbacher (2019), in a multimodal analysis of EFL textbook representation, shows how the visual and verbal construction of social actors in textbooks both reinforces and challenges students' sense of ownership of legitimacy in language learning spaces.

The Sufficient Rating on inclusivity given by the author does not indicate that this book is exclusive. This ranking indicates that representations of gender diversity, regional cultural variations, and differences in learner needs can be designed more deliberately across units. Glas et al., (2023) in their research showed that inclusive material design based on the principles of Universal Design for Learning (UDL)

and Differentiated Instruction (DI) can be achieved in the context of a resource-constrained national curriculum, when designers approach diversity as a pedagogical resource, not as a category of administrative compliance.

Criterion 5.5, activities feasible to implement in terms of time, resources, and class size, obtained Good (4) (pp. 48, 100), confirming that the tasks can be realistically carried out under standard Indonesian public school conditions. This is not a trivial finding: many sophisticated textbooks theoretically fail at the limits of implementation because their assignments presuppose small classes, reliable internet access, or preparation time that is not available to the majority of teachers in the field. This feasibility rating hints that the design team is calibrating the assignment to the actual realities of the classroom.

A Good Rating on Eligibility combined with a Sufficient rating on inclusivity defines the priorities of the supplement most clearly for practitioners: the textbook can be implemented as designed, but teachers need to develop differentiated variants of assignments, particularly for the blended proficiency classes that are the norm in Indonesian public secondary schools. The flexibility mechanism of the Independent Curriculum, namely the teacher-made Teaching Module, the P5 project, and competency-based Learning Outcomes, creates a formal space for this differentiation work.

Synthesis and Pedagogical Implications

The five dimensions as a whole reveal a textbook that provides structural robustness, communicative task variety, and strong curricular alignment, while also having three systematically recurring weaknesses: uneven density of HOTS assignments, monotony of inter-unit formats, and underdeveloped inclusivity scaffolding. These three

weaknesses are not isolated defects. All three are interconnected expressions of a single design tension: the demands of national curriculum standardization confronted with the learner-responsiveness demands of contemporary ELT pedagogy.

The ELT research tradition offers a principled path to resolve these tensions at the classroom level. The framework for material development Tomlinson (2011) argues that effective language learning materials should stimulate learners cognitively and affectively, provide meaningful interaction with authentic input, and create space for individual responses. This textbook meets these criteria partially, not evenly. Field (2009) in listening pedagogy, and Nation (2001) in vocabulary and reading, both emphasize that skill development requires dispersed and repetitive practice across different types of tasks, a design logic that is directly relevant to this study's concern with cross-unit format repetition.

In the context of EFL Indonesia specifically, Setiawan (2026) identifies a structural paradox that successive curriculum reforms articulate increasingly sophisticated pedagogical aspirations, but classroom practice remains constrained by teachers' dependence on the material and the pressure of institutional exams. The Independent Curriculum formally acknowledges this paradox through its shift from content coverage to depth of competence, and thus creates the conditions that allow for the textbook-supplement partnership recommended by the findings of this study.

Overall Good Rating (80.0%) is on target: this textbook is a significant improvement over its predecessor editions and deserves to be used as a primary instructional resource for Class X English under the Independent Curriculum. Its use needs to be

supplemented by teacher-made HOTS assignment sequences, deliberate rotation of formats across units, and inclusivity-oriented assignment adaptations, especially for schools that serve linguistically and socioeconomically heterogeneous student populations. Future editions will benefit from three concrete improvements: a systematic mapping of the revised Bloom Taxonomy across the entire task population, an explicit 21st-century skills progress thread in unit-level objectives, and a diversification of representations of social actors in the context of tasks as well as visual materials.

Alignment of Content with Curriculum Goals

The Alignment dimension scored highest among the five at 22 out of 25 (88.0%), between Good and Very Good. CP mapping and skill sequencing drive this result, and both represent design choices the 2023 edition makes more deliberately than its K-13 predecessors.

Criteria 4.1 and 4.3, covering alignment with Phase E Learning Outcomes and language skill integration, were both rated Very Good (5). Each unit opens with communicative goals mapped to Kemdikbudristek's (2021) Capaian Pembelajaran descriptors for Phase E. All four language skills (listening, speaking, reading, and writing) appear in each unit in a receptive-to-productive sequence (p. 2, p. 42; p. 10, p. 55; p. 20, p. 70). Nation and Macalister (2010) identify explicit learning outcome specification and skill integration as the most reliable markers of curriculum-aligned language programs. The 2023 edition meets both criteria.

Criteria 4.2 and 4.4, covering CEFR calibration and CLT task orientation, were rated Good (4). Vocabulary load, grammatical complexity, and input text difficulty

align with the A2–B1 range across units (p. 5, p. 30, p. 65). Learners use language for genuine communicative purposes (negotiating, describing, narrating, arguing) rather than completing discrete-item accuracy drills (p. 25, p. 68; p. 30, p. 82). Harmer (2015) holds that CLT-informed coursebooks should prioritize meaning negotiation over structural display; this edition does. The contrast with earlier Ministry-issued textbooks is measurable: Herliana et al. (2020) documented the form-focused task bias that dominated predecessor editions.

Criterion 4.5, alignment with P5 and 21st-century skill development, was rated Good (4). Project tasks and collaborative activities across units address the Pancasila Student Profile competencies of critical thinking, creativity, and collaboration (p. 38, p. 92). These activities form a consistent strand, but they function as unit-closing supplements rather than competencies built progressively across units. Pardede (2020) shows that collaborative and creative tasks produce stronger learning outcomes when embedded in recurring structures rather than placed as one-off unit closers. Future editions should apply this design logic to the P5 strand.

CP alignment, CEFR calibration, skill sequencing, and CLT task design together explain why this dimension outscores the other four. Where the dimension falls short of Very Good is P5 integration: the 21st-century competencies Kurikulum Merdeka prioritizes appear in the coursebook but are not scaffolded progressively across units. Teachers can close this gap by treating project tasks as entry points for extended P5-aligned inquiry sequences, building the scaffolding the coursebook does not yet provide (Berlin & Wirza, 2026).

Conclusion

This study evaluated *Bahasa Inggris untuk SMA/MA Kelas X* using an instrument adapted from Cunningsworth (1995), McGrath (2002), Tomlinson (2011), and Miekley (2005) across five dimensions. Each framework validates a specific finding: Cunningsworth's (1995) curriculum-alignment criteria underpin the Alignment dimension's 88.0% score, while Tomlinson's (2011) SLA-grounded standard for authentic communicative task design explains the CLT quality rating. The coursebook scored 104 out of 125 points (83.2%), placing it in the Good category.

However, three systematic weaknesses were identified, namely the uneven distribution of HOTS tasks across units, the monotony of the format between units due to the repetition of genre cycles, and inclusivity that did not adequately reflect Indonesia's sociocultural heterogeneity. All three are expressions of a fundamental design tension between the demands of national curriculum standardization and the learner-responsiveness demands of contemporary ELT pedagogy.

For teachers, the Teaching Module and P5 mechanisms in the Independent Curriculum provide space to compensate for these three weaknesses through HOTS task supplements, format rotation, and UDL-based differentiation. For material developers, it is necessary to systematically map the Bloom Taxonomy across all units as well as diversify socio-cultural representations in future editions. For the Ministry of Education and Culture, these findings confirm the urgency of integrating a multidimensional evaluation protocol based on an established ELT framework into the national textbook procurement cycle.

From the results of this study, the author concludes that this textbook is suitable for use as a primary resource,

but requires planned supplements. The Independent Curriculum provides space for that, then the responsibility to fill it lies with reflective teachers, responsive developers, and evidence-based policies.

Reference

- Azizah, N. (2020). An Evaluation of K-13 English Textbooks: Limitations in Pronunciation Pedagogy and Learner Autonomy. *Journal of English Education and Linguistics Studies (JEELS)*, 7(2), 213–234. <https://doi.org/10.30762/jeels.v7i2.2451>
- Bağ, H., & Gürsoy, E. (2021). The Effect of Critical Thinking Embedded English Course Design to The Improvement of Critical Thinking Skills of Secondary School Learners. *Thinking Skills and Creativity*, 41, 100910. <https://doi.org/10.1016/j.tsc.2021.100910>
- Berlin, A. W., & Wirza, Y. (2026). From Curriculum Ideals to Textbook Practice: Learner Autonomy in Indonesia's Government-Issued EFL Textbook. *Voices of English Language Education Society*, 10(1), 250–263. <https://doi.org/10.29408/veles.v10i1.33058>
- Bhuana, B. G., & Lande, S. M. (2021). *Bahasa Inggris: Work in Progress untuk SMA/SMK/MA Kelas X*. Pusat Kurikulum dan Perbukuan, Kemdikbudristek. <https://buku.kemendikdasmen.go.id/katalog/Bahasa-Inggris-Work-in-Progress-untuk-SMASMKMA-Kelas-X>
- Chan, H., & Cheuk, H. N. J. (2020). Revisiting the notion of ESL: A corpus-based analysis of English textbook instructional language. *Ampersand*, 7, 100066.

- <https://doi.org/10.1016/j.amper.2020.100066>
- Cunningsworth, A. (1995). *Choosing Your Coursebook*. Heinemann.
- Demir, S., & Töngür, A. N. (2026). Lexical collocations in English as a Foreign Language (EFL) coursebooks: A corpus-based analysis of the usefulness score and pedagogical value. *Applied Corpus Linguistics*, 6(1), 100194. <https://doi.org/10.1016/j.acorp.2026.100194>
- Dörnyei, Z., & Ushioda, E. (2011). *Teaching and Researching Motivation* (2nd edn). Pearson Education.
- Egbert, M., Yufu, M., & Hirataka, F. (2016). An investigation of how 100 articles in the Journal of Pragmatics treat transcripts of English and non-English languages. *Journal of Pragmatics*, 94, 98–111. <https://doi.org/10.1016/j.pragma.2016.01.010>
- Erlingsson, C., & Brysiewicz, P. (2017). A hands-on guide to doing content analysis. *African Journal of Emergency Medicine*, 7(3), 93–99. <https://doi.org/10.1016/j.afjem.2017.08.001>
- Febriyani, R. A., Yunita, W., & Damayanti, I. (2020). An Analysis on Higher Order Thinking Skill (HOTS) in Compulsory English Textbook for the Twelfth Grade of Indonesian Senior High Schools. *Journal of English Education and Teaching*, 4(2), 170–183. <https://doi.org/10.33369/jeet.4.2.170-183>
- Field, J. (2009). *Listening in the Language Classroom*. Cambridge University Press.
- Glas, K., Catalán, E., Donner, M., & Donoso, C. (2023). Designing and providing inclusive ELT materials in times of the global pandemic: A Chilean experience. *Innovation in Language Learning and Teaching*, 17(1), 114–129. <https://doi.org/10.1080/17501229.2021.1940187>
- Hamied, F. A., & Lengkanawati, N. S. (2018). English as a medium of instruction in Indonesia: A case study. In R. Barnard & Z. Hasim (Eds), *English medium instruction programmes: Perspectives from South East Asian universities* (pp. 55–69). Routledge.
- Hanifa, R. (2018). EFL Published Materials: An Evaluation of English Textbooks for Junior High School in Indonesia. *Advances in Language and Literary Studies*, 9(2), 166. <https://doi.org/10.7575/aiac.all.v.9n.2p.166>
- Harmer, J. (2007). *The Practice of English Language Teaching*. Pearson Longman. <https://books.google.co.id/books?id=Q-S1oAEACAAJ>
- Harmer, J. (2015). *The Practice of English Language Teaching* (5th edn). Pearson Education Limited.
- Herliana, L., Djajanegara, S., & Suendarti, M. (2020). Students' and Teachers' Perspectives on EFL Textbook For Senior High School Students. *DEIKSIS*, 12(01), 44. <https://doi.org/10.30998/deiksis.v12i01.5585>
- Javadi, Y., & Tahmasbi, M. (2020). Application of Sociocultural Identity Theory to Education in EFL Textbooks Development. *Theory and Practice in Language Studies*, 10, 536–541. <https://doi.org/10.17507/tpls.1005.07>

- Khodabakhshi, M. (2014). Choose a Proper EFL Textbook: Evaluation of “Skyline” Series. *Procedia - Social and Behavioral Sciences*, 98, 959–967. <https://doi.org/10.1016/j.sbspro.2014.03.505>
- Lavrenteva, E., & Orland-Barak, L. (2023). Conceptual-analytical framework for exploring culture in EFL coursebooks: Analysis of teaching materials from a multimodal perspective. *Social Sciences & Humanities Open*, 7(1), 100441. <https://doi.org/10.1016/j.ssaho.2023.100441>
- Litz, D. R. A. (2005). Textbook Evaluation and ELT Management: A South Korean Case Study. *Asian EFL Journal*, 48(1), 1–53. <https://asian-efl-journal.com/wp-content/uploads/mgm/downloads/95912900.pdf>
- Maman, Priatna, J., & Mulyaningsih, I. (2021). *Buku Panduan Guru Bahasa Indonesia Tingkat Lanjut: Cakap Berbahasa dan Bersastra Indonesia untuk SMA/MA Kelas XII*. Pusat Kurikulum dan Perbukuan, Kemdikbudristek. <https://buku.kemendikdasmen.go.id/>
- McGrath, I. (2002). *Materials Evaluation and Design for Language Teaching*. Edinburgh University Press.
- Miekeley, J. (2005). ESL Textbook Evaluation Checklist. *The Reading Matrix*, 5(2), 9–17. http://www.readingmatrix.com/reading_projects/miekeley/project.pdf
- Motschenbacher, H. (2019). Non-nativeness as a dimension of inclusion: A multimodal representational analysis of EFL textbooks. *International Journal of Applied Linguistics*. <https://doi.org/10.1111/ijal.12237>
- Mrah, I. (2017). Developing Higher Order Thinking Skills: Towards a Rethinking of EFL Coursebooks in Moroccan High Schools. *Journal of English Language Teaching and Linguistics*, 2(3), 225. <https://doi.org/10.21462/jeltl.v2i3.79>
- Mu, H., Chen, M., Yang, J., & Zhang, F. (2026). Affording critical thinking through text-based reading comprehension questions: An evaluation of Chinese College English textbooks. *Thinking Skills and Creativity*, 62, 102248. <https://doi.org/10.1016/j.tsc.2026.102248>
- Musthafa, B. (2010). Teaching English to Young Learners in Indonesia: Essential Requirements. *Educationist*, 4(2), 120–125. <https://scispace.com/pdf/teaching-english-to-young-learners-in-indonesia-essential-mn4315fw65.pdf>
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Nation, I. S. P., & Macalister, J. (2010a). *Language Curriculum Design*. Routledge.
- Nation, I. S. P., & Macalister, J. (2010b). *Language Curriculum Design*. Routledge. <https://doi.org/10.4324/9780203847558>
- Pardede, P. (2020). Integrating the 4Cs into EFL Integrated Skills Learning. *JET (Journal of English Teaching)*, 6(1), 71–85.

- <https://doi.org/10.33541/jet.v6i1.190>
- Patton, M. Q. (2015). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice* (4th edn). SAGE Publications.
- Rahimah, S. (2021). EFL Textbook Evaluation: Investigating the Gaps in Critical Thinking Tasks and Autonomous Learning Activities. *Indonesian Journal of Applied Linguistics (IJAL)*, 11(1), 89–102.
<https://doi.org/10.17509/ijal.v11i1.34567>
- Rehman, N., Huang, X., Mahmood, A., AlGerafi, M. A. M., & Javed, S. (2024). Project-based learning as a catalyst for 21st-Century skills and student engagement in the math classroom. *Heliyon*, 10(23), e39988.
<https://doi.org/10.1016/j.heliyon.2024.e39988>
- Saad, A., & Zainudin, S. (2022). A review of Project-Based Learning (PBL) and Computational Thinking (CT) in teaching and learning. *Learning and Motivation*, 78, 101802.
<https://doi.org/10.1016/j.lmot.2022.101802>
- Setiawan, B. N. (2026). Exploring interculturality in Indonesian English language classrooms: Reciprocal dialogues between local, national, and global English languages and cultures. *Asian Englishes*, 28(1), 170–192.
<https://doi.org/10.1080/13488678.2025.2602101>
- Tomlinson, B. (2011). *Materials Development in Language Teaching* (2nd edn). Cambridge University Press.
- Tyas, M., Nurkamto, J., & Marmanto, S. (2020). Cultivating Students' Higher-Order Thinking Skills in EFL Classes: The Role of the Teacher and the Textbook. *International Online Journal of Education and Teaching*, 7, 267–276.
- Yuliani, S., & Aziza, N. (2022). Content Analysis of K-13 English Textbooks: A Focus on Pronunciation, Critical Thinking, and Learning Autonomy. *Journal of English Language Teaching (JELT)*, 11(3), 411–426.
<http://unp.ac.id>