Ta'dibuna Jurnal Pendidikan Islam
P-ISSN: 2252-5793
E-ISSN: 2622-7215
Vol15, No.5, Oktober 2026, hlm. 776-794
DOI: 10.32832/tadibuna.V15i5.24548
http://ejournal.uika-bogor.ac.id/index.php/TADIBUNA/

Readability Progression and Students' Perceptions of Gontor Press English Lesson Textbooks in Indonesian Pesantren

Waridatul Hikmah[1], Feisal Aziez[2], John Louie Galang[3]
[1,2]Universitas Muhammadiyah Purwekerto, Indonesia
[3]Nueva Ecija University of Science and Technology, The Philippines
*feisalaziez@ump.ac.id

Abstract

This study investigates the readability level of English textbooks published by Gontor Press and used in pesantren affiliated with Darussalam Gontor. The study aims to evaluate the readability of the textbooks using the Flesch Reading Ease (FRE) and Flesch–Kincaid Grade Level (FKGL) formulas and to examine students' perceptions of the texts. Employing a descriptive quantitative case study design, the research analyzed 18 reading passages from three English Lesson textbooks (Grades 1, 2, and 3) and collected questionnaire responses from 400 students in pesantren located in Central Java. The findings revealed that most reading texts fell within the easy to very easy categories based on FRE scores, indicating relatively low estimated linguistic complexity. The FKGL analysis showed that the texts ranged from lower to intermediate grade levels, with variations in difficulty and an inconsistent progression across grades. The questionnaire results further indicated that the majority of students perceived the texts as easy and understandable. These findings suggest that the English textbooks published by Gontor Press generally contain accessible reading materials, although formula-based readability estimates do not always correspond to students' perceived reading difficulty. Therefore, a more systematic progression of text complexity across grade levels may be necessary to provide appropriate linguistic challenges and support continuous language development. The study highlights the importance of considering both textual characteristics and learners' reported experiences when reviewing the selection and sequencing of reading passages in pesantren EFL contexts.

Keywords: *Readability; Flesch Reading Ease; Flesch-Kincaid Grade Level; English Textbook*

Abstrak

Penelitian ini mengkaji tingkat keterbacaan buku pelajaran bahasa Inggris yang diterbitkan oleh Gontor Press dan digunakan di pesantren-pesantren yang berafiliasi dengan Darussalam Gontor. Penelitian ini bertujuan untuk mengevaluasi keterbacaan buku-buku pelajaran tersebut menggunakan rumus Flesch Reading Ease (FRE) dan Flesch–Kincaid Grade Level (FKGL), serta untuk mengkaji persepsi siswa terhadap teks-teks tersebut. Dengan menggunakan desain studi kasus kuantitatif deskriptif, penelitian ini menganalisis 18 bacaan dari tiga buku pelajaran bahasa Inggris (Kelas 1, 2, dan 3) serta mengumpulkan tanggapan kuesioner dari 400 siswa di pesantren yang berlokasi di Jawa Tengah. Temuan penelitian menunjukkan bahwa sebagian besar teks bacaan termasuk dalam kategori mudah hingga sangat mudah berdasarkan skor FRE, yang

Diserahkan: 14-07-2026 **Disetujui**: 20-10-2026. **Dipublikasikan**: 03-11-2026

Kutipan: Hikmah, W., Aziez, F., & Galang, J. L. (2026). Readability Progression and Students' Perceptions of Gontor Press English Lesson Textbooks in Indonesian Pesantren. Ta'dibuna Jurnal Pendidikan Agama Islam, 15(5), 776–794. https://doi.org/10.32832/tadibuna.V15i5.24548

mengindikasikan kompleksitas linguistik yang relatif rendah. Analisis FKGL menunjukkan bahwa teks-teks tersebut berkisar dari tingkat kelas bawah hingga menengah, dengan variasi tingkat kesulitan dan perkembangan yang tidak konsisten antar kelas. Hasil kuesioner lebih lanjut menunjukkan bahwa mayoritas siswa menganggap teks-teks tersebut mudah dan dapat dipahami. Temuan ini menunjukkan bahwa buku teks bahasa Inggris yang diterbitkan oleh Gontor Press secara umum berisi bahan bacaan yang mudah dipahami, meskipun perkiraan keterbacaan berbasis rumus tidak selalu sesuai dengan tingkat kesulitan membaca yang dirasakan oleh siswa. Oleh karena itu, perkembangan kompleksitas teks yang lebih sistematis di seluruh tingkatan kelas mungkin diperlukan untuk memberikan tantangan linguistik yang sesuai dan mendukung perkembangan bahasa yang berkelanjutan. Penelitian ini menyoroti pentingnya mempertimbangkan baik karakteristik teks maupun pengalaman yang dilaporkan oleh peserta didik saat mengevaluasi pemilihan dan urutan bacaan dalam konteks pembelajaran bahasa Inggris sebagai bahasa asing (EFL) di pesantren.

Kata kunci: Keterbacaan; Flesch Reading Ease; Flesch-Kincaid Grade Level; Buku Teks Bahasa Inggris.

I. Introduction

The Islamic boarding schools (pesantren) are among the oldest and most influential Islamic educational institutions in Indonesia. Initially established to promote Islamic education and prepare future religious scholars, pesantren have progressively integrated general subjects, including English, into their educational curricula (Nuryani et al., 2018; Aziez, 2018). The term pesantren is derived from the Arabic word funduq, meaning dormitory or lodging, while Islamic boarding schools refer to educational institutions where students study Islamic knowledge within a residential learning environment (Rohmah & Roqib, 2024). One prominent example is Pondok Modern Darussalam Gontor, which implements the Kulliyyatul Mu'allimin Al-Islamiyyah (KMI) curriculum. This curriculum integrates religious and general education while emphasizing foreign language learning, particularly Arabic and English (Rohmah & Roqib, 2024). To support English instruction, Gontor Press publishes the English Lesson textbooks, which serve as instructional materials in Gontor-affiliated pesantren. Given their role in English language instruction, evaluating the linguistic characteristics of these textbooks is important for understanding their suitability as learning materials. The Gontor Press textbook is the primary source of the KMI curriculum. This curriculum integrates religious and general education through the principle of 100% religious knowledge and 100% general knowledge, emphasizing the equal importance of both fields within Islamic education. Foreign language learning, particularly Arabic and English, forms an integral part of this educational approach (Rohmah & Roqib, 2024). To support English instruction, Gontor Press publishes the English Lesson textbooks, which serve as instructional materials in Gontor-affiliated pesantren. Given their role in the integrated KMI curriculum, evaluating the readability of these textbooks is relevant not only to English language instruction but also to the broader evaluation of teaching materials within Islamic education.

As a systematic teaching material, a textbook or module provides planned learning experiences, minimum objectives, content, and evaluations to help students master specific goals (Pulukadang et al., 2020). While a learning scenario is incomplete without it, a skilled teacher can fully maximize the potential of a textbook despite any flaws it may have (Mohammadi & Abdi, 2014). Textbook evaluation is an essential component of effective language instruction. Tomlinson (2012) emphasizes that teaching materials should be evaluated according to learners' needs and instructional objectives, while Richards (2001) highlights the importance of considering learners' proficiency and educational contexts. Systematic textbook evaluation is therefore necessary to determine whether instructional materials are appropriate for a particular educational context (Porter et al., 1995). Since textbooks serve as a major source of language input and classroom guidance (Harmer, 2001), one important aspect of textbook evaluation is readability, which concerns the linguistic characteristics that influence the relative ease or difficulty of reading a text.

Readability can be estimated through standardized formulas, including Flesch Reading Ease (FRE) and Flesch–Kincaid Grade Level (FKGL), which assess textual complexity using sentence length and syllabic characteristics (Kincaid et al., 1975). However, formula-based readability estimates do not fully capture the factors influencing how individual learners experience reading materials. Crossley et al. (2008) emphasize that other linguistic and reader-related characteristics should also be considered when evaluating textual difficulty. In addition to readability analysis, researchers have increasingly emphasized the importance of incorporating learners' perceptions when evaluating instructional materials because students are the primary users of textbooks (Mukundan et al., 2011; Tomlinson, 2012). Similarly, DuBay (2004) argues that readability serves as an important indicator for determining whether instructional materials correspond to learners' linguistic competence.

Previous studies have examined English instructional materials from different perspectives. Hakim et al. (2021) and Zainurrahman et al. (2024) investigated textbook readability in an Indonesian senior high school context by considering readability measurements and students' perceptions. Meanwhile, studies involving Gontor Press textbooks have examined their educational content and instructional characteristics, including the representation of Islamic culture (Nuryani et al., 2018) and second language acquisition principles (Saputri, 2025). These studies provide important insights into textbook evaluation but address different aspects of instructional materials. In particular, the relationship between standardized readability estimates and students' perceptions across successive levels of Gontor Press English Lesson textbooks remains insufficiently explored. This research gap is relevant to Islamic education because evaluating the linguistic accessibility of English textbooks can provide evidence for improving instructional materials within the integrated KMI curriculum. Unlike studies focusing primarily on Islamic cultural representation or textbook content, the present study

examines the progression of textual complexity and students' perceived reading difficulty as complementary aspects of educational material evaluation in pesantren. Furthermore, previous research has identified challenges in the implementation of Gontor Press English textbooks. Muhith (2018) reported concerns regarding lengthy reading texts and insufficient vocabulary support for students. Pedagogically, these textbooks mismatch national curriculum standards and prompt a heavy reliance on the traditional Grammar-Translation Method (GTM) over communicative approaches. While these qualitative concerns are documented (Muhith, 2018), there remains a critical gap in objective, empirical evaluations of the textbooks' actual readability levels.

This gap is particularly relevant because the three levels of English Lesson textbooks are used within the KMI educational context, where learners encounter progressively organized English instructional materials. Evaluating readability across successive textbook levels can provide information about whether linguistic complexity increases consistently. Furthermore, incorporating students' perceptions may reveal differences between the estimated difficulty of reading passages and the difficulty experienced by learners. Such information is relevant to textbook evaluation because reading difficulty is not determined exclusively by sentence length and word characteristics but may also be associated with vocabulary familiarity, textual organization, and background knowledge (Crossley et al., 2008; Nguyen et al., 2024).

This study is crucial because selecting appropriate textbooks is essential for ensuring effective English language learning. Textbooks with suitable readability levels can help students comprehend learning materials more easily, increase their motivation, and support the achievement of learning objectives. Moreover, understanding students' perceptions can provide valuable insights into how learners experience and evaluate the textbook in practice (Ehara,2021). Therefore, this study aims to (1) determine the readability levels of reading passages in Gontor Press English Lesson textbooks using the Flesch Reading Ease (FRE) formula, (2) examine their estimated grade levels and difficulty progression using the Flesch–Kincaid Grade Level (FKGL) formula, and (3) investigate students' perceptions of the readability of these passages. By integrating formula-based readability analysis with students' reported perceptions across three textbook levels, the study seeks to provide a more detailed evaluation of English instructional materials used in Gontor-affiliated pesantren and identify aspects that may require further consideration in textbook development. In the broader context of Islamic education, the findings may contribute to improve the selection and sequencing of English reading materials within the integrated KMI curriculum, thereby supporting more context-sensitive instructional material evaluation in pesantren.

II. Research Methods

This This study employed a descriptive quantitative case study design to investigate the readability of English textbooks published by Gontor Press and students' perceptions of the reading passages. Two sources of quantitative data were examined: readability scores obtained using the Flesch Reading Ease (FRE) and Flesch–Kincaid Grade Level (FKGL) formulas and students' responses to a structured questionnaire. The FKGL formula is widely recognized as an objective measure for estimating the educational grade level required to comprehend a text and has frequently been used to evaluate the appropriateness of instructional materials (Ponnudurai et al., 2025). Descriptive quantitative analysis was employed to summarize the distribution of readability scores and students' perceptions across the three textbook levels, consistent with the quantitative research procedures discussed by Libhaber (2025). The two sources of data were subsequently compared at the individual-passage level to identify similarities and discrepancies between estimated linguistic complexity and perceived reading difficulty.

The study was conducted in three Gontor-affiliated Islamic boarding schools located in Purwokerto, Cilacap, and Kebumen, Central Java, Indonesia. The target population was estimated at approximately 15,000 students across Gontor-affiliated pesantren. The minimum sample size was calculated using Slovin's formula, as discussed by Adhikari (2021, p. 26), with an assumed margin of error of 5%, resulting in approximately 390 respondents. The research sites were subsequently selected through purposive sampling based on their implementation of the Gontor educational curriculum and consistent use of English Lesson textbooks published by Gontor Press. All 400 eligible students in the three selected institutions were included in the questionnaire survey, resulting in 193 students (48.25%) using English Lesson 1, 136 students (34.00%) using English Lesson 2, and 71 students (17.75%) using English Lesson 3. Thus, the study combined purposive selection of research sites with total sampling of eligible students within those institutions. The findings were interpreted within the context of the participating pesantren rather than statistically generalized to the broader population.

The primary textual data consisted of 18 eligible reading passages from three English textbooks published by Gontor Press, comprising nine passages from English Lesson 1, four passages from English Lesson 2, and five passages from English Lesson 3. Each passage served as a unit of analysis. The reading passages were extracted from the textbooks without intentional modification, while non-reading components, such as exercises and instructional directions, were excluded from the readability assessment.

Students' perceptions were collected using a structured questionnaire adapted from Hakim et al. (2021) and adjusted to the reading passages in each textbook. The questionnaire employed a four-point Likert-type scale ranging from Very Difficult (1) to Very Easy (4), following the response format discussed by Dörnyei and Taguchi (2009). Students were also asked to identify the easiest and most difficult passages and select reasons for their judgments. Three questionnaire versions were administered according

to the textbook levels used by the respondents, comprising 13 items for English Lesson 1, eight for English Lesson 2, and nine for English Lesson 3. The distribution and structure of the questionnaire items are presented in Table 1. As no separate validity or reliability testing was reported for the adapted instrument, the responses were interpreted descriptively as indicators of students' perceived reading difficulty rather than direct measures of reading comprehension.

These examples illustrate how students' perceptions were collected and subsequently used to complement the formula-based readability analysis.

Table 1. Distribution and Structure of the Students' Perception Questionnaire

Questionnaire Component	English Lesson 1	English Lesson 2	English Lesson 3
Respondents	193	136	71
Reading passages	9	4	5
Readability rating items	9	4	5
Easiest text selection	1	1	1
Reason for easiest text	1	1	1
Most difficult text selection	1	1	1
Reason for most difficult text	1	1	1
Total items	13	8	9

The readability of each passage was analyzed using the Flesch Reading Ease (FRE) and Flesch–Kincaid Grade Level (FKGL) formulas (Kincaid et al., 1975). The readability score also could be seen at www.readable.com (Darwin, 2022). Both formulas estimate textual complexity based on average sentence length (ASL) and average syllables per word (ASW), as follows:

Flesch Reading Ease (FRE):

$$FRE = 206.835 - 1.015(ASL) - 84.6(ASW)$$

Flesch–Kincaid Grade Level (FKGL):

$$FKGL = 0.39(ASL) + 11.8(ASW) - 15.59$$

Higher FRE scores generally indicate easier reading materials, whereas higher FKGL scores indicate greater estimated linguistic complexity. The reading passages were assessed using an online readability analysis tool, and the resulting scores were recorded and classified according to established readability categories. The scores were compared across the three textbook levels to examine variations and patterns in estimated linguistic difficulty.

Questionnaire responses were analyzed using descriptive statistics, particularly frequencies and percentages. The percentage of each response category was calculated by dividing its frequency by the total number of valid responses and multiplying by 100. Students' selections of the easiest and most difficult reading passages and their reported reasons were also summarized using percentage distributions.

Finally, the FRE and FKGL scores were compared descriptively with students' perceptions at the individual-text level to identify similarities and discrepancies between estimated linguistic complexity and perceived reading difficulty. This comparison was descriptive rather than correlational. The readability formulas were treated as indicators of textual characteristics, while questionnaire responses represented students' subjective judgments. Neither measure was interpreted as direct evidence of students' actual reading comprehension or instructional effectiveness

III. Result and Discussion

A. Readability of English Textbooks based on Flesch Reading Ease (FRE)

This section addresses the first research question by examining the readability level of English lesson textbooks published by Gontor press, used in grade 1,2, and 3 in pesantren affiliated to Gontor through the use of Flesch Reading Ease (FRE) formula. The Flesch Reading Ease (FRE) analysis revealed variations in the estimated readability of 18 reading texts across three English textbooks published by Gontor Press. The scores ranged from 76 to 101 in English Lesson 1, 71 to 95 in English Lesson 2, and 78 to 87 in English Lesson 3. Most texts were categorized as easy or very easy, although several passages demonstrated relatively greater linguistic complexity. The distribution of the scores indicates that the textbooks generally contain structurally accessible reading materials, but the degree of readability differs both within and across instructional levels. The individual FRE scores of all 18 reading passages, together with FKGL scores are presented in Table 2 to facilitate comparison across the three textbook levels.

Table 2. FRE and FKGL Scores Across the Three English Lesson Textbooks

Textbook	Reading Text	FRE	FKGL
EL1	Seventh Lesson	84	3.85
EL1	Article of Clothing	96	1.44
EL1	Ninth Lesson	95	1.68
EL1	Tenth Lesson	97	1.54
EL1	Twelfth Lesson	97	1.84
EL1	Thirteenth Lesson	101	1.43
EL1	The Clock and the Watch	89	3.48
EL1	Reading and Conversation 1	95	2.68
EL1	Reading and Conversation 2	76	7.78
EL2	The Weather	95	2.47
EL2	An Invitation to Dinner	71	8.05
EL2	At the Seaside	90	4.54
EL2	Traveling	80	4.48
EL3	The Arrival	86	3.71
EL3	In London	87	4.15
EL3	In the Country	81	4.41

EL3	Animal	78	6.53
EL3	Human Beings	85	4.02

In English Lesson 1, the FRE scores varied considerably among the nine reading texts. Thirteenth Lesson obtained the highest score of 101. Regarding the Flesch Reading Ease (FRE) score of 101 calculated for the Thirteenth Lesson, the metric remains mathematically valid as the original algebraic formula is not artificially constrained to the conventional 0–100 interpretive range; instead, text segments characterized by exceptionally short sentences and words with few syllables minimize the subtractive variables against the algorithm's baseline constant, linearly driving the final output above 100 (DuBay, 2008; Kincaid et al., 1975). Crucially, this exceptional value must be interpreted strictly as an index of high structural readability and mechanical accessibility rather than as empirical evidence of students' actual cognitive reading comprehension, which remains dependent on non-textual factors such as prior knowledge, motivation, and conceptual density (DuBay, 2008). The next FRE scores were recorded for Tenth Lesson and Twelfth Lesson, both scoring 97. In contrast, Reading and Conversation 2 recorded the lowest score of 76, placing it in the fairly easy category. The difference between these scores indicates that the first textbook does not maintain a uniform level of textual complexity. Although most passages contain linguistic characteristics associated with high readability, Reading and Conversation 2 presents relatively greater estimated difficulty. This variation is important because students using the same textbook may encounter reading passages with substantially different structural demands.

A similar pattern emerged in English Lesson 2, where FRE scores ranged from 71 to 95. The Weather obtained the highest score of 95, whereas An Invitation to Dinner recorded the lowest score of 71. The remaining texts, At the Seaside and Traveling, obtained scores of 90 and 80, respectively. These differences demonstrate that the second textbook incorporates reading materials with varying levels of estimated linguistic complexity. Notably, An Invitation to Dinner represents a comparatively demanding passage, even though it appears alongside considerably easier texts. This variation suggests that reading difficulty is not necessarily distributed consistently within the textbook.

In English Lesson 3, the FRE scores were more concentrated, ranging from 78 to 87. In London recorded the highest score of 87, followed by The Arrival (86), Human Beings (85), and In the Country (81), while Animal obtained the lowest score of 78. Unlike the first two textbooks, which included several very easy texts, English Lesson 3 consisted predominantly of passages classified as easy, with one fairly easy passage. This narrower distribution indicates a more consistent level of estimated linguistic difficulty in the third textbook. Nevertheless, the overlap of scores across the three textbooks suggests that the progression of readability is not strictly hierarchical.

These findings are consistent with Ningrum et al. (2023) and Hanifa (2018), whose studies, as discussed in the present thesis, identified relatively high readability levels in English instructional materials. However, the present findings highlight an additional issue concerning the distribution of difficulty across successive textbook levels. While most texts demonstrate high readability, the presence of relatively demanding passages in earlier textbooks indicates that textual complexity does not always increase systematically. This observation corresponds with Permatasari et al. (2022), who emphasized the importance of evaluating the progression of reading difficulty across educational levels.

From a pedagogical perspective, the predominance of highly readable texts may provide conditions favorable for comprehensible language input. Krashen's (2008) Comprehensible Input Hypothesis emphasizes the importance of exposing learners to language that is understandable while offering appropriate linguistic challenges. However, high FRE scores alone cannot establish whether a passage represents appropriate input for a particular group of learners. Such an interpretation also requires information about students' existing proficiency and their responses to the materials.

Furthermore, Trumbull's (1984) Stages of Reading Development emphasizes the importance of gradually increasing textual complexity as learners develop their reading abilities. In the present study, the overlap of FRE scores among the three textbooks raises questions about the consistency of this progression. Rather than indicating a uniformly increasing level of difficulty, the findings suggest that accessible and relatively demanding texts coexist across different instructional levels. This pattern highlights the need to consider the sequencing of reading materials in addition to their individual readability classifications.

Nevertheless, the interpretation of these findings must acknowledge the limitations of the FRE formula. As explained by Kincaid et al. (1975), FRE primarily estimates readability through average sentence length and syllabic complexity. Crossley et al. (2008) further emphasized that these measures do not fully represent other factors affecting reading difficulty, including vocabulary familiarity, textual cohesion, and readers' background knowledge. Therefore, the high FRE scores observed in this study should be interpreted as indicators of relatively low structural complexity rather than direct evidence that students can successfully comprehend every passage.

Overall, the FRE findings reveal that Gontor Press English textbooks predominantly contain texts with high estimated readability, although their difficulty progression is not entirely consistent. The findings support the importance of systematic textbook evaluation emphasized by Tomlinson (2012) and Richards (2001), particularly in ensuring that reading materials balance linguistic accessibility with appropriate progression across instructional levels.

B. Readabilityi of English Textbooks Based on Flesn-Kincaid Grade Level (FKGL)

The Flesch–Kincaid Grade Level (FKGL) analysis provided further insights into the estimated linguistic complexity of the reading texts across the three English textbooks. The scores ranged from 1.43 to 7.78 in English Lesson 1, 2.47 to 8.05 in English Lesson 2, and 3.71 to 6.53 in English Lesson 3. These results indicate that the texts represented different estimated reading grade levels, with substantial variation within individual textbooks and considerable overlap across instructional levels. As presented in Table 2, the FKGL scores varied considerably across individual reading passages, revealing overlapping difficulty levels among the three textbooks.

In English Lesson 1, eight of the nine reading texts obtained FKGL scores below 4.00, indicating relatively low levels of estimated linguistic complexity. Thirteenth Lesson recorded the lowest score of 1.43, followed by Article of Clothing (1.44) and Tenth Lesson (1.54). However, Reading and Conversation 2 obtained a considerably higher score of 7.78, corresponding approximately to an eighth-grade reading level. This substantial difference indicates that the linguistic complexity of reading passages within the first textbook was not evenly distributed. The presence of a relatively demanding text alongside predominantly simple passages suggests that students may encounter different levels of structural difficulty within the same instructional material.

English Lesson 2 demonstrated a similarly uneven distribution of FKGL scores. The Weather obtained the lowest score of 2.47, while An Invitation to Dinner recorded the highest score of 8.05. Meanwhile, At the Seaside and Traveling obtained scores of 4.54 and 4.48, respectively. An Invitation to Dinner was therefore the most linguistically demanding text across all three textbooks according to the FKGL formula. This result is particularly noteworthy because the highest estimated grade level occurred in the second textbook rather than the third. The finding indicates that increasing textbook levels do not necessarily correspond to a continuous increase in the structural complexity of individual reading passages.

In English Lesson 3, the FKGL scores ranged from 3.71 to 6.53. The Arrival recorded the lowest score of 3.71, followed by Human Beings (4.02), In London (4.15), and In the Country (4.41). Animal obtained the highest score of 6.53. Compared with the previous textbooks, the third textbook showed a narrower distribution of estimated grade levels, with most texts concentrated around the fourth-grade level. Although English Lesson 3 generally contained fewer passages with extremely low FKGL scores, none exceeded the highest scores observed in English Lessons 1 and 2.

The comparison across the three textbooks reveals an important distinction between the overall tendency of difficulty progression and the complexity of individual texts. Although the third textbook contained a more concentrated range of moderately complex

texts, Reading and Conversation 2 in English Lesson 1 (7.78) and An Invitation to Dinner in English Lesson 2 (8.05) exceeded the highest FKGL score recorded in English Lesson 3 (6.53). This overlap indicates that reading difficulty was not organized in a strictly progressive sequence across the three instructional levels.

The variation in FKGL scores across the three textbooks also raises questions about how linguistic complexity is distributed within a sequence of instructional materials. Although the textbooks are designed for successive educational levels, their numerical classifications do not establish a consistently increasing pattern of structural difficulty. In particular, the presence of relatively demanding passages in earlier textbooks suggests that progression should be evaluated through the distribution of individual reading texts rather than exclusively through their placement in a particular grade level. This distinction is important because the instructional sequence may expose learners to different degrees of linguistic complexity within and across textbook levels.

Another important consideration concerns the interpretation of FKGL scores in the context of English as a Foreign Language instruction. Although FKGL expresses textual complexity through estimated grade levels, these classifications were originally developed to represent characteristics of English texts rather than the specific proficiency profiles of Indonesian pesantren students. A numerical grade-level estimate should therefore not be treated as equivalent to students' actual educational placement or reading ability. This distinction is particularly relevant in pesantren settings, where learners' exposure to English may differ from the educational environments in which traditional readability formulas were developed.

From the perspective of instructional material evaluation, examining progression requires attention not only to the difficulty of individual passages but also to their relationship within a broader learning sequence. A relatively demanding passage may be pedagogically appropriate when students receive sufficient preparation and instructional support. Conversely, a sequence dominated by structurally simple passages may warrant further examination to determine whether it offers opportunities for increasingly challenging language exposure. However, neither possibility can be established solely through readability scores. Additional information about classroom instruction, learning objectives, and students' language proficiency would be necessary to evaluate the appropriateness of the observed patterns.

These considerations extend the interpretation of FKGL beyond the identification of estimated grade levels. The findings demonstrate the usefulness of passage-level analysis for identifying differences in structural demands, while also emphasizing the importance of contextual interpretation. Rather than assuming that textbook levels necessarily correspond to progressively higher readability demands, educators and material developers may benefit from reviewing the sequencing of reading passages in relation to intended learning outcomes. Such evaluation could contribute to a more systematic

consideration of textual complexity without reducing instructional suitability to numerical readability classifications alone.

This finding is relevant to Trumbull's (1984) Stages of Reading Development, which emphasizes gradual exposure to increasingly complex texts as learners advance. The distribution of FKGL scores suggests that the analyzed textbooks do not consistently follow such progression at the level of individual reading passages. However, this does not necessarily indicate that students experience inappropriate reading difficulty, since FKGL scores represent estimated linguistic complexity rather than direct measurements of learners' abilities.

The findings also extend the discussion of textbook readability presented by Khodadady (2017) and Permatasari et al. (2022), particularly regarding the importance of examining the relationship between textual complexity and instructional levels. Rather than relying exclusively on the overall readability classification of a textbook, the present analysis demonstrates the value of examining individual passages to identify differences in estimated difficulty. This approach provides a more detailed understanding of how linguistic demands are distributed throughout instructional materials.

From the perspective of textbook evaluation, Richards (2001) and Tomlinson (2012) emphasize that instructional materials should correspond to learners' needs, proficiency levels, and educational objectives. The considerable variation in FKGL scores observed in this study suggests that the progression of linguistic complexity warrants further evaluation. In particular, the presence of highly demanding passages in earlier instructional levels raises questions about whether textbook sequencing sufficiently reflects the gradual development of learners' reading abilities. Nevertheless, determining the actual appropriateness of these passages would require additional evidence concerning students' language proficiency and reading comprehension.

Furthermore, the FKGL formula has methodological limitations. As noted by Crossley et al. (2008), traditional readability formulas primarily measure surface linguistic characteristics and do not fully capture factors such as discourse organization, lexical familiarity, and background knowledge. Consequently, FKGL values should not be interpreted as precise indicators of the grade level at which EFL students will successfully comprehend a passage. Instead, they provide estimates that are useful for comparing the structural demands of reading materials.

Overall, the FKGL findings demonstrate that the English textbooks published by Gontor Press contain reading texts with varied levels of estimated linguistic complexity. Although most passages fall within relatively low to moderate grade-level estimates, the substantial overlap across textbook levels indicates an inconsistent progression of difficulty. These findings suggest that future textbook evaluation should consider the

sequencing of reading passages alongside their individual readability scores to support a more systematic progression of linguistic complexity.

C. Students' Perception of the Text in English Textbook Published by Gontor Press

Students' perceptions of the reading passages were examined to complement the readability estimates obtained from the Flesch Reading Ease (FRE) and Flesch–Kincaid Grade Level (FKGL) formulas. The questionnaire assessed students' perceived reading difficulty and asked them to identify the easiest and most difficult passages in their respective textbooks. Across the three instructional levels, the proportion of responses in the easy category ranged from 47% to 63% in English Lesson 1, 45% to 63% in English Lesson 2, and 37% to 55% in English Lesson 3. These responses indicate that many students perceived the reading materials as relatively accessible. However, the selection of individual passages as the easiest or most difficult revealed differences between students' judgments and formula-based readability estimates. Table 3 presents the distribution of these selections across the three textbooks.

Table 3. Comparison of Students' Perceived Reading Difficulty Across Three English Lesson Textbooks.

Textbook	Reading Text	Easiest Text (%)	Most Difficult Text (%)
English Lesson 1	Seventh Lesson	38	6
	Article of Clothing	19	7
	Ninth Lesson	7	9
	Tenth Lesson	10	3
	Twelfth Lesson	1	31
	Thirteenth Lesson	6	10
	The Clock and the Watch	6	4
	Reading and Conversation 1	3	18
	Reading and Conversation 2	10	12
English Lesson 2	The Weather	69	3
	An Invitation to Dinner	10	29
	At the Seaside	13	25
	Traveling	8	43
English Lesson 3	The Arrival	22	18
	In London	10	25
	In the Country	10	20
	Animal	51	7
	Human Beings	7	30

The comparison between students' perceptions and readability scores revealed different patterns across the three textbook levels. In English Lesson 1, the Seventh Lesson was identified as the easiest passage by 38% of respondents, although its FRE score of 84 and FKGL score of 3.85 did not indicate the lowest estimated linguistic complexity in the textbook. Conversely, the Twelfth Lesson, with an FRE score of 97 and an FKGL score of 1.84, was selected as the most difficult passage by 31% of students. This

contrast suggests that passages characterized by relatively simple sentence structures and syllabic patterns were not necessarily perceived as the easiest by learners.

In English Lesson 2, the findings showed both correspondence and divergence between the two approaches. The Weather, which obtained an FRE score of 95 and an FKGL score of 2.47, was selected as the easiest passage by 69% of students. This result was consistent with its relatively low estimated linguistic complexity. In contrast, Traveling was identified as the most difficult passage by 43% of respondents, despite having a lower FKGL score (4.48) than An Invitation to Dinner (8.05). A more pronounced discrepancy appeared in English Lesson 3, where Animal obtained the lowest FRE score (78) and the highest FKGL score (6.53), yet was selected as the easiest passage by 51% of respondents. Meanwhile, Human Beings, which had an FRE score of 85 and an FKGL score of 4.02, was most frequently identified as the most difficult passage (30%). These differences demonstrate that the relative difficulty indicated by readability formulas did not consistently correspond to students' reported experiences with individual passages.

Students' explanations provided additional insight into these contrasting evaluations. In English Lessons 1 and 2, background knowledge (40%) and vocabulary knowledge (30%) were prominent reported reasons for identifying passages as easy. In English Lesson 3, students associated the perceived ease of Animal with short and grammatically simple sentences (38%) and familiarity with the topic (24%). Conversely, students attributed difficulty to factors including unfamiliar vocabulary, complex sentence structures, and insufficient knowledge of the subject matter. These responses suggest that students considered a broader range of textual and reader-related characteristics when evaluating reading difficulty than those directly represented by FRE and FKGL.

The differences between formula-based estimates and students' perceptions can be interpreted in relation to the limitations of traditional readability measures. FRE and FKGL estimate linguistic complexity primarily through sentence length and syllabic characteristics, whereas students' experiences of reading may also involve lexical familiarity, discourse organization, and background knowledge. Crossley et al. (2008) emphasized the relevance of linguistic dimensions beyond traditional readability indicators when evaluating text difficulty. The present findings illustrate this distinction, particularly in passages where relatively high estimated structural complexity coincided with students' reports of reading ease. Nevertheless, the questionnaire results do not establish the relative contribution of each factor to students' judgments, since the study employed descriptive rather than explanatory or correlational analysis.

Another consideration is that students' perceptions may reflect their familiarity with the instructional context in which the reading materials are used. Since the participants regularly encountered English Lesson textbooks within the pesantren curriculum, their judgments may have been shaped by previous exposure to similar vocabulary, topics, and

learning activities. Although these factors were not independently measured, they suggest the importance of examining readability within the educational context rather than relying exclusively on numerical estimates.

From a pedagogical perspective, these findings are relevant to the evaluation of instructional materials in English as a Foreign Language (EFL) context. Tomlinson (2012) and Richards (2001) emphasize the importance of considering learners' characteristics, educational needs, and instructional objectives when evaluating teaching materials. In the present study, the discrepancies between readability estimates and perceived difficulty indicate that formula-based classification alone may not adequately represent how particular reading passages are experienced by students. A passage with relatively demanding structural characteristics may remain familiar and manageable to learners, whereas a structurally simpler passage may be perceived as difficult when its vocabulary, topic, or organization presents additional challenges.

The findings also offer a qualified connection to Krashen's (2008) Comprehensible Input Hypothesis, which emphasizes the importance of accessible linguistic input in language learning. Although many reading passages received relatively easy readability classifications, such scores do not establish whether the materials provide an appropriate level of challenge for individual learners. Similarly, students' perceptions of reading ease cannot be equated with demonstrated comprehension or language acquisition. Evaluating instructional suitability therefore requires attention to both the textual characteristics of reading materials and learners' reported experiences, while recognizing that neither measure directly assesses reading comprehension.

The discrepancies between formula-based readability estimates and students' perceptions also have practical implications for teachers and textbook developers. Reading passages displaying substantial differences between readability estimates and students' perceptions warrant closer examination, particularly in relation to vocabulary familiarity, topic knowledge, and discourse organization. Teachers may address students' reported difficulties by introducing unfamiliar vocabulary, activating relevant background knowledge, and providing guidance on relationships between ideas in a text. Textbook developers may also consider both the progression of estimated linguistic complexity across instructional levels and students' evaluations of individual passages when reviewing the selection and sequencing of reading materials. Such an approach would allow readability assessment to inform instructional decisions without treating formula-based scores as definitive indicators of students' reading abilities.

Taken together, the findings indicate that FRE, FKGL, and students' perceptions contribute distinct but complementary information to textbook readability evaluation. While the formulas identify selected structural characteristics of reading passages, students' responses provide contextual information about their perceived difficulty. Integrating these perspectives enables a more nuanced assessment of reading materials

used in Gontor-affiliated pesantren, particularly in identifying passages that may require further pedagogical consideration.

IV. Conclusion

This study investigated the readability progression of 18 reading passages across three Gontor Press English Lesson textbooks by examining Flesch Reading Ease (FRE), Flesch–Kincaid Grade Level (FKGL), and students' perceptions. The FRE findings indicated that most passages were classified as easy or very easy, while FKGL scores revealed variations in estimated linguistic complexity within and across textbook levels. Although the textbooks were intended for successive instructional stages, the distribution of readability scores did not demonstrate a consistently increasing pattern of difficulty. This finding suggests that the progression of textual complexity requires consideration beyond the instructional level assigned to each textbook.

The comparison between formula-based readability estimates and students' perceptions revealed both similarities and discrepancies. While some passages showed correspondence between estimated linguistic complexity and perceived difficulty, others received contrasting evaluations. For instance, Animal, which had the highest estimated linguistic complexity in English Lesson 3, was identified as the easiest passage by the largest proportion of students. Students' reported reasons further highlighted the relevance of topic familiarity, vocabulary knowledge, and perceived sentence complexity in their judgments. These findings indicate that readability formulas and students' perceptions represent different dimensions of reading difficulty and should not be interpreted as interchangeable measures. The contribution of this study lies in demonstrating how their integration provides a more context-sensitive understanding of textbook readability within the pesantren EFL learning environment.

The findings have practical implications for teachers and textbook developers. In addition to examining readability scores, textbook evaluation should consider the sequencing of reading passages and students' reported difficulties, particularly when formula-based estimates and learners' perceptions differ. Teachers may provide additional vocabulary support, activate relevant background knowledge, and guide students in understanding relationships between ideas, while textbook developers may review the distribution of linguistic complexity across instructional levels to support more systematic progression. Nevertheless, this study was limited to 18 reading passages from three textbooks and relied on formula-based estimates and self-reported perceptions without directly assessing students' reading comprehension or language proficiency. Future research could incorporate comprehension assessments, classroom observations, and broader textbook samples to examine how estimated textual complexity relates to learners' actual reading performance.

V. **Bibliography**

Adhikari, G. P. (2021). Calculating the Sample Size in Quantitative Studies. Scholars' Journal, 14–29. https://doi.org/10.3126/scholars.v4i1.42458

Aziez, F. (2018). An analysis of interlanguage performed by students of an Islamic boarding school in Tasikmalaya. ELT Perspective: Jurnal Pendidikan Bahasa Inggris, 4(2). https://doi.org/10.33603/perspective.v4i2.1543

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. TESOL Quarterly, 42(3), 475–493. https://doi.org/10.1002/j.1545-7249.2008.tb00142.x

Darwin, K. P. P. (2022). Readability Level of Reading Texts in The English Textbook Published by Intan Pariwara for Junior High School Grade IX at Junior High School 2 Grogol Kediri. Edulink : Education and Linguistics Knowledge Journal, 4(1), 62. https://doi.org/10.32503/edulink.v4i1.2383

Dörnyei, Z., & Taguchi, T. (2009). Questionnaires in second language research: Construction, administration, and processing (2nd ed.). Routledge. https://doi.org/10.4324/9780203864739

DuBay, W. H. (2008). "Unlocking Language: The Classic Readability Studies," in IEEE Transactions on Professional Communication, vol. 51, no. 4, pp. 416-417, Dec. 2008 https://doi.org/10.1109/tpc.2008.2007872

Ehara, Y. (2021). Readability analysis and linguistic complexity in English educational texts. Language Testing in Asia, 11(3), 1–15. https://doi.org/10.1186/s40468-021-00125-7

Hakim, A. A., Setyaningsih, E., & Cahyaningrum, D. (2021). Examining the Readability Level of Reading Texts in English Textbook for Indonesian Senior High School. Journal of English Language Studies, 6(1), 18. https://doi.org/10.30870/jels.v6i1.8898

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), 166–172. https://doi.org/10.7575/aiac.alls.v.9n.2p.166

Harmer, J. (2001). Book Review: The Practice of English Language Teaching. RELC Journal, 32(1), 135–136. https://doi.org/10.1177/003368820103200109

Khodadady, E. (2017). Evaluating two high-intermediate EFL and ESL textbooks: A comparative study based on readability indices. Sociology International Journal, 1(3). https://doi.org/10.15406/sij.2017.01.00016

Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel (Research Branch Report No. 8-75). Naval Technical Training Command. https://doi.org/10.21236/ADA006655

Krashen, S. D. (2008). The comprehension hypothesis extended. In T. Piske & M. Young-Scholten (Eds.), Input matters in SLA (pp. 81–94). Multilingual Matters. https://doi.org/10.21832/9781847691118-007

Libhaber, E. (2025). A simple guide to analyse data: Descriptive statistics in quantitative research. SA Heart Journal, 22(3), 188–190. https://doi.org/10.24170/22-03-7664

Mohammadi, M., & Abdi, H. (2014). Textbook Evaluation: A Case Study. Procedia - Social and Behavioral Sciences, 98, 1148–1155. https://doi.org/10.1016/j.sbspro.2014.03.528

Muhith, A. (2018). Evaluating the quality of three English textbooks published by Gontor Press: A content analysis on EFL textbooks for Islamic boarding school of Daarul Falah Serang Banten (Master's thesis, Syarif Hidayatullah State Islamic University Jakarta). Institutional Repository UIN Syarif Hidayatullah Jakarta.

Mukundan, J., Hajimohammadi, R., & Nimehchisalem, V. (2011). Developing An English Language Textbook Evaluation Checklist. Contemporary Issues in Education Research (CIER), 4(6), 21. https://doi.org/10.19030/cier.v4i6.4383

Nguyen, T. T., Baker, J. R., & Le, T. Q. (2024). A partial least squares structural equation modeling exploration of EFL learners' perceptions of what contributes to the readability of model paragraphs. SAGE Open, 14(2). https://doi.org/10.1177/21582440231211802

Ningrum, A. S. B., Lestari, U. M., & Murti, N. R. (2023). Content analysis on two English textbooks for the eighth grader of junior high school. Magister Scientiae, 51(1), 47–62. https://doi.org/10.33508/mgs.v51i1.4413

Nuryani, Fitriyah Z. A., M., & Nurjanah, N. (2018). Representation of Islamic Culture in Textbooks at Islamic Boarding Schools: Study on Textbooks at Pondok Modern Darussalam Gontor. Proceedings of the 1st International Conference on Recent Innovations, 1386–1393. https://doi.org/10.5220/0009928213861393

Permatasari, M., Durriyah, T. L., Eviyuliwati, I., & Defianty, M. (2022). The readability level of English textbooks published by Kemendikbud for grade XI. In Proceedings of ICRI 2021 (pp. 31–36). https://doi.org/10.5220/0011240200003376

Ponnudurai, G., Caszo, B., & Gnanou, J. (2025). Readability analysis as a tool for evaluating English proficiency in first-year medical students. BMC Medical Education, 25(1). https://doi.org/10.1186/s12909-025-07348-8

Porter, P. A., McDonough, J., & Shaw, C. (1995). Materials and Methods in ELT: A Teacher's Guide. The Modern Language Journal, 79(2), 279. https://doi.org/10.2307/329642

Pulukadang, W. T., Uno, H. B., Panal, H., & Panjaitan, K. (2020). Integrated learning module development on PGSD students at Gorontalo State University, Indonesia. International Journal of Advanced Engineering, Management and Science, 6(7), 347–355. https://doi.org/10.22161/ijaems.67.7

Richards, J. C. (2001). Curriculum Development in Language Teaching. https://doi.org/10.1017/cbo9780511667220

Rohmah, N., & Roqib, M. (2024). Integration in Modern Islamic Boarding Schools; Kulliyatul Mu'allimin Al-Islamiyyah (KMI) Gontor Curriculum. Transformasi Manageria: Journal of Islamic Education Management, 4(2). https://doi.org/10.47467/manageria.v4i2.5965

Saputri, A. P. (2025). Exploring Modern Islamic Boarding School English Textbooks Based on the SLA Principles. Journal of English Language Teaching and Cultural Studies, 8(1), 55–66. https://doi.org/10.48181/jelts.v8i1.29637

Tomlinson, B. (2012). Materials development for language learning and teaching. Language Teaching, 45(2), 143–179. https://doi.org/10.1017/s0261444811000528

Trumbull, E. (1984). Stages of reading development. Jeanne Chall. New York: McGraw-Hill, 1983. Pp. 293. Applied Psycholinguistics, 5(3), 285–292. https://doi.org/10.1017/s0142716400005166

Zainurrahman, Yusuf, F. N., & Sukyadi, D. (2024). Text readability: Its impact on reading comprehension and reading time. Journal of Education and Learning, 18(4), 1422–1432. https://doi.org/10.11591/edulearn.v18i4.21724.