

READABILITY OF PRE-TERTIARY ENGLISH COURSE BOOKS IN GHANA: UNPACKING THE METRIC-GRADER ASSESSMENT INTERFACE

Justine Bakuuro¹, Africanus Diedong², Damasus Tuurosong³

¹justinebakuuro16@gmail.com, ²langyin45@gmail.com, ³dtuurosong@gmail.com

^{1,2,3}Simon Diedong Dombo University of Business and Integrated Development Studies, Wa, Ghana

Received: 18th January 2024

Revised: 30th April 2024

Accepted: 28th May 2024

ABSTRACT Readability of texts is generally determined by the application of readability metrics. In most cases, the learner (grader) is not directly involved in assessing the readability of texts assigned to his/her grade. This study sought to determine the extent of reconcilability of metric and grader readability assessment values on the same texts. The textual data was culled from the *Global Series* English textbook series used in Ghanaian Senior High Schools for the teaching of the English language. Applying the mixed methods research, primary data was also collected from a total of 150 graders across the 3 grades of SHS in Ghana. Using a descriptive research design, within the constructivist research paradigm, the study reveals that metric and grader readability assessments do not agree to a very large extent. Only two percent (2%) of the texts were found to be suitable for their intended grades at the senior high school. The grader readability assessment of texts was found to be at variance with the metric readability assessment. Whilst most graders found most texts to be readable (over 96%), metric readability assessment found the same texts to be unreadable (about 98%). Both metric and grader readability assessments need to be conducted on the same texts to conclude properly on their realistic readability status before assigning them to respective grades. This is a result of the wide metric-grader readability assessment disparity. Where the grader readability assessment reveals a text to be readable and the metric readability assessment finds the same text to be unreadable, the grader assessment should be considered as that is a more realistic assessment.

Keywords: Readability, Pre-Tertiary English Course Books, Metric Readability Assessment, Grader Readability Assessment, Metric-Grader Assessment

INTRODUCTION

According to readability theory, a text's readability is determined by how simple it is to read and comprehend. It is the "easy degree" to which a text can be understood, according to Gunning (1952). In other words, readability describes how natural it is to relate to written material. Gunning (1952) is largely recognized for having invented the Gunning Fog Index (GFI), which is responsible for this idea. This idea counts "difficult" words, adds them to the average amount of words in a sentence, and multiplies the result by 0.4. This indicates the text's level of readability. This process is used to analyze certain data sets for this investigation.

Halliday (1985b) cites lexical density as a vital consideration of the difficulty of the documentary language. Eggins (2004) identifies grammatical intricacy as one major factor in text complexity. Gunning (1952) believes that readability is yet another major factor in determining text complexity. As already stated, Halliday (1985) has recognized lexical density as a significant factor in the complexity of written language. This relates to the notion of lexico-grammar in Halliday's Systemic Functional Linguistic Theory.

On this dimension of written language, Wiredu (2012) opines that editorial language is complex and that there was overwhelming reliance on complex sentences in the study he conducted using newspaper texts as data. He added that the complexity arises from among other things, the dominance of nominal words or processes. He further states that argumentative editorials tend to attempt to lure readers to reason in similar ways as the paper reasons or argues. Again, Wiredu (2016) still commenting on written language, argues that pidgin language is perceived as complex and incomprehensible. His study finds this claim as a mere perception. Agreeing with Wiredu (2012) on the complex nature of written language, Frimpong (2017) observes that because writing a text requires extensive preparation and revision, written languages typically have more intricate linguistic structures. As a result of this intricacy, understanding calls for close reading, according to him.

The aforementioned authorities corroborate the propositions of Thorndike (1921), a founding scholar on readability, who in that seminal study, opines that course books are a

valuable source of information for teachers in the classroom. Therefore, one of the most essential considerations when selecting course books is text difficulty, and one useful metric for gauging text difficulty is text readability. To that end, he underscored the fact that a readability study anchored on LD is a more formidable study than a mere direct application of readability metrics.

From the foregoing discussion on text complexity, it is abundantly clear that indeed, written language is complex, according to Wiredu (2012) and Frimpong (2017), and this is largely attributable to the choice and use of the lexicon. More specifically, lexical density and text readability are identified to be major factors of text complexity. This study has closely examined the complexity/difficulty levels of texts utilized in the English senior high school textbooks in Ghana and their suitability to the intended grades. Readability metrics are applied to sampled texts to determine text-grade suitability on one hand. On the other hand, grader assessment of readability is done, to make room for comparison with the metric assessment.

This study's underlying presumption is that the readability assessment of texts using metrics should generally reconcile with the readability assessment made by learners as in their comprehension of texts assigned to them. Grader readability assessment can chiefly only be done by testing their degree of understanding of a given text. This is because there is obviously an intrinsic nexus between readability and text comprehension. That is, for us to conclude that a text is readable, it must, as a consequence, be comprehensible. Thus, this research aims to examine the nature of that link, by comparing readability values produced by applying metrics to texts and readability values produced by assessing learners' general comprehension of the same texts. More specifically, the research objectives are to investigate the metric readability assessment values of texts in senior high school English course books in Ghana; the grader readability assessment values of texts in pre-tertiary English textbooks in Ghana; how metric and grader readability assessment values reconcile in their application to same texts.

LITERATURE REVIEW

Theoretical Framework

This study is adequately underpinned by two primary theories; Reader-Response Theory and Schema theory. The reader-response theory emphasizes the reader's active role in constructing textual meaning and highlights the diversity of reader reactions. Just like the schema theory, this theory applies to research question three (3) which centers on metric and grader readability assessment. Readability is influenced by several factors as indicated in the literature in this study. To that end, the strict adherence to readability metrics to determine text-grade suitability is problematic. How the individual grader responds to a text during reading varies depending on their varied peculiar characteristics and privileges. For instance, whilst graders may be in the same grade if grader **A** lives with a highly educated family in a predominantly elite environment, grader **A** is most likely to find a text more readable than grader **B**, who lives with an illiterate family in a predominantly non-elite environment. Their levels of active engagement with the text would definitely not be the same. To that extent, one may understand why the metric and grader readability assessment values in this study are in sharp disparity.

The Schema theory, within the realm of cognitive science, delves into how the brain organizes knowledge. A schema, constituting an organized knowledge unit for a subject or event, draws from past experiences and is accessed to guide current understanding or actions. Graders' ability to read and process meaning from texts depends also on their past experiences and how their brains contextualize new knowledge within existing knowledge (schemata). In the analysis of this study, the shocking revelation of metric and grader readability assessment being in sharp disparity may stem from this assumption of the schema theory. Graders' experiences or existing knowledge may have aided their effort to read and understand the sampled texts. Therefore, whilst the metric assessment bluntly uses syllables, content words,

grammatical words, ranking clauses, and sentences among other readability variables to arrive at readability values, graders' schemata or background knowledge may override these abstract considerations.

Text(book) Readability

Fata et al (2022) attempt to evaluate the reading content in two distinct English textbooks that were released by various publishers in terms of readability. Several investigations have examined the reading materials found in English textbooks, and the results have shown that the reading materials weren't good enough for the students' levels. The study analyzed, interpreted, and described the data using a descriptive content analysis method. As a tool, the Flesch Reading Ease formula was used. The research's data came from the English textbooks "Bahasa Inggris" by The Ministry of Education and Culture and "Pathway to English" by Erlangga. The conclusions of the study show that five of the fifteen reading books correspond with the first textbook's students' level (a fairly difficult level) to three texts out of thirteen, whereas the second textbook's students' level (a fairly difficult level) corresponds to three texts out of thirteen. Additionally, the texts in the first textbook are classified as "standard" by the average readability score, whereas those of the second textbook were classified as "fairly complex." Based on the study's findings, educators are advised to apply the readability formulas to assign pupils to suitable reading levels and to do additional research on readability in a broader context.

The following readability tools were utilized by Kim et al. (2019): Gunning Fog Index, Flesch-Kincaid Grade Level, Coleman-Liau Index, Automated Readability Index, and Simple Measure of Gobbledygook Index. By calculating the mean of these five indices, an average reading grade level (AGL) was determined. Findings revealed that the whole texts of the hundred most cited papers appear more readable than their respective abstracts. This according to the study, suggested that abstracts are generally more coded with technical language in scientific writing than their full texts. While articles published in journals with greater impact factors were found to be less legible, experimental research and technique papers were also shown to be more readable than reviews or meta-analyses. That study applied various readability metrics on the same texts and was also interested in comparing the readability values of full texts and their excerpts (abstracts). The current study used whole texts and compared readability values using different readability formulae much the same way as Kim et al. (2019).

In 2020, Hidayatillah and Zainil conducted a study on the readability of an Indonesian course textbook on semantics and pragmatics. Students offering this course in Indonesia complained of poor understanding of texts in the textbook, resulting in their poor performance in the course. The study therefore sought to probe the relationship between the students' low test scores and the textbook texts' readability. Using three pages from each of the three textbook chapters (Chapters 9, 10, and 11) as well as questionnaires, interviews, and online readability instruments, the researchers measured the readability level of the textbooks, the factors that influenced readability levels, and the best way to use the textbooks. The study concluded that the textbook's readability level was, in fact, too high (73.09), making it difficult for students to read the text. Unfamiliar or complex vocabulary and lack of reading interest were identified as the leading factors influencing poor readability by students. The main aim of that study and this one are the same—determining text-grade suitability.

O'Sullivan et al. (2020) used validated and commonly used readability criteria (Flesch 1948 reading ease scale; Gunning 1952 Fog Index), which offer a comprehensive assessment of written communication, to quantify the degree of difficulty of clinical research Patient Information Leaflets/Informed Consent Forms (PILs/ICFs). Additionally, they sought to compare their results with recommendations for best practices. Fifty-four of the 176 PILs that were gathered could be evaluated. The American Medical Association recommends a mean reading age of 12 years or older, but none of the PILs/ICFs met this requirement. Of PILs/ICFs,

7.1 percent were deemed to have "plain English," 40.3% were deemed to be fairly tough, 51.3 percent were deemed to be challenging, and 1.3% were deemed to be extremely difficult. Only two PILs/ICFs matched all six best practice literacy indicators, and none of them had a CCI of less than 90. Examining the PILs/ICFs in the study using both health literacy-based methods and conventional readability criteria revealed that they were unduly complex. Compliance with guidelines was also found to be poor on the part of literacy agencies. This study relates to the current study in its application of more than one readability formula to the same texts. This provides for formula comparability.

Slippe & Gyasi (2019) investigated the English language textbooks' readability for University of Cape Coast diploma students at the Center for Distance Learning. For the study, three textbooks for year one through year three diploma students were used. For readability studies, selected passages from each textbook were used. Using the Flesch – Kincaid grade level (FKGL) and Flesch reading ease (FRE) indexes, the textbook readability scores were determined. The data were analyzed using bootstrapping, one-sample T-tests, and measures of central trends. According to the results, the three textbooks fell between the categories of "fairly difficult" or "difficult" to understand ($m = 52.73$ and 40.69). It was shown that there is a statistical difference between the textbooks' readability scores and the recommended ratings for publications that are publicly accessible. Additionally, the results suggest that for all three textbooks, the average word count per sentence should be between 22 and 25. This might have had a part in the textbooks' poor readability. The researcher suggests that the textbooks be revised for them to fulfill their intended function and support efficient English language instruction and learning.

Hakim et al (2021) make an effort to use the Coh-Metrix to evaluate an English textbook's reading level. Since a textbook is the object of analysis, content analysis is used. Through the development of a relationship between the Coh-Metrix test result and the students' perceived readability, this study additionally builds on earlier research on readability metrics. Using cohmetrix.com, an automated Coh-Metrix readability assessment was carried out to collect the data. Additionally, 35 pupils in the tenth grade were given a questionnaire on their perceived readability to obtain a deeper study. Next, Miles and Huberman's interactive model was used to analyze the gathered data. Despite the small discrepancy between the Coh-Metrix-generated reading level and the perceived readability level, most of the texts in the book are still too difficult for the pupils to understand, according to the study's findings. The textbook offers intelligible input; thus, it may have some promise for language learning.

However, the goal of Hidayat's (2016) study was to find out how readable a textbook was for senior high school students in Indonesia in Grade XI, Semester 1. The study's research topic was, "What is the readability level of the reading passages in the English textbook?" This book was the first English textbook used under the Kurikulum 2013 and the newest at the time. The Flesch Reading Ease Formula was utilized to examine the reading texts included in the research. The formula was used to determine the readability level. The results indicate that the book, with a score of 69.392, was at the Standard level. Stated differently, the reading materials were largely appropriate for kids in the eleventh grade.

The main objective of Kolahi & Shirvani's (2012) study was to evaluate the readability levels of Persian and English translation textbooks using the Gunning Fog Index. To this end, five translation textbooks—one in Persian and one in English—were included in the study's corpus. Since there aren't many textbooks available in Persian, these ones were chosen. A total of 284 example texts were chosen at random and assessed for readability, comprising 142 translations from English translation textbooks and another 142 translations of the matching English sample texts. The study's conclusions showed that whereas the average Fog Index of textbooks translated into English was 16.4, the average Fog Index of those translated into Persian was 20.1. This suggests that Persian-translated textbooks have a reading level that is 3.7-grade levels greater than that of their English equivalents. When the Fog Index increases,

the writing gets harder to read. The study's findings showed that textbook translations into Persian are not as readable as those into English.

For Swedish school years 7-9, Langeborg (2010) set out to assess the readability of texts from four English textbook series and determine if text difficulty increases within the series with each given difficulty level and grade level. The study also attempted to evaluate various textbook series to determine whether they are similar in terms of average readability. The readability formulas Flesch Reading Ease and Flesch-Kincaid in Microsoft Word 2007 were used to determine the readability grade levels of a sample of 231 texts from the series Good Stuff, Happy, Time, and Wings. According to the results of both algorithms, texts often get harder with each passing school year and level. The study did reveal, however, that there is a wide range in text complexity and that there are variations in average difficulty levels among the four series.

Maryansyah (2016) examined the readability of texts used to teach reading to pupils in the ninth grade at an Islamic secondary school that is owned by the state. The study employed a quantitative approach and was descriptive in nature. Its goal was to find out how readable the texts were. The 63 texts formed the data for that study. The Fry readability formula (graph) was the tool utilized to gather data. The data were interpreted using the percentage technique. MTsN 2 Kota Bengkulu grade IX students were given 63 texts; of them, 54% were easy, 27% were difficult, 10% were invalid, and 9% were appropriate. Considering the results of the study, the following suggestions have been made: It is advised that readability analysis be applied to texts by English teachers at MTsN 2 Kota Bengkulu before using them to teach reading, that writers of English textbooks intended for grade IX students be aware of the readability of the texts they incorporated into their textbooks, and that other researchers conduct follow-up research on the findings of this study using alternative readability techniques.

RESEARCH METHODS

Research Design

This study deploys the mixed methods approach to research, embodying a descriptive qualitative content analysis with the use of numerical data. A research technique called content analysis is used to find specific words, themes, or concepts in a given set of qualitative data, or text. The design of this study is analytical in nature; therefore, one may describe the study as having a descriptive or analytical design. This research is partly qualitative content analyses using a descriptive mode of analyzing data and adopting quantitative procedures in the course of data analysis. The heavy use of numerical data, coupled with the qualitative content analysis using the descriptive mode, lends this study to the mixed methods research approach. The relevance of applying the descriptive research design in this study lies in the fact that it has allowed for an effective analysis of the numerical data, making it possible to ascertain the relationship between LD and readability, the suitability of texts to intended grades and the agreeability of readability and LD metrics in their application to same texts, among other objectives. Therefore, this study employs quantitative tools for the presentation of data but it uses the mixed methods approach to research. The study does not seek to extrapolate but to understand the phenomenon under study—text readability. This is done with the aid of descriptive statistical data, giving a fair balance between qualitative and quantitative research (mixed methods research).

Sampling and Sample Size

A total of fifty (50) graders from each academic grade (SHS1-3) constituting an overall total of one hundred and fifty (150) graders across the three (3) academic grades, formed the research participants for this segment of the study. The fifty (50) graders for each grade were selected from five (5) different schools. Ten (10) graders were randomly selected from each of the five (5) separate schools. In the random selection procedure, the total number of graders at each grade was taken and divided by ten (10). Then the resultant figure was used to count

graders and pick only those on whom the lot fell. This was done from class to class and the ten graders were eventually selected for each grade from the five schools.

Data Analysis

In this study, to display the qualitative data, the researcher employed quantitative methods. This was carried out to facilitate data analysis and help the researcher make sense of the corpus data. The textual analysis plan is applied in this study, as previously mentioned. The texts utilized as data in this study are taken from the Global Series (Nelson, 2016) English textbook series. Ghanaian SHS students use the English textbook Global Series. The Queen's language has been taught in many Ghanaian schools for many years thanks to its endorsement and approval by the Ghana Education Service. Using a Microsoft Word processor, the texts were retyped from the textbook series. The texts were painstakingly revised and retyped to match the originals in the textbook. After that, readability metrics and the four (4) LDs were used to examine them. To convert the text data into numerical data, Textalyser, Microsoft Word, Microsoft Excel, and Lexicool were the tools utilized. Words, syllables, sentences, clauses, phrases, and letters were all counted using these instruments. The examined text data was thoroughly cross-checked by the researcher to guarantee the reliability, validity, and accuracy of the final results.

Ethical Considerations

A collection of guidelines that direct the design and procedures of research are known as ethical considerations in research. Voluntary involvement, informed consent, anonymity, secrecy, risk of damage, and results communication are some of these guiding concepts. Thus, the set of guidelines that direct the overall technique and research design in this study is covered in this segment. In order to safeguard study subjects' rights, improve the validity of their findings, and uphold the general integrity of science, among other reasons, researchers are required to do their work in accordance with ethical norms. The department provided an official letter of introduction, which the researchers used for their study. He then emailed the author of Global Series to request permission to utilize the book. Before the researcher used the texts for the analysis, the author duly gave permission. The selection of graders as research participants was done seamlessly without any shred of discrimination or unfair treatment in whatever form.

Trustworthiness of Data

One strategy used by researchers to convince readers and themselves that their work merits attention is trustworthiness (Lincoln & Guba, 1985). By adding the criteria of credibility, transferability, dependability, and confirmability to the traditional quantitative assessment criteria of validity and reliability, Lincoln and Guba (1985) improved the idea of trustworthiness. This research employs a strict textual analysis grounded in established theoretical and philosophical frameworks. The human aspect is completely missing from this study, with the exception of potential human error in the computational analysis, about which the researcher is extremely cautious. The analytical methodological techniques utilized in this work are well-researched, tried-and-true methods in the field. Their suitability and dependability as research tools or methods in doing textual analysis (readability analysis) are highly commendable.

The sampled texts included in this study's data came from a popular English textbook series that is recommended for senior high school use in Ghana by the Ghana Education Service. This verifies the accuracy of the data that was analyzed. Conclusions and findings are therefore valid, and the validity and reliability of results are guaranteed. Data inter-rater reliability was examined. According to Creswell (2014), it is one of the verification tools used to assess the reliability and validity of study findings. To help with the data trustworthiness assessment for this study, the researchers employed the services of two (2) research assistants from the university. To be more precise, these research assistants and the researchers independently verified the number of syllables in the words used in the texts, the number of

words in each text, the number of ranking phrases per clause, the number of ranking clauses per sentence, and the number of sentences per text. Additionally, the research assistants helped with the independent confirmation of the textual analysis technique used in the study. Before the researchers used the data, the findings of each of the three verifiers—two research assistants and the four researchers themselves—were compared and validated.

Instruments

Software adopted

In this study, Microsoft Word and Excel were employed to deal with some counting and abstract statistics. First, Microsoft Word was used to calculate the total words in each text and to show the readability statistics realized by the various formulae. Later, Microsoft Excel was employed to analyze the data and determine the correlation between measurements. Other key online tools used for the analysis are discussed below.

Online Text Analyzer (Lexicool)

This text analysis tool offers statistics on word frequency and character counts in addition to information on a text's difficulty. With the help of this application, you may count the amount of words, phrases, and characters in a text to examine it. Additionally, it offers statistical data regarding the frequency of terms and phrases. It's simple to use this word counting and text analyzer online. The way this software functions is as follows: Choose the language for optimization (English, Spanish, French, or Italian), copy and paste the text, or type it into the input box, and click "Go." The word repetition analysis is subjected to a filter that excludes specific short "irrelevant" terms if a language for optimization is chosen. It is preferable to choose "none" if the language you are using is not available in the drop-down list. With this analytic tool, texts in any language can be evaluated. Teachers can use this tool to gauge the amount of difficulty and ability needed to comprehend a book. It is also meant for educators. Additionally, they can quickly extract a study list of terminology.

Syllable Counter (SyllableCounter.net)

A straightforward and cost-free web application called Syllable Counter can be used to determine how many syllables are in a phrase or sentence. This is helpful for syllable checking when writing. It is a teaching and learning aid for English syllables and grammar. The way this approach works is as follows: A basic algorithm is employed to determine the overall count of syllables. To see the number of syllables in each line, you would enter the sentences line by line. It should be made very clear that the purpose of this online tool is merely to verify that the number of syllables in words and sentences as a whole matches the Merriam-Webster dictionary's syllable count. The Merriam-Webster dictionary is used when there is a disagreement between the two regarding syllable identification and number. With the aid of this instrument, researchers can satisfy the requirements listed in Flesch's (1948) Readability Formula: $84.6 \times ASW - (1.015 \times ASL) = 206.835$ is the FRE.

ASL = Average sentence length in ASL

ASW = Average syllable count per word

The Merriam-Webster.com Dictionary

This was the instrument used in the research to tally the number of syllables in words. This software operates as follows: A word is searched using the Merriam-Webster dictionary and the breakdown of syllables in the word is displayed alongside its right pronunciation and meaning among other linguistic features. The Online Text Analyzer tools are used in calculating the total number of words, syllables, content words, grammatical words, clauses, and sentences. Findings are presented according to each research question. That is, the theme of each research question is the prime focus of the analysis and appropriate findings from the analyses have been identified as answers to research questions. To this end, the results as revealed by the analysis of data have been carefully discussed to properly contextualize the findings in respect of each research question in this paper.

FINDINGS AND DISCUSSION

Metric and Grader Readability Assessments

The comparative metric-grader readability assessment values are shown side in Table 1.

Table 1: Metric and Grader Readability Assessment Values

	Text 1		Text 2		Text 3		Text 4		Text 5		Averages	
	Gunning Vs Graders	Flesch Vs Graders	Gunning Vs Graders	Flesch Vs Graders	Gunning Vs Graders	Flesch Vs Graders	Gunning Vs Graders	Flesch Vs Graders	Gunning Vs Graders	Flesch Vs Graders	Gunning Vs Graders	Flesch Vs Graders
Narrative texts-SHS1	12/16	77/23	08/38	43/28	14/32	38/39	24/41	68/37	34/35	46/26	18/32	54/31
Narrative texts-SHS2	09/39	77/44	18/18	51/34	18/26	41/33	21/34	71/38	31/35	51/28	19/30	58/35
Narrative texts-SHS3	10/33	80/34	15/31	44/38	21/33	36/21	18/37	72/40	30/47	53/37	19/36	57/34
Descriptive texts-SHS 1	09/35	83/43	19/32	34/46	23/21	35/35	19/18	81/30	29/36	64/49	20/28	59/41
Descriptive texts-SHS 2	11/44	77/39	21/25	46/36	18/36	68/24	25/30	76/43	28/36	67/32	21/34	67/35
Descriptive texts-SHS 3	14/26	68/48	20/31	41/34	15/34	63/35	28/39	72/41	17/34	71/23	19/33	63/36
Expository texts-SHS 1	15/34	80/37	14/37	56/40	18/39	71/39	30/20	69/41	11/29	69/36	18/32	69/39
Expository texts-SHS 2	14/36	71/39	16/41	47/37	18/41	77/48	18/40	68/39	09/29	64/42	15/37	65/41
Expository texts-SHS 3	10/38	77/42	13/37	40/39	20/27	65/36	27/45	67/35	14/38	60/44	17/37	62/39

From Table 1, the readability values of the sampled texts for each of the five texts for each academic grade and genre have been indicated. The readability value using each metric is first written, then followed by the readability value using grader assessments. The readability values using grader assessments were the averages of the fifty (50) graders for each grade. For example, the 12/16 readability assessment label for Narrative texts-SHS1 (see first horizontal column from top of the table) means, the 12 is the readability value using the average readability assessment of the 50 graders for SHS1.

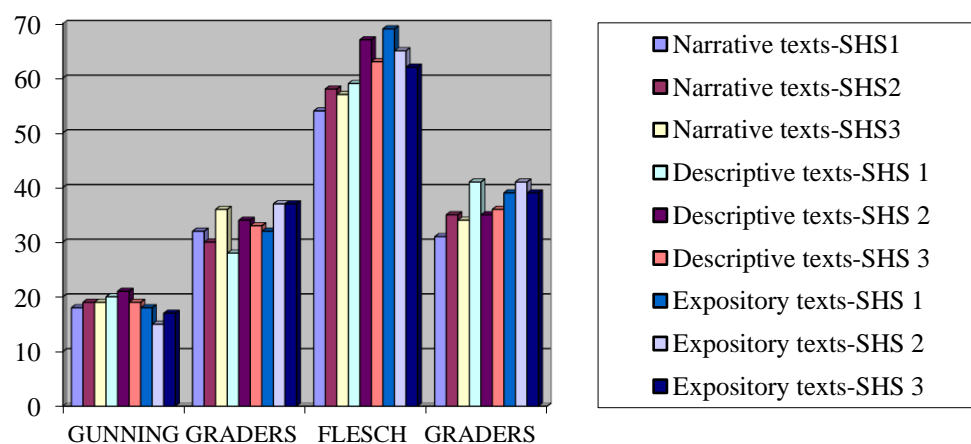
After that, overall averages of metric readability values and grader readability values are computed at the right end of the table. These overall averages are collated in Table 2 and represented graphically below the table in Figure 1.

Table 2: Average Metric and Grader Readability Assessment Values

GENRE-GRADE CATEGORY	GUNNING FORMULA	GRADERS (SHS1-3)	FLESCH FORMULA	GRADERS (SHS1-3)
Narrative texts-SHS1	18	32	54	31
Narrative texts-SHS2	19	30	58	35
Narrative texts-SHS3	19	36	57	34
Descriptive texts-SHS 1	20	28	59	41
Descriptive texts-SHS 2	21	34	67	35
Descriptive texts-SHS 3	19	33	63	36
Expository texts-SHS 1	18	32	69	39
Expository texts-SHS 2	15	37	65	41
Expository texts-SHS 3	17	37	62	39

Using the Flesch-adapted readability scale, the comparison of metric and grader readability values reveals a significant contradiction as can be seen in Table 2. Whilst some texts were assessed by graders to be “readable” or “very readable”, metric assessments found the same texts to be “unreadable” or “very unreadable”, as the case may be. This revelation in the analysis of data is very curious.

Figure 1: Average Metric and Grader Readability Assessment Values



From the foregoing analysis, metric and grader readability assessment values do not agree when applied to the forty-five (45) texts under this study. The analysis in Tables 1 and 2 clearly indicates this. There is a large discrepancy seen in metric and grader judgments of readability. This is supported by the finding that, whereas less than 5% of the sampled texts were determined to be intelligible by metric readability evaluation, over 90% of the texts were found to be readable by grader assessment. This presents some important discoveries about the evaluation of readability using metrics and graders. Certain texts were judged to be unreadable by metric readability assessments, while they were found to be readable by grader readability assessments, and vice versa. The readability values for metrics are higher than those for graders. This presupposes that the books were generally harder to read in the metric evaluations than they were in the grader assessments.

Based on the Flesch adapted readability index, this evaluation was made. When the Flesch readability measure is compared to the Grader readability assessment, the Flesch readability values are higher, suggesting that metric assessment is higher and inconsistent with grader assessment as per the objective of this study. This partially affirms the findings of Langeborg (2010) and Maryansyah (2016). When the Gunning readability measure is compared to the graders' readability judgment, the Gunning readability values are found to be lower. This creates the presumption that the texts were harder to read for the graders than they were using the Gunning readability metric. When compared to Gunning and Flesch's readability metrics, the results of grader readability assessments show that the sampled texts have an averagely high readability. When compared to readability metrics generally, this indicates a high degree of consistency in the grader readability assessment. These revelations contradict the findings of Langeborg (2010) and Maryansyah (2016) which revealed significant inconsistencies in grader readability assessments.

CONCLUSION

In sum, there is a very significant discrepancy between the readability assessments made by graders and those made using readability metrics. Whilst the graders' readability assessments generally conclude that the assigned texts were readable according to the Flesch Reading Ease Scale, metric readability assessments show that the same texts were "very unreadable". The disparity in scores reaches 40% in certain instances. To that extent, the validity and reliability of the Flesch and Gunning readability metrics are called into question. These findings imply that metric assessment of readability is not wholly reliable. To that end, it is recommended that metric assessment must always be juxtaposed with grader assessment to ascertain the reliability of results. The study is however limited by the fact that this grader

assessment was made by students from one institution. The outcome could possibly have been different with students of another institution. Therefore, a comparative grader assessment of the same texts is recommended in further studies. When metric and grader readability assessment scores conflict, the grader readability score naturally wins out.

REFERENCES

- Bakuuro, J. (2023). *Text Readability and Grade Propriety: An Analysis of Senior High School English Textbooks in Ghana*. [Unpublished Doctoral dissertation]. University of Education, Winneba
- Creswell, J.W. (2014). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches* (4th ed.). Sage.
- Eggs, S. (2004). *An introduction to systemic functional linguistics*. Continuum.
- Elkind, D. (2004). The problem with constructivism. *The Educational Forum*, 68 (4), 306–312.
- Fata, I.A., Komariah, E., & Alya, A.R. (2022). Assessment of Readability level of reading materials in Indonesia EFL Textbooks. *Lingua Cultura*, 16(1), 97-104.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. <https://doi.org/10.1037/h0057532>
- Frimpong, G.K. (2017). Subordination across Ghanaian and British newspaper editorials: A register perspective. *Ghana Journal of Linguistics*, 6(1), 75-119. <https://dx.doi.org/10.4314/gjl.v6i1.59>
- Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.
- Gyasi, W.K., & Slippe, D.P. (2019). Readability of English language textbooks for diploma students of the University of Cape Coast. *International Journal of Research*, 8(1), 107-115.
- 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), 18-35.
- Halliday, M.A.K. (1985). *Spoken and written language*. Oxford University Press.
- Hidayat, R. (2016). The readability of reading texts on the English textbook. In: *International Conference: Role of International Languages toward Global Education System* (pp. 120-128), Digital Library IAIN Palangka Raya, Central Kalimantan, Indonesia.
- Hidayatillah, N., & Zainil, Y. (2020). The readability of students' textbook used in semantic and pragmatic course in English language education program of UNP. *Journal of English Language Teaching*, 9(1), 144-159.
- Kim, C., Wang, K., & Zhang, L. (2019). Readability of 10-K reports and stock price crash risk. *Contemporary accounting research*, 36(2), 1184-1216.
- Kolahi, S. & Shirvani, E. (2012). A comparative study of the readability of English textbooks of translation and their Persian translations. *International Journal of Linguistics*, 4(4), 344-361.
- Langeborg, L. (2010). Readability: An analysis of English textbooks for Swedish school years 7-9. [Bachelor thesis]. University of Gavle Repository. <https://www.diva-portal.org/smash/record.jsf?dswid=-407&pid=diva2%3A352171>
- Lincoln, Y.S. & Guba, E.G. (1985). *Naturalistic Inquiry*. Sage Publications.
- Maryansyah, Y. (2016). An analysis on readability of English reading texts for grade IX students at MTsN 2 Kota Bengkulu. *Premise: Journal of English Education and Applied Linguistics*, 5(1), 69-88.
- Nelson, R. (2016). *Global Series: English language for senior high schools 1,2&3*. Approachers (Ghana) Ltd.
- O'Sullivan, L., Sukumar, P., Crowley, R. McAuliffe, E. & Doran, P. (2020). Readability and understandability of clinical research patient information leaflets and consent forms in

- Ireland and the UK: A retrospective quantitative analysis. *BMJ Open* 2020, 10(9), e037994 <https://doi.org/10.1136/bmjopen-2020-037994>
- Thorndike, E. L. (1921). *The teacher's word book*. Bureau of Publications, Teachers College, Columbia University.
- Wiredu, J.F. (2012). A grammar of newspaper editorial language: The complex Sentence. *Legon Journal of the Humanities*, 23, 75-124.
- Wiredu, J.F. (2016). The dependent clause in Ghanaian English pidgin. *Research on Humanities and Social Sciences*, 6(11), 50-63