IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Comparative Study of Readability Tools for EFL Assessments

[1]Golda Cherian, [2]Dr. Sunil Kumar Jha

[1]PhD Scholar, Banasthali Vidyapith University, Rajastan & English Lecturer, Sultan Qaboos University, Muscat,

[2] Assistant Professor, Department of English & Modern European Languages, Banasthali Vidyapith University

Abstract: Creating appropriate reading and listening materials is crucial in English as a Foreign Language (EFL) education, particularly for assessment purposes. Readability tools are a potential solution to know text difficulty by providing the quantitative measures of text complexity. This study compares the efficacy of six widely-used readability tools—Flesch Reading Ease, Flesch-Kincaid Grade Level, Gunning Fog Index, SMOG Index, Coleman-Liau Index, and the Automated Readability Index—in assessing the readability of texts designed for EFL assessments. A set of exam texts from a higher education institution in Oman was analyzed using these tools. Results were compared to identify discrepancies and determine the reliability of each tool. Findings show some variations among the tools' readability scores, highlighting the challenges of relying on a single tool for accurate readability assessment in EFL contexts. The study provides practical recommendations for educators and test developers to ensure the appropriateness of reading materials based on multiple readability metrics, enhancing the alignment between text difficulty and learners' capabilities.

Keywords: readability tools, EFL assessment, text difficulty, language education, Flesch-Kincaid, Gunning Fog Index, SMOG Index, Coleman-Liau, Automated Readability Index, test writers, material developers, higher education institutions in Oman

Introduction

According to Gunning (2003), readability plays a significant role in teaching and learning, helping to select appropriate materials for EFL learners. Readability tools evaluate the complexity of a text and assign it a numerical score. Some internationally recognized scales include Flesch Reading Ease, Gunning Fog Index, Flesch-Kincaid Grade Level, SMOG Index, Coleman-Liau Index, and the Automated Readability Index. A readability score provides insights into the reading age, word usage, number of syllables, sentence length, and overall comprehension level that is necessary to understand the content (“Readability Formulas”).

Measuring readability involves assessing how easy a text is to read by evaluating elements like complexity, familiarity, legibility and typography. Readability tools use formulas to assess sentence length, syllable density, and word familiarity as part of their calculations (“About Readability”). A readability score tells us what level of education a person needs to read a piece of text easily. In other words, the readability score indicates the grade level corresponding to the number of years of education a person has completed. Grade level in readability scores corresponds to the USA education system. Upon completing high school, the reading level approximates the 10-12 grade level in the readability system. A grade level of around 8 is apt for texts aimed at the public. This will ensure that almost 85% of the public can read and comprehend the text. Popular fictions usually score at or below a grade level of 6 (“What is readability?”).

Readability is a vital factor in producing quality content. It also impacts Google ranking, accessibility, and usability. Readability has an impact in every industry. Good readability inspires trust in the reader. For example, good readability improves commercial businesses. In government, documents with good readability reduce citizen enquiries, saving time and money. Laws in plain language communicate with people transparently and continue to influence the public with wider societal impact.

There are many advantages to optimizing the readability of texts. Texts created at the readability level of the audience are quicker to read and easier to understand.

Different Readability Formulas

Flesch Reading Ease: Rudolf Flesch was a pioneer in readability formulas. In 1946, he created a tool to calculate text readability, which became known as Flesch Reading Ease. His tool was used in both public and private sectors. The Flesch Reading Ease formula uses a ranking scale of 0-100. The higher the score, the easier a text is to read. This differs from the most readability scores, where a lower score indicates an easier text. Low scores indicate that the text is complicated to understand ("Flesch Reading Ease and the Flesch Kincaid Grade Level"). The formula he used was:
Flesch Reading Ease = 206.835 – 1.015 x (words/sentences) – 84.6 x (syllables/words).
A reading score of 60 to 70 is equivalent to a grade level of 9, so a text with this score should be understood by 13 to 15-year-olds. A score between 70 to 80 is equivalent to a school grade level of 8. This means text should be fairly easy for the average adult to read. The Flesch Reading Ease is used by marketers, research communicators and policy writers etc. to assess how easily a piece of text will be understood by readers.

The Gunning Fog index: In the 1960s, Robert Gunning developed a readability formula to improve his own writing. He called it the Gunning Fog Index, and it was the first readability formula to use an American grade level. Therefore, it was more user-friendly compared to the Flesch Reading Ease. The Gunning Fog formula generates a grade level between 0 and 20, estimating the education level required to understand the text. A Gunning Fog score of 6 indicates readability for sixth-graders. Text aimed at the public should target a grade level of around 8. Texts scoring above 17 are at a college graduate level ("The Gunning Fog Index").
The formula for Gunning Fog: 0.4 x [(words/sentences) + 100 (complex words/words)].
Complex words are those containing three or more syllables. The Gunning Fog Index is still used to ensure clarity and simplicity.

Flesch-Kincaid Grade Level: As technology advanced, readability measures also evolved. The US Navy wanted to improve the readability formulas by making tabulation automated for better test-retest reliability. In 1976, J. Peter Kincaid, at the request of the US Navy, recalculated the Flesch Reading Ease and came up with a new tool. The aim of the new tool was to help sailors read Navy training manuals faster and understand them better ("Measure reading levels with readability indexes"). The result was a new formula, known as the Flesch-Kincaid Grade Level:
Flesch-Kincaid Grade Level: (0.39 x words/sentences) + (11.8 x syllables/ words) – 15.59
The Flesch-Kincaid Grade indicates the American school grade required to comprehend the material on the page. If the score is lower, the easier the text is to read and comprehend. After rigorous validation, Flesch-Kincaid Grade became the US military standard.
A Flesch-Kincaid grade level between 8 and 10 is accessible to the public. The Flesch-Kincaid formula still remains as one of the most widely-used readability tools ("Flesch Reading Ease and the Flesch Kincaid Grade Level").

The SMOG Index: SMOG is the acronym for 'Simple Measure of Gobbledygook'. It estimates the number of years of education required for an average person to understand a text. This tool is most suitable to analyze texts containing 30 sentences or more. This was the length of text sampled when the formula was initially created. This readability tool was created by a clinical psychologist, G. Harry McLaughlin. It was first published in 1969 and is still in use today. It was a simple formula. He counted ten sentences from the beginning of the text, ten from the middle and ten from near the end, totaling 30 sentences. He then counted every word with three or more syllables in those sentences, calculated the square-root of this number and rounded it to the nearest whole number and added three to this figure.
SMOG Index: 3 + √ polysyllabic count

Although SMOG formula is very simple, it is widely used, especially in the healthcare sector. Its framework effectively measures overall comprehension. Being clear in your message is key to audience understanding. Doctors now use this tool to check the readability of their written materials for patients. This practice has reduced the number of follow-up appointments patients make with their doctors because patients are now able to comprehend the terms used ("The SMOG Index").

The Coleman-Liau Index: This readability tool was created by Coleman and Liau in 1975. They did not agree with syllabic counting techniques, finding it difficult for machine tabulation to detect syllabic nuances. Therefore, Coleman and Liau developed their new tool based on the belief that word length measured in letters is a better predictor of readability than depending on the number of syllables.
Coleman and Liau Index: (0.0588 x L) – (0.296 x S) – 15.8 where L = average number of letters per 100 words and S = average number of sentences per 100 words
In the Coleman-Liau Index, a score of 6 corresponds to the 6th grade in the US schooling system. A grade level of around 8-10 is appropriate for the public. The Coleman-Liau Index is widely used in schools and for medical documents ("The Coleman Liau Readability Index").

The Automated Readability Index: The Automated Readability Index (ARI) assesses the U.S. grade level required to read and comprehend a piece of text. Rather than counting syllables, it counts characters. The more characters a word has, the more difficult it is considered. This feature makes the ARI formula distinct from others. The ARI was designed in 1967 for military use. It was used for real-time evaluation of readability on typewriters. The counting method was automated. The readability scores it produced are similar to the results of other tools, demonstrating its accuracy.
The Automated Readability Index = 4.71 (characters/words) + 0.5 (words/sentences) – 21.43
An Index Tabulator attached to a typewriter for real-time tabulation, counts the number of keystrokes, words and sentences in any passage.

The ARI is well-suited to technical writing due to its speed and efficiency, making it particularly useful for creating trainee resources ("The Automated Readability Index").

Automated Readability Index		**Coleman-Liau Index & Flesch-Kincaid Grade**		**Gunning Fog**		**SMOG Index**		**Flesch Reading Ease**	
Grade Level	**Score**	**Grade Level**	**Score**	**Grade Level**	**Score**	**Grade Level**	**Score**	**Score**	**Level**
Kindergarten	≤_1	6th Grade	≤_6	6th Grade	≤_6	6th Grade	≤_6	90-100	Very easy (5th grade or lower)
1st Grade	2	7th Grade	7	7th Grade	7	7th Grade	7	80-90	Easy (6th grade)
2nd Grade	3	8th Grade	8	8th Grade	8	8th Grade	8	70-80	Fairly easy (7th grade)
3rd Grade	4	High School Freshman	9	High School Freshman	9	High School Freshman	9	60-70	Plain English/Standard
4th Grade	5	High School Sophomo	10	High School Sophomo	10	High School Sophomo	10	50-60	Fairly difficult (10th to 12th
5th Grade	6	High School Junior	11	High School Junior	11	High School Junior	11	30-50	Difficult (College)
6th Grade	7	High School Senior	12	High School Senior	12	High School Senior	12	0-30	Very difficult (College Graduate)
7th Grade	8	College Freshman	13	College Freshman	13	College Freshman	13		
8th Grade	9	College Sophomo	14	College Sophomo	14	College Sophomo	14+		
9th Grade	10			College Junior	15				

10Th Grade	11			College Senior	16				
11th Grade	12			College Graduate	17+				
12Th Grade	13								
College	14								

Table 1: A comparison of Readability Scores ("Readability Tests")

A lower Flesch Reading Ease readability score indicates a higher level of readability difficulty, whereas a lower score in readability tools means that the text is easier to understand. For example, texts scoring under 50 with the Flesch Reading Ease are considered 'difficult' or 'very difficult', while texts scoring 7 or below with other tools are deemed 'easy' or 'very easy. In terms of difficulty, Flesch Reading Ease has an inverse relationship with the reading score. According to Flesch Reading Ease, the lower the score, the harder a text is to read, whereas according to other readability tools, the lower the score, the easier a text is to read.

Literature Review

Rakes (1973) conducted a comparative assessment on the readability of 29 most commonly used books in Adult Basic Education in Tennessee and found no significant correlation between most formulae pairs. The only significant correlation was between the Gunning Fog Index and the Fry Readability Graph. However, there was no significant difference in the mean scores for around 69% of the materials as shown by different readability scores.

Chen (2012) conducted a study to examine the most common readability formulas using three children's books to compute the numerical or grade-level score. Results showed that readability formula embedded in a Microsoft Word processing program was the easiest way to obtain accurate readability scores by typing the entire book into a word document. The scores differed slightly depending on the readability tool used. However, the ranking of the three books was consistent among the different tools. The readability scores showed similar results for the tool, the Fry (calculated by hand) and the Flesch-Kincaid from Microsoft Word and from free web-based calculators. Both formulas used the same variables - sentence length and number of syllables per word - to calculate readability levels.

Acar et al (2017) compared the readability scores of various technical and scientific English texts and their corresponding Turkish translation using a variety of readability tools. To check comprehensibility levels, 43 English lecturers completed a checklist of source texts and their corresponding questions. The comprehensibility level ranged from 0 to 100. The study found that Flesh Reading Formula and Ateşman Reading Ease Formula were compatible to each other. A statistical difference was also found in the readability and comprehensibility levels between the texts.

Zhou et al (2017) conducted a study on the consistency of the readability equations in estimating the readability of design standards and found that different equations calculate readability scores differently for the same texts. It is assumed that defining a syllable, word, or sentence is straightforward, and tools that calculate these readability equations should yield the same score for the same text passage. However, the differences in the results were substantial. Disparity in results might have arisen from the way hyphens, contractions, digits, dates, acronyms, abbreviations, and other text elements like punctuations were treated in word and syllable counts. The rules for counting them vary between equations. Therefore, the researchers found that while the equations provided useful insights, their scores should not be treated as precise estimates. The study did not suggest any strong preference for any particular readability equation.

According to Srisunakrua et al (2019), readability plays a significant role in English language teaching because it provides an overall picture of the difficulty level of a text. Considering readability is practical for making decisions about teaching and testing materials. The study explored the readability levels of reading passages in the Thai National Education English textbooks and English Tests through Coh-Metrix computational tool. The findings showed a mismatch between the readability levels and linguistic characteristics, with passages in the English test being more difficult than those in the textbooks. The study

recommended that teaching and testing materials be aligned with appropriate readability levels to ensure that students benefit from them and their performance in testing is not adversely affected.

Redmiles et al (2019) found that the readability of a digital text can affect people's ability to learn new information from digital resources. Readability also affects search rankings, and is used to assess the performance of Natural Language Processing (NLP) systems. The study presented a comparison of the well-known readability measures and a new approach, Smart Cloze across four different corpora. The researchers assessed the convergent and content validity of the measures on these corpora. These results provided a basis for more precise readability measurements and better evaluation of new NLP systems and tools.

Nattam et al (2023), published a study which found that Patient Education Materials (PEMs) did not meet the readability recommendations of the American Medical Association (AMA) and National Institutes of Health (NIH), making them very difficult to read and comprehend for patients with low health literacy. This study emphasized the need to improve the readability of patient education materials to help patients make informed decisions. Readability should be checked for all medical and health documents. Content creators who write health education materials should use validated readability tools.

Cherian & Jha (2024) compared the difficulty level of texts used EFL (English as a Foreign Language) teaching and exams, focusing on Reading and Listening skills. The texts were assessed for text complexity using Flesch-Kincaid readability scores and CEFR (Common European Framework of Reference for Languages) levels. The study indicated that teaching materials and test materials matched in complexity and reading difficulty. However, the tools sometimes showed disparities between CEFR level checkers and readability score checkers. The study suggested that exam writers and content developers use both tools for a comprehensive view of text complexity, while relying on a single metric for reliability and consistency, especially for test materials.

Methodology

This study compares the readability scores of four reading texts used in Reading exams by analyzing the scores generated by different readability checkers. Although material developers and test writers globally use readability scores to check the text difficulty, differences in results when checked by various tools continue to be underexplored. Specifically, the study investigates the consistency of results generated by different readability tools. It is crucial to use reliable tools to assess the readability of texts that are used for teaching and assessment, especially in EFL contexts, to ensure that students are taught and tested at an appropriate level.

The Sample of the Study

As test samples, the researcher used texts she prepared for the Reading Exams in the University of Technology and Applied Sciences – Ibra for the Level 3 Mid Semester Exam (MSE) and Final exam (FE) in the first semester of the academic year 2017-2018. The university was previously known as Ibra College of Technology. The researcher used her own texts created for the exams to avoid plagiarism or copyright issues. Moreover, these exams have been used as practice exams and so, the exams are not currently stored in the university's exam bank. Therefore, using these texts will not affect exam confidentiality and security. Level 3 exams are intended to have an $8^{th/}$ 9^{th} grade readability or 'standard/plain English', as students are EFL learners.

Tools Used

There are various tools to measure readability, but this study focuses on six tools, namely Flesch Reading Ease, Flesch-Kincaid Grade Level, Gunning Fog Index, SMOG Index, Coleman-Liau Index, and the Automated Readability Index, since these are the most popular readability tools used by educators. Understanding the relationship between scores from readability tools can help educators select and prepare materials that are appropriately challenging for students at various proficiency levels of English language, especially in the EFL context.

This study analyzes four different Reading texts to check their readability scores using online tools such as 'Readability Scoring System' (https://readabilityformulas.com/readability-scoring-system.php), and 'Readability Test' (https://www.webfx.com/tools/read-able/) to identify variations between results created by different tools and websites.

Analysis of the texts

Appropriate Difficulty Level for Level 3 Exams Based on the Course Materials: Readability Score: 80-60 (Flesch Reading Ease), (Standard-Fairly Easy)																								
	Flesch Reading Ease				Flesch-Kincaid Grade Level				Gunning Fog Index				SMOG Index				Coleman-Liau Index				Automated Readability Index			
	words	sentences	score	grade	words	sentences	score	grade	words	sentences	score	grade	polysyllables	sentences	score	grade	words	sentences	score	grade	words	sentences	score	grade
ME1	501	40	64	8 & 9	501	40	7.41	7	501	40	10.1	10	69	40	7.73	8	501	40	9.22	9	501	40	7.48	7
ME1	502	40	66.2		502	40		7.1	502	40		9.5		40		7.3	502	40		11.5	502	40		6.7
ME2	279	19	60	8 & 9	279	19	8.53	9	279	19	12.5	college	52	19	9.63	10	279	19	10.85	11	279	19	9.46	9
ME2	282	23	62.1		282	23		7.6	282	23		10.4		23		8.4	282	23		12	282	23		6.9
FE1Tool	483	35	63	8 & 9	483	35	7.79	8	483	35	10.2	10	58	35	7.58	8	483	35	9.98	10	483	35	8.47	8
FE1Tool	483	35	62.1		483	35		8	483	35		10.8		35		8	483	35		12	483	35		7.7
FE2	286	25	76	7	286	25	5.54	6	286	25	9.6	10	36	25	7.1	8	286	25	7.99	8	286	25	6.21	6
FE2	286	25	75.1		286	25		5.6	286	25		9.3		25		7.1	286	25		10.4	286	25		5.3

Table 2: Analysis of Readability for Reading Text Samples from L3 Mid Semester & Final Exams

Observations

1. All tested L3 exams correspond to the suggested Flesch Reading Ease readability score range of 80-60 (standard).
2. There are variations in grade levels shown by different tools. This is due to the differing formulas used for calculations.
3. Sometimes, different websites show disparities in counting the number of words and sentences because of the variations in how they calculate characters and hyphenated words.
4. Tool 1 shows both scores and grade levels for all readability checkers, whereas Tool 2 displays only the Flesch Reading Ease score for readability ease but provides grade levels for other readability checkers. Additionally, Tool 2 doesn't account for polysyllables.

Implications for Test Developers and Content Creators

1. Exam writers and material developers should ensure that the difficulty level of the texts aligns with their students' proficiency level by checking their readability using readability checkers.
2. Discrepancies can occur between results from different readability score checkers. Therefore, it is crucial to validate test specifications for developing tests by providing clear instructions on which readability score is to use to ensure consistency.
3. Exam writers and material developers should cross check results from one tool or website with a second tool or website to ensure reliability and consistency.
4. Exam writers and material developers can adjust text readability by modifying sentence length. Longer sentences and multi-syllabled words increase difficulty level and affect readability.

Conclusion

Creating appropriate materials for classroom teaching and testing according to the level of the learners is crucial, especially in the EFL context since students study a foreign language. Therefore, exam writers and material developers are obliged to ensure the complexity of the texts they create are appropriate for their students. Readability checkers are efficient tools for helping them achieve this goal by evaluating text readability.

References

1. "About Readability." *Readable*, 2020, https://readable.com/readability/.

2. Acar, Arif, and Kadir Ulkar İŞİSAĞ. "Readability and Comprehensibility in Translation Using Reading Ease and Grade Indices." *International Journal of Comparative Literature and Translation Studies*, vol. 5, no. 2, 2017, pp. 47-52, https://doi.org/10.7575/aiac.ijclts.v.5n.2p.47

3. Chen, H. "How to Use Readability Formulas to Access and Select English Reading Materials." *Journal of Educational Media & Library Sciences*, vol. 50, no. 2, Winter 2012, pp. 229-254. https://eds.p.ebscohost.com/eds/detail/detail?vid=0&sid=90fcd128-ad78-4d6d-82f6-19d746dff667%40redis&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3d%3d#AN=86000616&db=lxh.

4. Cherian, G., and Jha, S.K.. "Impact of Readability and CEFR Levels on EFL Materials." *International Journal of Emerging Technologies and Innovative Research*, vol. 11, no. 6, June 2024, pp. ppa44-a51, http://www.jetir.org/papers/JETIR2406007.

5. "Flesch Reading Ease and the Flesch Kincaid Grade Level." *Readable*, n.d., https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/.

6. Gunning, G. T. "The Role of Readability in Today's Classrooms." *Topics in Language Disorders*, vol. 23, 2023, pp. 175-189.

7. "Measure Reading Levels with Readability Indexes." *Wylie Communications*, 2021, https://www.wyliecomm.com/2021/11/measure-reading-levels-with-readability-indexes/#_ftn2.

8. Nattam, Akhilesh, et al. "Assessing the Readability of Online Patient Education Materials in Obstetrics and Gynecology Using Traditional Measures: Comparative Analysis and Limitations." *Journal of Medical Internet Research*, vol. 25, 2023, e46346, https://www.jmir.org/2023/1/e46346/.

9. "Readability Formulas." *Readable*, n.d., https://readabilityformulas.com/.

10. "Readability Tests." *Siteimprove*, n.d., https://help.siteimprove.com/support/solutions/articles/80000448325-readability-tests.

11. Redmiles, Elissa, et al. "Comparing and Developing Tools to Measure the Readability of Domain-Specific Texts." *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Association for Computational Linguistics, 2019, pp. 4831-4842, https://aclanthology.org/D19-1489.

12. Rakes, Thomas A. "A Comparative Readability Study of Materials Used to Teach Adults to Read." *Adult Education*, vol. 23, no. 3, 1973, pp. 192-202, https://doi.org/10.1177/074171367302300302.

13. Srisunakrua, Thanaporn, and Tipamas Chumworatayee. "Readability of Reading Passages in English Textbooks and the Thai National Education English Test: A Comparative Study." *Arab World English Journal*, vol. 10, no. 2, 2019, http://dx.doi.org/10.2139/ssrn.3418124.

14. "The Gunning Fog Index." *Readable*, n.d., https://readable.com/readability/gunning-fog-index/.

15. Wylie Communications. "Measure Reading Levels with Readability Indexes." *Wylie Communications*, 2021, https://www.wyliecomm.com/2021/11/measure-reading-levels-with-readability-indexes/#_ftn2.

16. Zhou, Shuang, et al. "How Consistent Are the Best-Known Readability Equations in Estimating the Readability of Design Standards?" *IEEE Transactions on Professional Communication*, vol. 60, no. 1, Mar. 2017, pp. 97-111, https://doi.org/10.1109/TPC.2017.2669398.