Journal of Health Care Communications

ISSN: 2472-1654

Open access Research Article

The Accuracy of Readability Formulas in Health Content: A Systematic Review

Brittany U. Carter[1*], Kinjalbahen Nayak[2], Isabella Vembenil[3]

[1]*Department of Health and Research, Wellsource, United States*

[2]*Department of Public Health, Kent State University, United States*

[3]*The Ursuline School, United States*

ABSTRACT

Background: Readability formulas are commonly used to assess the ease in which a reader can understand written text. It is unclear which is best to use when evaluating the readability of health content. This systematic review assessed the accuracy of readability formulas when evaluating health content targeting adults.

Methods: Searches of PubMed, Cochrane Library, and Education Resources Information Center from inception through January 31, 2024 were conducted. Two investigators independently reviewed abstracts and full-text articles against a set of a priori inclusion criteria. Studies evaluating the accuracy or validity of readability formulas when applied to health content for adults were included. Data was analyzed qualitatively.

Results: Three fair-quality studies were included. One study found readability formulas frequently underestimated the document's difficulty when compared to expert panel ratings. Another study found very low correlations between the readability formulas and user difficulty ratings. Another study found readability formulas were unable to consistently identify problematic health survey questions among question pairs.

Conclusion: Evidence is limited regarding the accuracy of readability formulas for health content. Study limitations and those associated with the readability formulas likely contributed to the poor performance. More studies are needed to determine which is best to use for health content.

Keywords: Health literacy; Readability formulas; Health content

INTRODUCTION

According to the National Assessment of Adult Literacy, 88% of U.S. adults are not proficient in health literacy [1], that is, the degree to which they have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions [2]. Low health literacy is associated with negative health outcomes including increased hospitalizations, poor medication adherence, and higher mortality rates [3,4]. The ease in which a reader can understand written text-or readability-is a contributing factor to health literacy as it affects the reader's ability to comprehend the content. The Centers for Disease Control and Prevention (CDC) and National Institutes of Health (NIH) recommend, and the Centers for Medicare and Medicaid Services (CMS) requires for its members, that health materials be written at a reading level no higher than an 8th grade, while the American Medical Association (AMA) recommends patient-facing material be written at a 6th grade reading level.

Readability can be determined by using various formulas such as the Flesch Reading Ease (FRE), Flesch-Kincaid Grade Level (FKGL) index, and the Simple Measure of Gobbledygook (SMOG) index. However, there are concerns, criticisms, and limitations regarding the use and accuracy of these readability formulas in general as well as when assessing health content [5-9]. For example, although the government agencies and

Received:	05-February-2024	**Manuscript No:**	IPJHCC-24-19014
Editor assigned:	07-February-2024	**PreQC No:**	IPJHCC-24-19014 (PQ)
Reviewed:	21-February-2024	**QC No:**	IPJHCC-24-19014
Revised:	26-February-2024	**Manuscript No:**	IPJHCC-24-19014 (R)
Published:	04-March-2024	**DOI:**	10.36846/2472-1654-9.1.9002

Corresponding author Brittany U. Carter, Department of Health and Research, Wellsource, United States, E-mail: brittany.u.burda@gmail.com

Citation Carter BU, Nayak K, Vembenil I (2024) The Accuracy of Readability Formulas in Health Content: A Systematic Review. J Healthc Commun. 9:9002.

medical associations provide or suggest a threshold at which content should not pass (i.e., a 6th or 8th grade reading level), studies have shown that the readability formulas when used for health content largely disagree by up to 5 reading grade levels on the same text [10]. This may be the result of the formulas using different components as well as not being designed, intended, or appropriate for use in evaluating health content. These issues with the readability formulas can lead to confusion as to which is best to use to assess the readability of health content to maximize health literacy. It also results in spending unnecessary resources to revise health content that might otherwise not need to be modified. There is currently no guidance as to which formula should be used in health content. The aim of this systematic review was to evaluate the evidence on the accuracy of readability formulas when assessing health content targeting adults to inform which to use.

MATERIALS AND METHODS

Scope of the Review

This systematic review addressed the key question: What is the accuracy of readability formulas when evaluating health content directed towards adults? This systematic review is not registered.

Data Sources and Searches

Comprehensive literature searches were performed in PubMed, Education Resources Information Center (ERIC), and the Cochrane Library from inception through January 31, 2024. The literature search strategies were developed by one of the investigators and are available in **Appendix A**. The database searches were supplemented by reviewing the reference lists from relevant studies and background articles. All references were managed using EndNote™ version 19 (Thomson Reuters, New York, NY).

Study Selection

Two investigators independently reviewed titles and abstracts using an online platform (Rayyan®, Rayyan Systems, Inc), and subsequently, full-text articles against pre-specified inclusion and exclusion criteria (**Appendix B**). Disagreements were resolved through discussion and when necessary, consultation with a third investigator.

Studies published in the English language that assessed the validity or accuracy of readability formulas in the evaluation of written health content directed towards an adult population (i.e., ages 18 years or older) were included. There were no limitations on the topic of the health content (e.g., nutrition, diabetes), the source used to deliver the written health content (e.g., website, brochure, medical journal), or the setting in which the health content was provided (e.g., healthcare organization, university-based). Studies that evaluated content that was unrelated to health, or any content that was directed towards children and/or adolescents (i.e., ages less than 18 years) were excluded. There were also no limitations on the study design or publication type.

Readability formulas are those that assess the ease with which a reader can understand written text and are typically reported as a grade level or a score that is translatable to a grade level equivalent to the U.S. education system. Only the following readability formulas were included as they are the most frequently used in health [10]: FRE, FKGL, SMOG, Dale-Chall, Spache, FORCAST, Fry Graph, Rate Index (RIX), Automated Readability Index (ARI), Gunning-Fox Index, and the Coleman-Liau Index. Readability formulas specifically designed to measure the readability of tables, charts, or graphs (e.g., Suitability Assessment of Materials [SAM]) were excluded. Formulas used to measure comprehensiveness of the content, reading comprehension (e.g., Cloze test), suitability or quality of the content, or to estimate health literacy (e.g., Rapid Estimate of Adult Literacy in Medicine [REALM]) were excluded.

Studies were included if the comparator group with whom to compare the results of the readability formulas) were standard health textbooks at a specific grade-level or a panel of individuals rating the content. Studies that compared the grade levels between readability formulas were excluded, as were studies that compared the readability of the content to the state or national average grade level or to the content developer's stated readability level.

Studies were included if they reported on the validity or accuracy of the readability formulas through correlations or measures of accuracy (e.g., sensitivity). Studies that reported on the agreement between two or more readability formulas were excluded.

Quality Assessment and Data Abstraction

Quality assessment of the included studies using an existing validated tool, such as those used for diagnostic accuracy or specific observational study designs [11,12], was not performed as these tools did not met the needs of this systematic review. In lieu of formal quality assessment, two investigators discussed the limitations of each study and reached consensus on a final quality rating of good, fair, or poor. Such study limitations included the sample size, the depth and breadth of health content evaluated including its selection, the comparator used, the approach to the analysis, and generalizability. Fair-quality studies often had minor flaws while a poor-quality study had significant design flaws that seriously called into question the validity of the study results, and thus, poor quality studies were excluded.

One investigator abstracted data from all included studies and a second investigator checked the data for accuracy. Abstracted data included information on the study design, descriptions of the health content, the readability formulas evaluated, the comparator used, validity and/or accuracy outcomes, and study limitations.

Data Synthesis and Analysis

Results from included studies are summarized qualitatively as there were too few studies and not enough data to allow for meta-analysis.

RESULTS

A total of 1,652 unique citations and 28 full-text articles were reviewed (**Figure 1**). Three fair-quality studies met the inclusion criteria; a list of excluded studies at full-text is available in

Appendix C [13-15]. In summary, two of the studies assessed the readability of diabetes content while the other assessed health survey questions (Table 1). The accuracy of FKGL was evaluated in three studies; SMOG and Gunning Fog Index were evaluated in two studies; and Dale-Chall and FRE were evaluated in one study each. No other readability formulas were evaluated in the studies. The comparator group of two studies consisted of a panel of individuals who assessed the ease or difficulty of the content while the other was designed to identify the problematic question of a question pair. All three studies found the readability formulas to not perform well. Further information about each of the included studies can be found below.

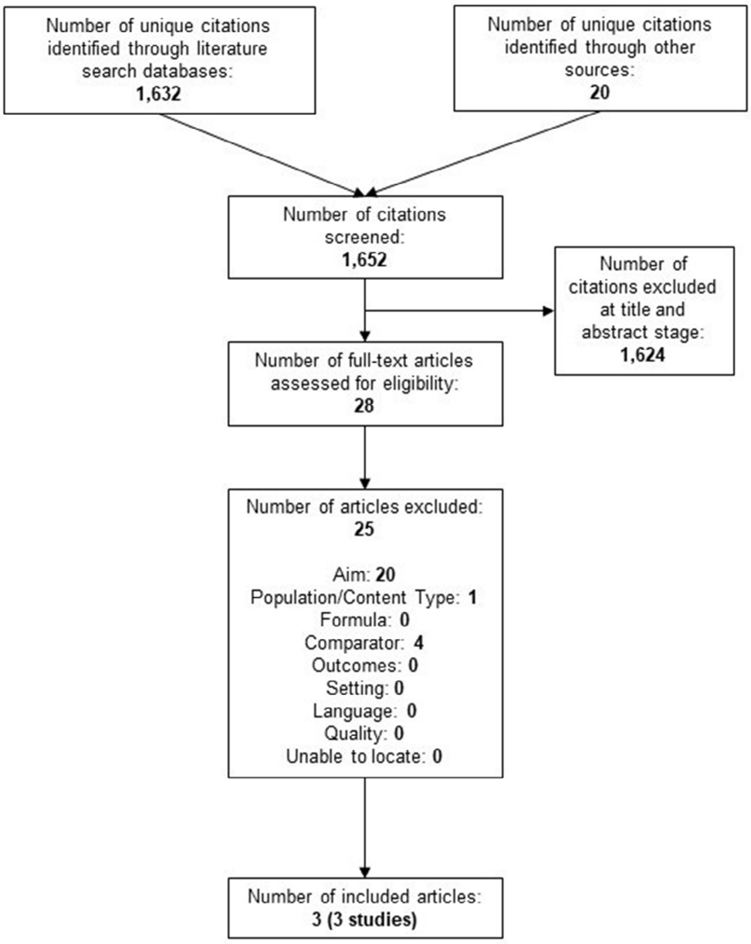
Number of unique citations identified through literature search databases: 1,632
Number of unique citations identified through other sources: 20
Number of citations screened: 1,652
Number of citations excluded at title and abstract stage: 1,624
Number of full-text articles assessed for eligibility: 28
Number of articles excluded: 25
Aim: 20
Population/Content Type: 1
Formula: 0
Comparator: 4
Outcomes: 0
Setting: 0
Language: 0
Quality: 0
Unable to locate: 0
Number of included articles: 3 (3 studies)

Figure 1: Literature flow diagram

Table 1: Characteristics of included studies

Study	Brief Methods	Health Content	Readability Formula(s)	Comparator	Quality
Kandula, 2008 [14]	324 diabetes-related documents manually reviewed by an expert panel to assign a readability score which was compared to the FKGL and SMOG scores	Diabetes mellitus consumer education materials (k=142), news stories (k=34), clinical trial records (k=39), clinical reports (k=38), scientific journal articles (k=38), consumer education materials for children (n=33)	FKGL SMOG	Expert panel (n=5) developed Likert scale (1-7)* to assess ease of reading material	Fair
Lenzner, 2014 [15]	Readability scores calculated for 71 pairs of health survey questions (one difficult and one improved) to determine if formulas correctly classified them as being more or less difficult	Health survey question pairs as published in journal articles or textbooks (k=15), or in the Q-Bank database (k=56)	FKGL FRE Gunning Fog Dale-Chall Problematic survey question (e.g., syntactically complex)	Problematic survey question (e.g., syntactically complex)	Fair

Zheng, 2017 [13]	Readability scores of 382 diabetes-related documents compared to laypeople's perceived difficulty of the content	General health information on diabetes (Wikipedia articles; k=140) and electronic health records notes with diabetes ICD-10 codes (k=242)	FKGL SMOG Gunning-Fog	Laypeople's (n=15) perceived difficulty level of the content on a Likert scale (1-10)†	Fair

FKGL=Flesch-Kincaid Grade Level; FRE=Flesch Reading Ease; ICD-10=International Classification of Disease 10th edition; SMOG=Simple Measure of Gobbledygook

*Lower rating indicates lower education level: 1=any; 3=high school graduate (12th); 4=some college; and 7=professional education only

†Lower rating indicates easiest to understand with highest score indicating the most difficult to understand

One fair-quality study used a panel of five health literacy and clinical experts to assess the readability of 324 documents on diabetes in comparison to two readability formulas (FKGL and SMOG) [14]. The documents included consumer education materials (k=142), news reports (k=34), clinical trial records (k=39), scientific journal articles (k=38), and clinical reports (k=38); 33 documents targeted children and are not discussed further. The expert panel developed a "gold standard" using a Likert scale (1-7) to assess the ease of the reading materials with lower scores indicating lower education levels required to understand the content. The "gold standard" had good interrater reliability and validity (data not shown). Across all documents (including those directed at children), the grades assigned by FKGL and SMOG had a Spearman's correlation coefficient of 0.54 and 0.55, respectively, with the expert ratings; both were statistically significant moderately strong correlations (p<0.0001) (Table 2). When evaluated by document type, the only significant difference between the expert ratings and the two readability formulas were the clinical reports with the latter demonstrating the content to be understandable at a lower grade level. The formulas frequently underestimated the document's difficulty across all document types when compared to the expert panel ratings. Limitations of this study include using a non-validated "gold standard" which relied on the use of expert knowledge as opposed to lay persons, having a small sample size of the experts providing the ratings, evaluating a few content sources where health information is delivered, and covering only one health topic.

Table 2: Correlations of Readability Formulas

Study	Document Type	Formula	Mean (SD)	Correlation
Kandula, 2008 [14]	All (k=324)*	FKGL	NR	0.54 (p<0.0001)
		SMOG	NR	0.55 (p<0.0001)
	News report (k=34)	FKGL	11.29 (2.22)	NR
		SMOG	13.68 (1.80)	
		Expert Panel†	4.26 (0.86)	
	Consumer education material (k=142)	FKGL	10.19 (2.05)	NR
		SMOG	12.62 (1.69)	
		Expert Panel	4.02 (1.18)	
	Clinical trial record (k=39)	FKGL	15.71 (2.37)	NR
		SMOG	17.10 (1.93)	
		Expert Panel	6.33 (0.89)	
	Scientific journal article (k=38)	FKGL	15.82 (1.43)	NR
		SMOG	17.24 (1.07)	
		Expert Panel	6.55 (0.72)	
	Clinical report (k=38)	FKGL	8.38 (1.78)	Significant difference between expert panel and FKGL/SMOG (p-value NR)
		SMOG	11.36 (1.44)	
		Expert Panel	6.13 (0.84)	
Zheng, 2017 [13]	Wikipedia articles (k=140)	FKGL	14.75	0.1758
		SMOG	11.07	0.4134
		Gunning Fog	12.33	0.2695
		User Rating‡	4.41	–
	EHR (k=242)	FKGL	9.87	0.2999
		SMOG	8.74	0.1024
		Gunning Fog	8.16	0.1272
		User Rating	5.35	–

EHR=electronic health record; FKGL=Flesch-Kincaid Grade Level; FRE=Flesch Reading Ease; NR=not reported; SD=standard deviation; SMOG=Simple Measure of Gobbledygook

*Includes the 33 documents directed towards children

†Lower rating indicates lower education level: 1=elementary school; 3=high school graduate (12th); 4=some college; and 7=professional education only

‡Lower rating indicates easiest to understand with highest score indicating the most difficult to understand

Another fair-quality study also compared the readability of three formulas (FKGL, SMOG, and Gunning Fog) against a difficulty rating as perceived by 15 Amazon Mechanical Turk users (Table 1) [13]. The Amazon Mechanical Turks had English as their primary language and were master workers; three had high school diplomas, 7 had an associate degree, four had a Bachelor's degree, and one did not report their education level. 140 Wikipedia articles containing general health information on diabetes and 242 de-identified Electronic Health Record (EHR) notes with diabetes codes (International Classification of Diseases 10th edition codes 250.00 to 250.93) were evaluated. Each user measured the level of difficulty of 20 randomly assigned paired analogous documents using a Likert scale (1-10) with lower scores indicating the easiest to understand. The readability formulas suggested the EHR notes to be significantly easier than the Wikipedia articles while the user ratings showed the opposite (Table 2). All correlations were very low between the user ratings and readability formulas suggesting the users' perceived difficulty and the readability formulas predictions were inconsistent (Table 2). Limitations of this study include using a non-validated Likert scale, having a small sample size of lay persons provide ratings in a population that may not be generalizable to the United States [16], evaluating only two content sources where health information is delivered, and covering only one health topic.

The final fair-quality study evaluated whether four readability formulas (FRE, FKGL, Gunning Fog Index, and Dale-Chall) correctly identified the problematic survey question from 71 question pairs each containing a problematic question and an improved version of the same question (Table 1) [15]. Problematic questions were described as syntactically complex or vague. An example of a problematic question was "Have you ever had a Pap test to check for cervical cancer?" while the improved question was "Have you ever had a Pap smear or Pap test?" The question pairs came from two sources: Journal articles or textbooks, and the Q-Bank database. The literature sources addressed questionnaire design and included examples of problematic survey questions together with recommendations for improving them; 15 pairs were reported in five publications. The Q-Bank database provides access to question evaluation research and contains pretested and improved survey questions. 56 questions originally identified having "problematic terms" (k=34) or were "overly complex" (k=22) were included with their improved version. The FRE and FKGL formulas performed better than the Gunning Fog Index and Dale-Chall formulas (Table 3), however, all showed low classification accuracy (less than 52%). Binomial testing showed none of the formulas performed significantly better than expected from random guessing, in fact, three of the formulas (FKGL, Gunning Fog Index, and Dale-Chall) performed worse than by chance alone (data not shown). When analyses were limited to from where the questions pair came, the classification accuracy was considerably better for questions from the Q-Bank database than from the literature (Table 3). Limitations of this study include relying on an assumption that the pairs contained one problematic and one improved question as they were not tested.

Table 3: Success Rates of Readability Formulas Correctly Classifying Question Pairs [15]

Health Survey Questions	Readability Formula	Success Rate	P-value
All (k=71)	FRE	51%	1
	FKGL	49%	1
	Gunning Fog	39%	0.1
	Dale-Chall	38%	0.06
From journal articles or textbooks (k=15)	FRE	27%	0.12
	FKGL	27%	0.12
	Gunning Fog	27%	0.12
	Dale-Chall	27%	0.12
From Q-Bank database (k=56)	FRE	57%	0.35
	FKGL	55%	0.5
	Gunning Fog	43%	0.35
	Dale-Chall	41%	0.23
From Q-Bank database and tagged as having a "problematic term" (k=34)	FRE	53%	0.86
	FKGL	50%	1
	Gunning Fog	44%	0.61
	Dale-Chall	44%	0.61
From Q-Bank database and tagged as being "overly complex" (k=22)	FRE	64%	0.29
	FKGL	64%	0.29
	Gunning Fog	41%	0.52
	Dale-Chall	36%	0.29

EHR=electronic health record; FKGL=Flesch-Kincaid Grade Level; FRE=Flesch Reading Ease; NR=not reported; SD=Standard deviation; SMOG=Simple Measure of Gobbledygook

DISCUSSION

Only three studies assessing the accuracy of readability formulas when evaluating health content were identified and only a few of the readability formulas were evaluated in these studies. The studies show that the readability formulas did not correlate well with panel ratings or accurately identify problematic health survey questions [13-15]. There are too few studies to make any definitive conclusions on the accuracy of readability formulas when assessing health content directed towards adults. And there is a multitude of reasons that the readability formulas performed poorly, albeit in the few included studies in this systematic review, including study design limitations and some of the previously mentioned preexisting concerns, criticisms, and limitations of using readability formulas to assess health content.

First, readability formulas were not created with health content in mind. These formulas were invented in the late 19th and early 20th centuries to assess the comprehensibility of general educational literature for children and adults-not health literature [17]. The developers used written text from school textbooks, newspapers, and magazines-not health-related sources-to identify components that would predict comprehension or assess reading difficulty. Components of formulas identified as the best to determine readability and used in the equations include the length of words, the length of sentences, and/or the frequency of "hard words". Health content, however, often contains polysyllabic words, scientific terminology, and medical abbreviations which results in it being assessed at a higher grade level. The components or coefficients of the readability formulas might differ if health content were used in their development. In fact, recent studies have shown different elements such as syntax and semantics to be more important when assessing health content than the components used in existing readability formulas [18,19]. Although one adult reading ease formula used passages on health topics from various sources to identify factors influencing the difficulty of reading level among adults with limited reading ability, and used three factors in the formula, it was not included in this review as it is not commonly used [20].

Second, the readability formulas also did not use health content in their validation. The McCall-Crabbs Standard Test Lessons in Reading (for children), the Cloze test, or other similar types of texts were used for validation-not health content [17]. Although the predictions of the Dale-Chall readability formula across 55 passages of health education materials were highly correlated with the judgments of readability experts [17]; there was no further information regarding this statement or validation study from the 1940's and thus not included in this review. Validating the readability formulas with health content may contribute to refinements in the equations when applied to this specific type of content.

Third, the assumptions of the readability formulas also do not hold true with regards to health content. According to the formulas, a word or sentence is harder if it is longer. As mentioned previously, health content tends to include polysyllabic words such as diabetes and hypertension which decreases readability. Attempts to improve the readability, however, such as substituting polysyllabic words with multiple smaller words (e.g., hypertension to high blood pressure) make it less readable per the formulas because sentence length increases. In addition, readability formulas do not consider if the words are familiar or the sentences are clear and cohesive [6]. Many health terms and abbreviations-such as cholesterol and HIV-are common to lay audiences, however, including these words or abbreviations makes the health content less readable per the formulas. And adding more words to explain these health terms more simply again counts against readability.

And finally, the source from which the health content is provided may further contribute to issues with using these formulas to assess readability. These readability formulas assume content is delivered in well written paragraphs with complete sentences of at least 100 words. A patient's EHR, however, may contain lists of clinical events, drug names, tests results with numbers and measurement units, medical abbreviations, and short and incomplete sentences, thus lowering the readability [13]. It is possible that the health content used in the included studies may not have met the minimum requirements for the readability formulas to accurately work thus contributing to the poor performance of the formulas. For example, the Dale-Chall formula requires at least a 100-word sample of text while FORCAST requires a 150-word sample [21]. When it comes to health survey questions, for example, most are less than 100 words and thus no readability formula may be suited for their assessment. Given the uncertainty that adequate health content was used in the included studies contributes to the lack of a definitive conclusion on the readability formulas accuracy. It is not possible to definitively conclude that health content might not be a good fit for the readability formulas without further good quality studies.

The results of this systematic review do not suggest that the readability formulas are useless or that the limited evidence suggests they should not be used as there must be a starting place for ensuring lay audiences understands health information. For example, one could use these formulas to estimate the ceiling of the content's readability and subsequently edit the content as needed. In addition, previous studies have shown that readability formulas are valid and reliable with other types of content such as fiction and non-fiction literature, general education textbooks, and governmental websites [17,21,22]. The choice of formula can have a strong influence on the conclusions drawn from readability and there is minimal guidance on which to use to assess health content. As mentioned previously, readability of the same health content can differ by up to five reading levels per the formulas [10]. One readability formula-SMOG-has been endorsed for use in healthcare as it is the only formula based on complete comprehension of the content meaning all individuals who read the content also understand it [10,21,23]. Such high comprehension is very important in healthcare to ensure individuals experience safe and effective outcomes. Expecting such high comprehension, however, means SMOG scores are often 1-2 grades higher than the results of other readability formulas [24] so more effort may be needed to bring the grade level down. There are lots of tools and suggestions on how to improve readability when one's health content is grading too high such as the Agency for Healthcare Research and Quality's Health Literacy Toolkit [25] or the Centers for Medicare and Medicaid Services Guidelines for Effective Writing and Plain Language [26,27].

Implications of the Findings for Health Communication Research and Practice

The greatest implication of these findings on health communication research and practice is there is no universally recommended formula for content developers to use that has proven validity and reliability in assessing the readability of health content. Every effort made to help improve health literacy (and subsequently health outcomes) by attempting to optimize this one aspect (i.e., readability) may be thwarted because the output from readability formulas may be incorrect. Health communication research should conduct future studies to determine which readability formula is best to use for health content either by validating and/or refining existing formulas or creating a *de novo* one. These future studies should also consider the source of the health content as written text in EHRs may differ from that provided on a drug label or a website about diabetes care-all of which may be read by a lay person. With regards to health communication practice, health content developers should continue to strive to ensure optimal readability by following proven guidance to ensure the reader can read, process, and understand the content. The use of readability formulas-although at risk of an inaccurate grade level being reported-can still occur as a starting point to see what content may require improvement to warrant maximum comprehension.

LIMITATIONS OF THE SYSTEMATIC REVIEW

There are a few limitations of this review. First, the literature search was not developed nor reviewed by a research librarian and only three databases were searched; both of which may have resulted in the searches missing relevant studies. The literature search also did not include terms for online readability calculators (e.g., readable.com) that use the underlying included readability formulas which may have also missed relevant studies. This limitation was mitigated by checking references of relevant studies and background articles to ensure none were missed. Second, some readability formulas were omitted such as the Golub Syntactic Density Score, Linsear Write, and the Lexile framework. These omissions were not concerning as the most prevalent readability formulas used in health were in the literature search strategies [10]. Targeted searches for studies on the omitted formulas also resulted in few-to-none being found and thus the literature searches were not modified from the initial search. Third, there is no universal gold standard [10] and thus it was defined in this review as either a panel of individuals assessing the content's difficulty to read or understand the materials or standard health text at a specified grade reading level. A more broadly defined comparator may have resulted in additional included studies. Fourth, no formal quality assessment was conducted; however, the investigators considered the limitations of the studies to ensure none had any significant flaws that called into question the validity of the results which should be sufficient to have trust in the conclusions of this review. And finally, most of the initial validity and reliability studies of the readability formulas were excluded as they were validated among children; used content targeted towards children, or used content that was not related to health.

CONCLUSION

Limited evidence contributes to the concerns, criticisms, and limitations of using existing formulas to assess the readability of health content. In lieu of having a better alternative, using existing readability formulas (especially SMOG) is a start to ensuring improved health literacy and subsequently health outcomes.

ACKNOWLEDGEMENT

None.

COMPETING INTERESTS STATEMENT

BUC was employed by Wellsource, a company who develops health content. KN and IV worked on this project as student interns of Wellsource.

FUNDING STATEMENT

None.

REFERENCES

1. Lopez C, Kim B, Sacks K (2022) Health literacy in the United States: Enhancing assessments and reducing disparities. Santa Monica, CA: Milken Institute.
2. Ratzan SC, Parker RM (2000) Introduction in: Selden CR, Zorn M, Ratzan SC and Parker RM, eds. Health Literacy.
3. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K (2011) Low health literacy and health outcomes: An updated systematic review. Ann Intern Med. 155(2):97-107.
4. Fan ZY, Yang Y and Zhang F (2021) Association between health literacy and mortality: A systematic review and meta-analysis. Arch Public Health 79(1):119.
5. U.S. Department of Health and Human Services (2010) Using readability formulas: A cautionary note.
6. Agency for Healthcare Research and Quality (2015) Tip 6. Use caution with readability formulas for quality reports.
7. Redish G (2000) How valid are readability formulas for technical material for adult readers? J Com Doc. 24:132-137.
8. Pichert JW, Elam P (1985) Readability formulas may mislead you. Patient Educ Couns. 17(2):181-191.
9. Danielson KE (1987) Readability formulas: A necessary evil? Reading Horizons: A Journal of Literacy and Language Arts. 27:178-88.
10. Wang LW, Miller M, Schmitt M, Wen F (2013) Assessing readability formula differences with written health informaiton materials: Application, results, and recommendations. Res Social Adm Pharm. 9(5):503-516.
11. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks

JJ, et al. (2011) QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 155(8):529-536.

12. Wells GA, Shea B, O'Connell D, Robertson J, Peterson J, et al. (2021) The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa: The Ottawa Hospital Research Institute.
13. Zheng J, Yu H (2017) Readability formulas and user perceptions of electronic health records difficulty: A corpus study. J Med Internet Res. 19(3):e59.
14. Kandula S, Zeng-Treitler Q (2008) Creating a gold standard for the readability measurement of health texts. AMIA Annu Symp Proc. 2008:353-357.
15. Lenzner T (2014) Are readability formulas valid tools for assessing survey question difficulty? Sociol Methods Res. 43(4):677-698.
16. Mortensen K, Hughes TL (2018) Comparing Amazon's mechanical turk platform to conventional data collection methods in the health and medical research literature. J Gen Intern Med. 33(4):533-538.
17. DuBay WH (2006) The classic readability studies. Costa Mesa, CA: Impact Information.
18. Rosemblat G, Logan R, Tse T, Graham L (2006) Text features and readability: Expert evaluation of consume health text.
19. Treitler QZ, Kim H, Goryachev S, Keselman A, Slaughter L, et al. (2007) Text characteristics of clinical reports and their implications for the readability of personal health records. Stud Health Technol Inform. 129:1117-21.
20. Dale E, Tyler RW (1934) A study of factors influencing the difficulty of reading materials for adults of limited reading ability. The Library Quarterly. 4(3):384-412.
21. Burke V, Greenberg D (2010) Determining readability: How to select and apply easy-to-use readability formulas to assess the difficulty of adult literacy materials. Adult Basic Edu Literacy Journ. 4(1):34-42.
22. Ley P, Florio T (1996) The use of readability formulas in healthcare. Psych Health Med. 1(1):7-28.
23. Friedman DB, Goetz HL (2006) A systematic review of readability and comprehension instruments used for print and web-based cancer information. Health Educ Behav. 33(3):352-373.
24. Badarudeen S, Sabharwal S (2010) Assessing readability of patient education materials: Current role in orthopaedics. Clin Orthop Relat Res. 468(2):2572-2580.
25. Brega AG, Cifuentes M, Barnard J, Mabachi NM, Albright K, et al. (2015) Guide to implementing the health literacy universal precautions toolkit. Agency for healthcare research and quality.
26. Centers for Medicare and Medicaid Services (2023) Guidelines for effective writing.
27. Plain Language Action and Information Network (2011) Federal plain language guidelines.