ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 2, No. 1, pp. 43-53, January 2012

doi:10.4304/tpls.2.1.43-53

Readability of Texts: State of the Art

Mostafa Zamanian
Department of Foreign Languages, Shiraz Branch, Islamic Azad University, Shiraz, Iran

Pooneh Heydari
Department of Foreign Languages, Shiraz Branch, Islamic Azad University, Shiraz, Iran
Email: pooneh.heydari@yahoo.com

***Abstract*—In TEFL, it is often stated that communication presupposes comprehension. The main purpose of readability studies is thus to measure the comprehensibility of a piece of writing. In this regard, different readability measures were initially devised to help educators select passages suitable for both children and adults. However, readability formulas can certainly be extremely helpful in the realm of EFL reading. They were originally designed to assess the suitability of books for students at particular grade levels or ages. Nevertheless, they can be used as basic tools in determining certain crucial EFL text-characteristics instrumental in the skill of reading and its related issues. The aim of the present paper is to familiarize the readers with the most frequently used readability formulas as well as the pros and cons views toward the use of such formulas. Of course, this part mostly illustrates studies done on readability formulas with the results obtained. The main objective of this part is to help readers to become familiar with the background of the formulas, the theory on which they stand, what they are good for and what they are not with regard to a number of studies cited in this section.**

***Index Terms*—comprehensibility, text variables, readability, readability formulas, validity of readability formulas**

I. Introduction

In order to best understand the importance of readability, it is important first to provide a definition for it. According to Richards, *et al.* (1992, p.306), readability means: "how easily written materials can be read and understood. This depends on several factors including the average length of sentences, the number of new words contained, and the grammatical complexity of the language used in a passage."

The creator of the SMOG readability formula, Harry McLaghlin (1969), defined readability as, "the degree to which a given class of people find certain reading matter compelling and comprehensible."

Generally, Dale and Chall's (1949) definition may be the most comprehensive: "The sum total (including all the interactions) of all those elements within a given piece of printed material that affect the success a group of reader have with it. The success is the extent to which they understand it, read it at an optimal speed, and find it interesting."

Therefore, a reader-text mismatch (for example, assigning a selection from the unabridged "Othello" for a 3rd grade reading exercise) can result in the user failing to use or ignoring the text. To avoid mismatch, educators would like a tool to check if a given text would be readable by its intended audience. Inventing such tools has been the primary focus of readability research for the past 90 years (Kondru, 2006).

To this end, readability formulas were originally created to predict the reading difficulty associated with text. To put it in another way, "a readability formula is an equation that gives an estimate of the readability of a text. The estimate is generally in terms of the number of years of education one needs to have to comprehend that text" (Kondru, 2006, p.7). Ability to predict text readability is useful because it helps educators select appropriate texts for students and authors write texts accessible to the audience they target.

All in all, readability studies are concerned with ensuring that a given piece of writing reaches and affects its audience in the way that the author intends. Indeed, as Seaton (1975 cited in Rezaei, 2000) asserts, communication presupposes comprehension, but the increasing variety, volume, and complexity of written materials make understanding more and more of a problem. Therefore, readability studies concentrate on the linguistic factors, in particular, word length and sentence length. In other words, the main purpose of readability studies is, in fact, to measure the comprehensibility of a piece of writing.

II. History of Readability

As Edgar Dale (1972 cited in Mosenthal & Kirsch, 1998) stated: "Readability is as old as the hills and the written stories that have described them." Of course, the notion of readability that he referred to was, "the ease of understanding based on an author's style of writing and organization of his/her ideas" (p.638).

The earliest investigations of readability were conducted by asking students, librarians, and teachers what seemed to make texts readable. It is convenient to locate the beginnings of the classical tradition of readability assessment in the

1921 publication of Thorndike's *Teachers' Work Book*, which provided a means for measuring the difficulty of words. Thorndike tabulated words according to the frequency of their use in general literature. It was assumed that words that were encountered frequently by readers were less difficult to understand than words that appeared rarely. Of course, familiarity breeds understanding. Thorndike's (1921) book was the first extensive listing of words in English by frequency. Other word lists and reading lessons were adapted to measure word difficulty later on. In fact, the knowledge of words has always been a strong measure of a reader's development in reading comprehension performance. As Chall and Dale (1995, p.84) wrote: "It is no accident that vocabulary is also a strong predictor of text difficulty."

Klare (1968), reviewing the research on word frequency, also concluded:

Not only do humans tend to use some words much more often than others, they recognize more frequent words rapidly than less frequent, prefer them, and understand and learn them more readily. It is not surprising therefore, that this variable has such a central role in the measurement of readability (p.12).

In addition to word factors, sentence length was also studied in the 1920s and became another factor included in the study of readability. Another contemporary of Thorndike, the psychologist Kitson (1921), published *The Mind of the Buyer*, in which he showed how and why readers of different magazines and newspapers differed from one another. He found that sentence length and word length, as measured by syllables, were good indicators of readability. He confirmed his theories through the analysis of newspapers and magazines.

Later, Kitson's claim was confirmed by other researchers and experts. As Catalano (1990), in his study, stated: "Readability and writing experts say sentence length is an appropriate gauge of difficulty because it measures relationships" (p.98).

Though such studies developed no readability formulas, they took the initial important steps which were to lead to the development of readability formulas. Since those early beginnings, the linguistic indicators of word and sentence length have remained the main factors of modern readability formulas which have been used extensively to classify reading materials. As Kirkwood and Wolfe (1980) declared, "Readability formulas contain a measure of vocabulary load and sentence length."

The credit for attempting to devise the first-ever readability formula goes to Lively and Pressey in 1923. They were concerned with the practical problem of selecting science textbooks for junior high school. The books were so overlaid with technical words that teachers spent all class time teaching vocabulary. Lively and Pressey (1923) argued that it would be helpful to have a way to measure the vocabulary burden of textbooks. They related the difficulty of a word to its frequency and attempted to develop a method for measuring vocabulary in textbooks as well as other reading materials to be used for school. Lively and Pressey assumed that the more common the word, the easier it is to understand. The lively and Pressey (1923) method was not a suitable instrument for measuring readability because they could not provide a scale to interpret the scores; but, their study marked the beginning of work on readability formulas that would continue unabated until the present time.

Readability formula is an analytical way to predict readability (Kondru, 2006). There are many readability formulas used to measure the readability level of the written materials, but some of them are better known and more popular. Popular readability formulas are based on extensive research and as Kondru (2006) implied, "their predictions correlate very well with the results of the actual readability measurements of expert judgments, comprehension tests, and the cloze procedures" (p.9).

In this section, some of the popular readability formulas are presented. Perhaps, the most common and the most publicized readability formula was credited to Rudolph Flesch (1948). The popularity of his formula made Flesch a leading authority on readability. Flesch Reading Ease Readability Formula (1948) is also used in Microsoft Office Word. Today, readability evaluation can be performed by computer. As such, most grammar or editing software today can determine the readability level of written materials. After Microsoft Office Word finishes the spell- and grammar-check, it can display the information about the readability level of the passage. Each readability rating is based on the average of the number of syllables per word and words per sentence. Flesch Reading Ease Readability Formula rates texts on a 100-point scale; the higher the score, the easier it is to understand the document. Most standard passages have approximately a readability score of 60 to 70.

The Flesch Reading Ease Readability Formula is:

$206.835 – (1.015 \times ASL) – (84.6 \times ASW)$

Where, ASL is the Average Sentence Length (the number of words divided by the number of sentences), and ASW is the Average of Syllables per Word (the number of syllables divided by the number of words).

As it was mentioned, in Flesch formula, the score ranges from 0 to 100, with 0 corresponding to the highest reading difficulty and 100 corresponding to the lowest reading difficulty. Table 1 provides interpretation of the Flesch Reading Ease Score.

TABLE 1:
FLESCH READING EASE SCORE

Reading Ease Score	Description	Predicted Reading Grade	Estimated Percentage of U.S. Adults
0-30	very difficult	college graduate	4.5%
30-40	difficult	college grade	33%
50-60	fairly difficult	10^{th}-12^{th} grade	54%
60-70	standard	8^{th}-9^{th} grade	83%
70-80	fairly easy	7^{th} grade	88%
80-90	easy	6^{th} grade	91%
90-100	very easy	5^{th} grade	93%

Another popular readability formula is the Dale-Chall (1948 cited in DuBay, 2004) Formula. The original Dale-Chall Formula was developed for adults and children above the 4^{th} grade level. They designed it to correct certain shortcomings in the Flesch Reading Ease Formula. It was a sentence-length variable plus a percentage of hard words—words not found on the Dale-Chall long list of 3000 easy words, 80 percent of which are known to fourth-grade readers.

The Dale-Chall Raw Score is given by,

Raw Score= 0.1579 PDW + 0.496 ASL + 3.6365

Raw Score=reading grade of a reader who can answer one half of the test questions on a passage,

PDW=Percentage of Difficult Words (words not on the Dale-Chall word list), and

ASL=Average Sentence Length in Words.

Raw Score is converted to school grade intervals using the conversion scheme shown in Table 2.

TABLE 2:
DALE-CHALL RAW SCORE TO GRADE INTERVAL CONVERSION

Raw Score	Grade Interval
4.9 and below	4th grade and below
5.0 - 5.9	5th – 6th grade
6.0 - 6.9	7th – 8th grade
7.0 - 7.9	9th – 10th grade
8.0 - 8.9	11th – 12th grade
9.0 - 9.9	Grade 13 through 15 (college)
10 and above	Grade 16 and above (college graduate)

After that, in *The Technique of Clear Writing*, Gunning (1952) published a readability formula developed for adults, the Fog-Index, which became popular because of its ease of use. It uses two variables, average sentence length and the number of words with more than two syllables for each 100 words.

Grade Level= 0.4 × (Average Sentence Length + Number of hard words)

Where:

A hard word is defined as a word that is more than two syllables long.

The Gunning's Fog-Index is shown in Table 3.

TABLE 3:
GUNNING'S FOG-INDEX

	Fog-Index	Estimated Reading Grades
	17	College graduate
	16	College senior
	15	College junior
	14	College sophomore
Danger line	13	College freshman
	12	High school senior
	11	High school junior
	10	High school sophomore
Easy	9	High school freshman
Reading	8	Eighth grade
Range	7	Seventh grade
	6	Sixth grade

The publication of the Flesch, Dale-Chall, and Gunning formulas conveniently marks the end of the first 30 years of classic readability studies.

DuBay (2004) claimed:

The authors of these formulas brought the issue of readability to public attention. They stimulated new consumer demands for documents in plain language. Finally, they stimulated new studies, not only on how to improve the formulas, but also on the other factors affecting reading success (p.25).

The new readability was a period of consolidation and deeper study. Investigators sought to learn more about how the formulas work and how to improve them. In the 1960s, several other developments accelerated the study of readability. One of the earliest investigations of that era was conducted by Fry (1968).

Fry (1968) created one of the most popular readability tests that used a graph. It was suitable for all ages, from infant to upper secondary. Estimation of text readability using the Fry Readability Graph is described in the following algorithm:

1. Select samples of 100 words from the text.
2. On the y (vertical) axis of the Fry Graph, plot the average sentence length of the sample.
3. On the x (horizontal) axis of the Fry Graph, plot the average word length.
4. The zone on the graph that includes a point (corresponding to a sample) shows the grade score associated with that sample. Take grade scores associated with at least three points on the graph and average them to get the average grade level associated with the entire text. Scores that appear in the shaded areas are invalid.

The Fry Graph is shown in Figure 1.

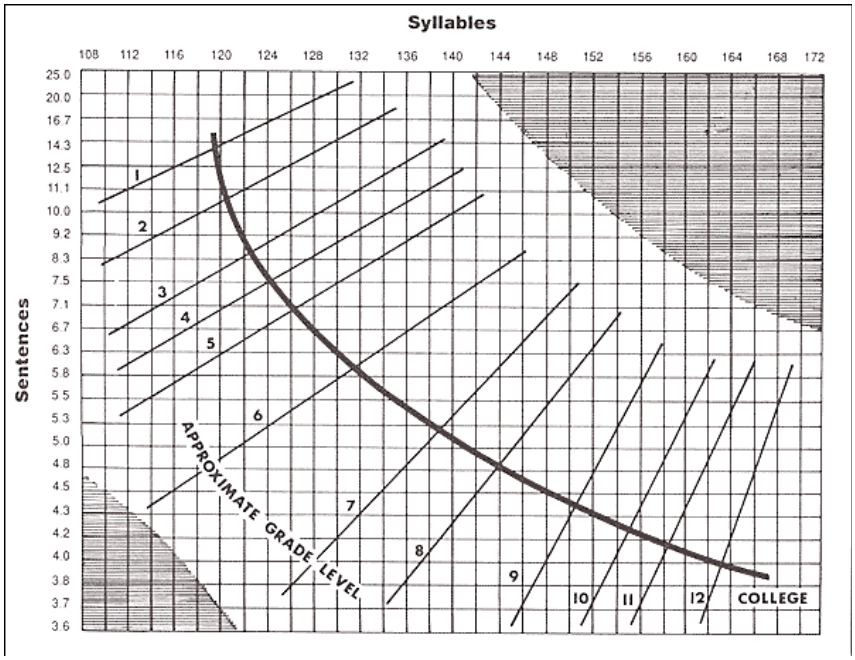

Figure 1: Fry Graph for estimating Reading Ages (in years)

After that, G. Harry McLaughlin (1969) published his SMOG (Simple Measure Of Gobbledygook) formula in the belief that word length and sentence length should be multiplied rather than added. By counting the number of words of more than two syllables (polysyllable count) in 30 sentences, he provided this simple formula:

SMOG Grading = 3 + square root of polysyllable count

Another known readability formula, the Flesch-Kincaid Formula (1975 cited in Greenfield, 1999), is a recalibration of the original Flesch Formula. It rates text on a U.S. grade school level. For example, a score of eight means that an eight grader can understand the document. For most documents, the writers aim for a score of approximately 7.0 to 8.0.

The formula for the Flesch-Kincaid Grade Level Score is:

$(0.39 \times \text{ASL}) + (11.8 \times \text{ASW}) - 15.59$

Where:

ASL is the Average Sentence Length (the number of words divided by the number of sentences), and ASW is the Average of Syllables per Word (the number of syllables divided by the number of words).

It converted the Reading Ease Score to a U.S. Grade School Level. In general, as Graesser *et al.* (2004, p.199) declared: "a text should have more than 200 words before the Flesch Reading Ease and Flesch-Kincaid Grade Level can successfully be applied."

Finally, one synthesis of the advances in readability studies has been the creation of Coh-Metrix (Graesser, McNamara, Louwerse, and Cai, 2004), a computational tool developed at the University of Memphis that measures cohesion and text difficulty at various levels of language, discourse, and conceptual analysis. This tool was designed with the goal of improving reading comprehension in classrooms by providing a means to improve textbook writing and to more appropriately match textbooks to the intended students (Graesser *et al.*, 2004; McNamara *et al.*, 2002). Coh-Metrix enhances conventional readability measures like Flesch-Kincaid and Flesch Reading Ease by providing detailed language and cohesion features, which will eventually match this textual information to the background knowledge of the reader (McNamara *et al.*, 2002). The system investigates lexicon, pattern classifiers, part of speech taggers, syntactic parsers, shallow semantic interpreters, and other components that have been developed in the field of computational linguistics (Jurafsky and Martin, 2002). After the user enters an English text, Coh-Metrix returns measures requested by the user. It analyzes text on over 200 measures of language, text, and readability including co-

referential cohesion, causal cohesion, density of connectives, latent semantic analysis metrics, and syntactic complexity. In addition, a facility allows the user to store the results of these analyses in data file (such as text, Excel, and SPSS).

Moreover, there are many other readability formulas devised to measure the readability level of text. Klare (1974) reported more than 40 different ones, most of which did not become widely adopted by the field. However, the differences in formulas are generally in weight assigned to different variables as well as the presence or absence of word lists (Wait, 1987). In other words, as Alderson (1986 cited in Rezaei, 2000, p.21) pointed out, "the rationale, construction and validity of these formulas are not very much different."

III. The Pros and Cons

The following part addresses itself to the pros and cons views toward the use of readability formulas.

Cons View: Even thought readability formulas are being used more widely than ever before in schools, libraries, newspapers, business, and government, there are critics who attack their use.

As it was mentioned earlier, readability formulas were originally created for testing the readability level of school textbooks. Therefore, at the outset, they were mainly applied to evaluate the readability of textbooks. Given the influence readability formulas had over the reading materials offered to children and young adults in schools and libraries, a closer examination of the principles that underlie them was in order. Educators and researchers asked themselves: Do these principles really enable readability formulas to offer a sound, scientific way of evaluating the difficulty of texts?

To this end, many of the critics were concerned about the limitation of the formulas and argued for their disuse. One of the first studies conducted in this respect was by Kirkwood and Wolfe (1980). They found that because readability formulas are composed of the variables of words and sentence length, they correspond to the surface structure of a passage rather than the deep syntactic and semantic structure. Readability formulas do not address the interaction between the reader and the texts. Therefore, they are not consistent with the psycholinguistic theory of reading.

In addition to what Kirkwood and Wolfe (1980) stated, Bertram and Newman (1981) discussed three different weaknesses of readability formulas. The first concern involves the belief that most formulas consider only sentence length and word difficulty. Thus, they ignore factors such as cohesion, complexity of ideas, and required schemata. The second flaw is the lack of accountability of readers' specific factors such as interest and purpose for reading. The third concern is the lack of statistical back-up for most readability formulas.

Smith (1988) seriously overstated the case when asserted that, "readability formulas based on word counts and sentence length have been generally discredited" (p.239).

In another study, Bailin and Grafstein (2001) re-examined the linguistic criteria that form the basis of readability scores and argued that the criteria commonly used in readability formulas do not constitute a satisfactory basis for assessing reading difficulty. In fact, the developers of readability formulas treated the issue of readability as if it were a monolithic phenomenon. Their underlying assumption was that how easy a text is to read is always based on the same criteria that is measurable by a statistical formula, and that it is reducible to a score returned by that formula. Bailin and Grafstein (2001) believed, on the other hand, that there is no single, simple measure of readability. They emphasized that how easy a text is for an individual to read is the result of the interaction of a number of different factors, reflecting properties both of texts and readers and the interaction between them.

Another problem with the formulas underlying principles is that they consistently failed to tell us about text comprehension. Chall (1958) and Klare (1963) defined readability in terms of writing style. They agreed that readability is the result of writing style that is legible, interesting and comprehensible. Moreover, Richards *et al.* (1992, p.306) defined readability as, "how easy written materials can be read and understood." The definition of readability as seen by Chall (1958), Klare (1963), and Richards *et al.* (1992) focused on comprehension. A fundamental assumption of the psycholinguistic model of reading is that comprehension is the primary goal of the reader (Smith, 1982), but "readability formulas do not address the phenomenon of comprehension" (Wait, 1987, p.13).

In this respect, Dreyer (1984) claimed that, "the focus of readability formulas on number of syllables and mean sentence length ignores the textual features that affect comprehension" (p.335). In particular, readability formulas disregard what linguists refer to as whole-text aspects (Schriver, 1989). Whole–text aspects are concerned with the positioning and organization of sentences and paragraphs in texts and with how information flows through the text.

In addition, a critical position argued that readability formulas do not measure a number of factors associated with readability and that they do not measure understandability or comprehension (Jones & Shoemaker, 1994).

It appears then that readability formulas due to their underlying principles are imperfect predictors of text readability and understandability or comprehension. Other limitations of the formulas will be described later on.

Educators and authors are often charged with the formidable task of selecting and writing reading material for individuals of widely diverse reading skills. What should an individual in a particular grade read? How difficult or easy should the texts be? Which texts should be suggested to advanced readers? Which texts should poor readers in a particular grade be encouraged to read? Of course, these are not easy questions and it is no surprise that many educators and authors have welcomed and endorsed tools that claim to use objective criteria to make selections and recommendations. However, some critics argue that readability formulas cannot prove to be valuable tools for producing, revising, and selecting written materials.

Frase, Rubin, Starr, and Plung (1981) found that readability formulas fail to meet expectations of expanded use due to their ignoring or violating much current knowledge about reading, shaky statistical basis, and inappropriateness as practical tools for matching children and texts or for providing guidelines for writers. Accordingly, the results demonstrated that in most uses, readability formulas violate the basic assumptions on their applicability.

Davison (1981 cited in Hewitt & Homan, 2004) contended that if a text is being rewritten or revised to match a particular level of reading ability, the changes may be made based on readability, not content. Davison stated that changes should be made because of inherent difficulty or problems of ambiguity not just to influence the score a text might receive from a readability formula.

Powell, Barry, and Redish (1981) shared the same concern expressed by Davison, when stated, "despite its aid, a readability formula cannot place good training in how to write clear, well-organized, audience-focused material" (p.43).

Davison and Kantor (1982), who studied specific changes in SRA laboratory materials, found rewritten materials sometimes more difficult to comprehend than the original text. They believed that rewritten materials were most successful when the writer was not trying to fit the text to a formula.

Wright (1982), using readability formulas in his study, rewrote biology materials in an attempt to reduce reading difficulties of 265 ninth and tenth graders. After four weeks of reduced materials, no significant differences in achievement were apparent.

Besides, some credible organizations like the International Reading Association and the National Council for Teachers of English had put out warnings on readability formula use. They did not just question the use of readability formulas but actively disparaged their use. For example, Goodman (1986 cited in Fry, 1989) speaking for the Reading Commission of NCTE stated:

Readability formulas used in selecting and rewriting materials do not produce appropriate, readable texts. In fact, tinkering with texts to produce acceptable levels may turn them into texts which are harder to read. That's why Sheila Fitzgerald as President of NCTE and Bernice Cullinan as President of IRA have issued a joint statement calling for the abandonment of use of readability formulas in preparing and choosing school texts (p.292).

So, it seems that although writers can fit their text to formula variables, this alone does not make material more readable.

While it is evident from the above inventory of the literature that readability formulas based on their underlying assumptions are discredited, it is certainly expected that classical readability assessment has come under increasing criticism. Educators and researchers asked themselves: Do they predict precisely enough how readable a given piece of writing will be for a given audience? Do they evaluate precisely enough how well the reader will understand the ideas in the texts? Are they an effective measure of text-readability for EFL learners? These questions are inextricably linked with the concept of validity. Thus, educators and researchers tried to re-evaluate the validity of readability formulas.

One of the first studies conducted in this respect is by Froese (1971). In his study, the validity of the Dale-Chall readability ratings for sixth-grade science textbooks when compared to an independent criterion of language difficulty expressed in cloze units was examined. Three hundred and sixty six sixth-graders participated in this study. Finally, he concluded that the Dale-Chall readability formula is not a valid measure of sixth-grade science textbook materials when the cloze procedure is used as a criterion.

In another study, Rezaei (2000) re-considered the validity of one popular readability formula — the Fog-Index of Readability. To do so, first some reading passages were selected and a number of multiple-choice items were devised for each passage. The text was administered to the participants and the data were collected. Then, the passages were respectively made more difficult or easier by increasing or decreasing the number of sentences so that the readability of the passages was changed. The second version of the text was also given to the subjects and the results were compared with those of the original version. Finally, the analysis of data obtained from the revised test revealed that the participants in the study did not find the passages significantly different. As a result, Rezaei (2000) concluded that this formula is not at all sensitive to difficulty/ease level of reading passages when their readability index is changed up to a certain degree.

Ardoin, Suldo, Witt, Aldrich, and McDonald (2005) in their study, re-examined the validity of eigth readability formulas — Spache, Dale-Chall, Fry, Flesch-Kincaid, Fog, Powers-Summer-Kearl (PSK), Smog, and Forcast — to determine how well they predict passage difficulty in R-CBM research. Curriculum-based measurement in reading (R-CMB) is an assessment procedure in which students read passages and the number of words read correctly in one minute (WRCM) is recorded. R-CBM was placed on the approved list by the Reading First Assessment Committee (Kame'enui, 2002 cited in Ardoin, *et al.*, 2005) meaning that it meets or exceeds standards of psychometric accuracy. R-CBM is an invaluable tool for many practitioners while its gold standard for determining passage difficulty is the readability formulas. Ardoin *et al.* (2005) investigated the validity of eight formulas most commonly employed in R-CBM research. The study was based on the premise that as grade levels assigned to passages by readability formulas increase, students' WRCM should decrease. Results indicated a modest relationship between reading fluency and passage difficulty as indicated by the eight readability formulas. Ardoin *et al.* found that the formulas most commonly employed in R-CBM research were the poorest predictors. It follows from what Ardoin *et al.* found that efforts beyond the use of currently published readability estimates are needed in order to procure equivalent forms for R-CBM research.

Reviewing the research on readability formulas, one might conclude that the domain and discourse of readability research has been almost entirely limited to native language reading, and mostly native English. At the same time, the newer theories and discoveries about the complex interacting factors operating on reading difficulty, furnish ample reasons to be suspicious of the validity of readability formulas derived from a native English population sample for use in an EFL context. In this case, Drury (1985) remarked that, "traditionally used readability formulas have drawbacks, especially when used with non-fluent users of English" (p.11).

Carrell (1987) elsewhere discussed both the importance of an accurate readability measures for EFL learners and the fault of shallow-based readability formulas such as Flesch Reading Ease formula, Fog-Index of readability, and Fry Grade Level. According to Carrell (1987), shallow-based readability formulas work to a degree for first language learners because they are developed from statistical formulas and intended for large samples of text. Carrell's major criticisms of using traditional readability formulas for L2 texts were their failure to work for smaller student populations, reader abilities, and text passages.

Indeed, only a few research articles have appeared dealing specifically with English L2 readability. The most significant research on EFL readability to date has been done by Brown (1998). In order to test the reliability of traditional readability formulas for second language learners, Brown (1998) used cloze procedures on fifty randomly selected library passages and over 2300 Japanese EFL students. Finally, he realized that readability formulas developed for L1 readers were likely not appropriate for L2 readers. Brown (1998) not only researched the effectiveness of L1 readability formulas for L2 readers, but also developed and designed an alternative L2 readability measure. Brown's criticism of L1 readability formulas for L2 readers was based on the idea that L1 readability formulas did not account for reader-based variables such as: language differences, education, age, or learning styles. He contended that readability formulas designed specifically for L2 readers should include the type, function, and frequency of words as well as word redundancy. Brown's EFL Readability Index was a small subset of variables that included the average number of syllables per sentence, the average frequency that the cloze item tested appeared elsewhere in the text, the percentage of words over 7 letters, and the percent of function words in the text. This EFL Readability Index, while not a precise estimate of readability, did have a high degree of association and accounted for more variance in L2 learners than traditional readability formulas.

Brown's EFL Readability Index is:

38.7469 + (.7823 x syllable/sentence)
+ (-126.1770 x passage frequency)
+ (1.2878 x Percent Long Words)
+ (.7596 x Percent Function Words)

One more criticism of the formulas must be treated here, and that is the discrepancy between the scores of different formulas. This problem had long been perplexing. Critics have often cited such discrepancies as indications of the lack of precision of the formulas. Kern (1979) argued that, "the discrepancies among the Kincaid and Caylor formulas deprived them of usefulness."

Chen (1986), in his study, compared the results of the readability formulas over the same textbooks in order to offer guidance for use of the formulas. Eleven frequently recommended elementary social studies textbooks were included in his study. An Nth name sampling technique was employed to randomly select thirty 100-word passages from each textbook. Then, seven readability formulas — Dale-Chall, Flesch, Fry, Fog, SMOG, Spache, and Power-Sumner-Kearl — were applied to the passages. The results showed (1) there was no universal agreement among the formulas as to the rank ordering of textbook difficulty, and (2) there were wide discrepancies among formulas that resulted in the same textbook being rated several grade levels apart. Hence it appears that the range of scores provided by different formulas remind us once more that they are not perfect predictors.

Pros View: Although critics have been arguing that readability formulas are not accurate or useful measures of the difficulty of texts, it does not mean that they do not measure anything. However, some educators and researchers not only do not defend the misuse of the formulas, but also employ them in order to control for their passage difficulty. Klare (1980) considered readability formulas; though far from perfect, far more accurate than human judgment. Of course, research in this area may yield fruitful results and as Bailin and Grafstein (2001, p.299) stated, "may well allow us to assess the difficulty of at least certain aspects of texts present for certain readers."

Existing readability formulas are based on countable aspects of the text such as average sentence length and average word length. While critics referred to these principles as the most important cause of formulas' limitations, Klare (1974) stated:

Unless the user is interested in doing research, there is little to be gained from choosing a highly complex formula. A simple 2 variable formula should be sufficient, especially if one of the variables is a word or semantic variable and the other is a sentence or syntactic variable (p.63).

Kintsch (1979), who is not particularly a fan of traditional readability formulas, shared the same concern when found that most of text difficulty could be accounted for by two factors, reinstatement searches, and traditional word frequency.

The two specific characteristics that are evident in the various readability formulas are: (1) the emphasis on how easy a text is to understand, and (2) the emphasis on quantification. These emphases have made readability a particularly

attractive concept for educators. The attractiveness of using formulas to measure readability lies in the belief that, in principle, they objectively and quantifiably evaluate the difficulty of written material without measuring characteristics of readers. Moreover, a readability formula can return a numerical score, giving the user the sense of knowing the precise level of difficulty of a text (Bailin & Grafstein, 2001).

Focusing on the issue of formulas' underlying principles, DuBay (2004) remarked that, "the variables used in the readability formulas show us the skeleton of a text" (p.61). He maintained that it is up to us to flesh out that skeleton with tone, content, organization, coherence, and design. In fact, readability formula has made us very aware of what we write at the level of words and sentences (Hargis, 2000).

Besides, readability formula can predict comprehension, oral reading errors, and inclination to continue reading. As Klale (1948 cited in Fry, 1989) stated, "their prediction ability is at least as good as reading tests, IQ tests, or more other psycho educational measures" (p.294).

Another limitation of readability formulas, in the opinion of some critics, is that they are not valuable tools for producing and revising written materials. In contrast, McClure (1987) emphasized that, "a readability formula is an evaluation tool, not a reading or writing tool" (p.12).

Fry (1989) neatly stated, "Readability formulas are not writability formulas" (p.293). He believed that they are not and have never been intended to be writer's guides. In his opinion, critics, bad writers, and lazy editors might blame readability formulas for poor quality textbooks, but this is not the formulas' fault.

Besides, some other reading researchers and technical communicators paid attention to the efficiency of readability formulas. Since the goal of technical writers is accurate and efficient communication, readability is one of their biggest concerns. Thus, various readability formulas have become the subject of technical writers' interest (Sharma, 1982).

For example, Powel, Barry, and Redish (1981, p.43) argued, "When used with understanding, readability formulas can be helpful to writers." Powel *et al.* (1981) asserted that through counting adaptation to the needs of publishers and educators, readability programs have become user-oriented and can be run on a variety of computers and in many other high-level computer languages. Powel *et al.* also maintained that as the current national trend toward writing for easy understanding builds momentum and extends to the technical fields, computerized readability analysis can be a convenient and vital aid to the generation of clear, understandable written material.

Connaster (1999) considered readability formulas as the subject of technical writers' interest because they are used to equalize the reading difficulty of texts used in experiments.

One more criticism of formulas must be addressed, and that is related to their validity. Unlike critics' viewpoint, there is much evidence, both old and new, that readability formulas are indeed valid. Fry (1989, p.295), who is particularly a fan of readability formulas, reported some of the validity measures that formulas are correlated with as follows:

1. comprehension assessed by traditional multiple choice questions,
2. comprehension assessed by cloze passages,
3. oral reading errors,
4. readership (a journalism concept pertaining to the number of readers for a particular article),
5. subvocalization,
6. eye-voice span,
7. function chaining (how many words a typist continues to type after the copy page is covered),
8. controlled subjective judgment, and
9. concurrent validity (formulas correlate with each other).

Focusing on the issue of validity, some researchers and educators examined the appropriateness of traditional readability formulas for second language learners.

Hamsik (1984) conducted the first validation study to examine whether readability formulas developed for the measuring of reading difficulty for native English readers are applicable to the measuring of ESL readability. The specific purpose of her study was to determine if four widely used readability formulas—the Flesch formula, the Dale-Call formula, the Fry Graph, and the Lorge formula—measure readability difficulty for ESL students. The subjects were forty Intensive English Center students at intermediate to advanced levels of English reading proficiency as measured by the TOEFL. From data analysis, it was found that a correlation does exist between the rank order of the passages as measured by the cloze scores of the EFL students and by the readability formulas. Furthermore, this correlation was meaningful. According to the data of the sample, it now seems possible to state that the four mentioned readability formulas do measure readability of text for ESL students. So, they can be used to select materials appropriate to the reading level of ESL students.

Elsewhere, Brown's (1998) aforementioned study had been criticized by Greenfield (1999), who, in an attempt to replicate Brown's study, used a corpus of 31 academic texts that had been evaluated for textual difficulty by administering cloze tests to L1 speakers (Bormuth, 1971 cited in Greenfield,1999). Using this corpus, and comparing L2 learners' performance to L1 readers' performance under a constant passage set, Greenfield found that traditional L1 readability formulas such as Flesch Reading Ease and Flesch-Kincaid Grade Level formula had strong correlation to L2 cloze test performance. Using statistical methods similar to Brown's, Greenfield also constructed an EFL Readability Index that was similar to traditional L1 readability formulas, but scaled for EFL learners. Greenfield's formula, called the Miyazaki EFL Readability Index is:

164.935 – (18.792 × letters per word) – (1.916 × words per sentence)

As stated before, some critics pointed to the discrepancies between the scores of different formulas as indications of the lack of precision, whereas DuBay (2004) asserted what these critics ignore are, "the correlations of the formulas with comprehension texts" (p.60). He maintained what is important is not how the formulas agree or disagree on a particular text, but their degree of consistency in predicting difficulty over a range of graded texts.

IV. Conclusion

If any conclusion is possible to draw from the studies reviewed on readability formulas, it is that there are two opposite views toward the use of readability formulas. Both of these two views have been advocated by different researchers and there is enough empirical evidence for each to be true. Thus, it can be stated that the formulas have both advantages and disadvantages.

Advantages of using readability formulas:

a) By definition, readability formulas measure the grade-level readers must have to read a given text. The results from using readability formulas provide the writer of the text with much needed information to reach his target audience.

b) Readability formulas do not require the readers to first go through the text to decide if the text is too hard or too easy to read. By readability formulas, one can know ahead of time if his readers can understand the material. This can save time, money and energy.

c) Readability formulas are text-based formulas; many researchers and readers find them easy to use.

d) Today, readability formulas can be performed by computer. As such, most grammar or editing software today can determine the readability level of written materials.

e) Readability formulas help writers convert their written material into plain language.

Disadvantages of using readability formulas:

a) Unfortunately, readability formulas are not of much help if one wants to know how well the target audience understands the text.

b) Due to many readability formulas, there is an increasing chance of getting wide variation in results of a same text.

c) Readability formulas cannot measure the context, prior knowledge, interest level, difficulty of concept, or coherence of text.

In connection to the effect of readability level on comprehension, TEFL educators and researchers interested in seeking the truth through research should re-evaluate the validity of readability formulas.

References

[1] Ardoin, S. P., Suldo, S. M., Witt, J., Ardrich, S., & McDonald, E. (2005). Accuracy of readability estimates' predictions of CBM performance [Electronic version]. *School Psychology Quarterly, 20*(1), 1-22. Retrieved December 23, 2007, from http://proquest.umi.com/pqdweb?did=818687841&sid=2&Fmt=4&clientId=46449&RQT=309&VName=PQD

[2] Bailin, A., & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: A critique. *Language & Communication, 21*, 285-301.

[3] Bertram, B., & Newman, S. (1981). Why readability formulas fail (Report No. 28). Illinois University, Urbana: Center for the Study of Reading. (Eric Document Service No. ED205915)

[4] Brown, J. D. (1998). An EFL readability index. *JALT Journal, 20*, 7-36.

[5] Carrell, P. (1987). Readability in ESL. *Reading in a Foreign Language, 4*, 21-40.

[6] Catalano, K. (1990). On the wire: How six news services are exceeding readability standards. *Journalism Quarterly, 67*(1), 97-103.

[7] Chall, J. S. (1958). Readability: An appraisal of research and application. Columbus, OH: The Ohio State University Bureau of Educational Research.

[8] Chall, J. S., & Dale, E. (1995). Readability revisited, the new Dale-Chall readability formula. Cambridge, MA: Bookline Books.

[9] Chen, W. S. (1986). A comparison of seven computerized readability formulas as applied to elementary social studies textbooks. Unpublished doctoral dissertation, Columbia University College, U.S. Retrieved November 10, 2008, from http://proquest.umi.com/pqdweb?did=771892981&sid=9&Fmt=4&clientId=46449&RQT=309&VName=PQD

[10] Connaster, B. R. (1999). Last rites for readability formulas in technical communication [Electronic version]. *Journal of Technical Writing and Communication, 29*(3). Retrieved December 23, 2007, from http://proquest.umi.com/pqdweb?did=46774455&sid=2&Ftm=2&clientId=46449&RQT=309&VName=PQD

[11] Dale, E. & Chall, J. S. (1949). The concept of readability. *Elementary English*, 26, 23.

[12] Davison, A., & Kantor, R. N. (1982). On the future of readability formulas to define readable texts. *Reading Research Quarterly, 17*, 187-209.

[13] Dreyer, L. G. (1984). Readability and responsibility. *Journal of Reading, 32*, 334-338.

[14] Drury, A. (1985). Evaluating readability [Electronic version]. *IEEE Transactions on Professional Communication, PC28*(4), 11-15. Retrieved October 13, 2008, from http://proquest.umi.com/pqdweb?did=11355684&sid=3&Fmt=4&clientId=46449&RQT=309&VName=PQD

[15] DuBay, W. H. (2004). The principles of readability. Retrieved February 12, 2008, from http://www.nald.ca/fulltext/readab/readab.pdf

[16] Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221-233.

[17] Frase, L. T., Rubin, A., Starr, K., & Plung, D. L. (1981). Readability formulas: Used or abused? [Electronic version]. *IEEE Transactions on Professional Communication, PC24* (1), 48-54. Retrieved December 23, 2007, from http://proquest.umi.com/pqdweb?did=1135513&sid=9&Ftm=2&clientld=46449&RQT=309&VName=PQD
[18] Froese, V. (1971). Cloze readability versus the Dale-Chall formula. Paper presented at the Meeting of the International Reading Association, Atlantic City, NJ.
[19] Fry, E. B. (1968). A readability formula that saves time. *Journal of Reading, 11*, 513-516.
[20] Fry, E. B. (1989). Reading formulas – maligned but valid. *Journal of Reading, 32* (4), 292-297.
[21] Graesser, G. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers, 36*, 193-202.
[22] Greenfield, G. R. (1999). Classic readability formulas in an EFL context: Are theyvalid for Japanese speakers? Unpublished doctoral dissertation, University of Temple, U.S.
[23] Gunning, R. (1952). The technique of clear writing. New York: McGraw-Hill.
[24] Hamsik, M. J. (1984). Reading, readability, and the ESL reader. Unpublished doctoral dissertation, The Florida University, U.S. Retrieved December 23, 2007, from http://proquest.umi.com/pqdweb?did=749006291&sid=1&ftm=2&clientld=46449&RQT=309&VName=PQD
[25] Hargis, G. (2000). Readability and computer documentation. *ACM Journal of Computer Documentation, 24*(3), 122-131.
[26] Hewitt, M. A., & Homan, S. P. (2004). Readability level of standardized test items and student performance: The forgotten validity variable [Electronic version]. *Reading Research and Instruction, 43*(2), 1-16. Retrieved December 23, 2007, from http://proquest.umi.com/pqdweb?did=771892981&sid=9&Fmt=4&clientld=46449&RQT=309&VName=PQD
[27] Jones, M. J., & Shoemaker, P. A. (1994). Accounting narratives: A review of empirical studies of content and readability. *The Journal of Accounting Literature, 13*, 142-184.
[28] Jurafsky, D., & Martin, J. H. (2002). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Upper Saddle River: Printice-Hall.
[29] Kern, R. P. (1979). Usefulness of readability formulas for achieving army readability objectives: Research and state-of-the-art applied to the army's problem. Retrieved December 10, 2008, from http://www.rbainformationdesign.com.au/downloads/Readability%20Formulas.pdf
[30] Kintsch, W. (1979). On modeling comprehension. *Educational Psychologist, 14*, 3-14.
[31] Kirkwood, K. J., & Wolfe, R. G. (1980). Matching students and reading materials: A cloze-procedure method for assessing the reading ability of students and the readability of textual materials. Toronto: Ontario Department of Education. (ERIC Document Reproduction in Service No. ED 195 928)
[32] Kitson, H. D. (1921). The mind of the buyer. New York: Macmillan.
[33] Klare, G. R. (1963). The measurement of readability. Ames, IA: Iowa State University Press.
[34] Klare, G. R. (1968). The role of the word frequency in readability. *Elementary English, 45*, 12-22.
[35] Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly, 10*, 62-102.
[36] Klare, G. R. (1980). A manual for readable writing. Glen Burnie, MD: REM Company.
[37] Kondur, J. (2006). Using part of speech structure of text in the prediction of its readability. Unpublished master's thesis, University of Texas, Arlington, U.S. Retrieved November 13, 2008, from http://proquest.umi.com/pdqweb?did=1216761731&sid=1&Fmt=2&clientld=46449&PQT=309&VName=PQD
[38] Lively, B. A., & Pressey, S. L. (1923). A method for measuring the vocabulary burden of textbooks. *Educational Administration and Supervision Including Teacher Training, 9*, 389-398.
[39] McClure, G. M. (1987). Readability formulas: Useful or useless? [Electronic version]. *IEEE Transactions on Professional Communication, PC30*(1), 12-16. Retrieved December 23, 2007, from http://proquest.umi.com/pdqweb?did=1135713&sid=9&Fmt=2&clientld=46449&PQT=309&VName=PQD
[40] McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading, 12*(8), 639-646.
[41] McNamara, D. S., Louwerse, M. M., & Graesser, A. C. (2002). Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension. Unpublished Grant proposal, University of Memphis, Memphis, Tennessee.
[42] Mosenthal, P. B., & Kirsch, I. S. (1998). A new measure for assessing document complexity: The PMOSE/ IKIRSCH document readability formula. *Journal of Adolescent & Adult Literacy*, *41*(8), 638-658.
[43] Powell, K. B., Barry, J. G., & Redish, J. C. (1981). Readability formulas: Used or abused? [Electronic version]. *IEEE Transactions on Professional Communication*, *PC24*(1), 43-48. Retrieved December 23, 2007, from http://proquest.umi.com/pdqweb?did=1135512&sid=9&Fmt=2&clientld=46449&PQT=309&VName=PQD
[44] Rezaei, A. A. (2000). The validity of the "Fog-Index of Readability". *Journal of Humanities of Islamic Republic of Iran*, *7*(4), 17-27.
[45] Richards, J. C., Platt, J., & Platt, H. (1992). Longman dictionary of language teaching and applied linguistics. London: Longman.
[46] Schriver, K. A. (1989). Evaluating text quality: The continuum from text-focused to reader-focused method. *IEEE Transactions on Professional Communication, 32*(4), 238-255.
[47] Sharma, B. C. (1982). Graphic evaluation of readability scores [Electronic version]. *IEEE Transactions on Professional Communication, PC25*(1), 34-36. Retrieved December 23, 2007, from http://proquest.umi.com/pqdweb?did=1135550&sid=9&Fmt=2&clientld=46449&RQT=309&VName=PQD
[48] Smith, F. (1982). Understanding reading. New York: Holt, Rinehart & Winston.
[49] Smith, F. (1988). Understanding reading (4th ed.). Hillsdale, New Jersey: Erlbaum.
[50] Thorndike, E. L. (1921). The teacher's workbook. New York, NY: Teacher's College, Columbia University.
[51] Wait, S. S. (1987). Textbook readability and the predictive value of the Dale-Call, comprehensive assessment program, and cloze. Unpublished doctoral dissertation, University of Florida, U.S.

[52] Wright, J. D. (1982). The effect of reduced readability text materials on comprehension and biology achievement. *Science Education, 66*, 3-13.

Mostafa Zamanian Ph.D. (mostafazamanian@yahoo.com) is an Assistant Professor in IAU, Shiraz Branch. He teaches M.A. and Ph.D. classes and is interested in teaching methodologies and curriculum development. He got his Ph.D. from U.S.A.

Pooneh Heydari (corresponding author: Pooneh.Heydari@yahoo.com) is a lecturer at Shiraz Azad University. She is interested in L1 and L2 acquisition, syllabus design and materials development. She is a Ph.D. student of TEFL in IAU, Shiraz Branch, Iran.