Text Complexity and Readability Measures: An Examination of Historical Trends

Technical Report · December 2014

CITATIONS
0

READS
1,079

1 author:

Laura Clark Briggs
Middle Tennessee State University
6 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Text Complexity and Readability Measures: An Examination of Historical Trends

A Review of the Literature

Laura Clark Briggs

Middle Tennessee State University

Author Note

This paper was prepared for LITS 7100: Historical Trends, Issues, & Methodologies of Literacy, Fall 2014. Correspondence concerning this article should be addressed to Laura Clark Briggs, College of Education, Middle Tennessee State University, Murfreesboro, TN 37132. Email: lcc2a@mtmail.mtsu.edu

Abstract

In recent years, much attention has been given to the readability and complexity of texts. With the move toward the use of Common Core State Standards (CCSS) by many states, policymakers and educators have directed attention to the ability of American students to successfully navigate complex text. This concern with text complexity and readability is not new; in fact, the earliest American attempt to use quantitative measures to examine text began in the late 19th century. This review examines how readability has been gauged in the United States in the 19th and early- to mid 20th century and whether mathematical formulas to determine readability have remained constant over the years or whether they have been improved upon, augmented, or discarded in favor of other mathematical models. This review further examines catalysts behind the attention, research, application, and sense of urgency to make text readable to Americans: war and the nation's economic interests.

Text Complexity and Readability Measures: An Examination of Historical Trends

A Review of the Literature

A decrease in average sentence length and text complexity in textbooks, newspapers, and magazines during the twentieth century has been blamed for everything from decreased SAT scores to a "dumbing down" of our nation's youth (Adams, 2011). It is uncertain whether text complexity and readability are major factors for the decline in average SAT scores over the years; it is equally uncertain whether the fact that printed material is less complex now than in generations past correlates to poorer performance on benchmark indicators of academic competences as compared to previous generations (Adams, 2011. Bejar & Blew, 1981). Additionally, there is no consensus of agreement on a definition of readability and text complexity, as the definition has been inconsistent over time. This paper examines the historical definitions of readability and text complexity from its beginnings in the late 19th century through the mid-20th century. An initial examination of early indicators of readability is necessary. This includes providing historical background surrounding readability research in the United States and how this research has evolved over time.

The Origins of Readability, Readability Formulas, and Measures of Complexity

19th Century

Lucius Sherman (1893), in his book *The Analytics of Literature*, examined sentence length and decrease in predication [length and complexity] (predication referring to the part of the sentence that contains the verb) in an effort to study how these characteristics had changed over time. He showed that writers were reasonably consistent in their sentence length. Sherman's systematic approach was applied to writers from the pre-Elizabethan and Elizabethan eras: these compositions, Sherman said, were characterized by writing that was composed of

sentences that were often laboriously painful to read. He added that, due to the heavy, weighty feel of the sentences, it was common and often necessary for the reader to reread a sentence several times to determine the meaning.

Beginning with Robert Fabyan, a London clothier, writer, and government official who died in 1513 (Jacob, 2012), Sherman studied the average sentence length of *Fabyan's Chronicle*, and found the average sentence length to be 63.02 words. Sherman examined not only Fabyan's work, but also that of Spenser, Hooker, Macaulay, Channing, and Emerson, and found a gradual and consistent decrease in average sentence length, from Fabyan's 63.02 average to Emerson's 20.58 average. In taking a closer look at the structure of individual sentences within passages, he narrowed his focus to that of the decrease of predication within sentences, calling the decrease an effect, rather than a cause of some force or forces that had not yet been determined (Sherman, 1893).

Examining the prose of writers over time (e.g., Chaucer, Spenser, Emerson, and Bartol), Sherman found that the average predication-per-sentence yield had decreased while the average percentage of simple sentences in passages had increased. To illustrate, Chaucer's writings yielded 5.24 predications per period with four percent of the sentences deemed simple sentences; at the other end of the prose continuum, Bartol's *Radical Problems: Genius, etc.* averaged 2.00 predications per sentence and 45% simple sentences (Sherman, 1893).

Sherman's scientific view of approaching literature, although criticized by some (Dubay, 2007a), became a seminal piece in the study of readability. Sherman's study was the first to look at literature from an objective and scientific perspective (Dubay, 2007a).

Readability Concerns Spurred on by World War I

As the United States entered World War I, much attention was focused on the issue of

literacy among troops. The United States was regarded as a global and an economic powerhouse, yet an estimated 25% of the nation's soldiers were not able to read (Dowling, 2005). In 1917, Robert Yerkes and others began administering tests that were designed to measure intelligence in new recruits. Army Alpha was the test instrument administered to recruits who were literate; Army Beta, a pictorial test, was used to assess intelligence of those who were non-native speakers or of those who were illiterate (Kalantzis & Cope, 2012).

In *The Classic Readability Studies* (2007a), editor William Dubay mentioned two major trends that brought forth new interest in readability: the first trend discussed was the increasing number of students whose parents were immigrants – in today's terms, English Language Learners; the second trend was the focus on measuring education issues scientifically and objectively. The number of children of immigrants with limited English proficiency was significant because of the increasing number of immigrants; this limited proficiency meant that traditional textbooks were too difficult for some students. The timing of using a scientific approach to measure readability was both auspicious and fortuitous in that the influx of English Language Learners required educators to re-evaluate textbooks that had been used in American classrooms.

Continued Research in Readability: Post-World War I Through the Depression

In 1921, psychologist E. L. Thorndike published *The Teacher's Word Book* (TWB). Thorndike had endeavored to rank the 10,000 most common words from 41 sources including classics, textbooks, the Bible, newspapers, and correspondence (Thorndike, 1921). Based on this analysis, Thorndike determined a range and a frequency of a word's appearance in the texts sampled. The range of a word pertained to how widely the word was used across texts; the frequency indicated how often the word was used within the texts sampled. Thorndike's TWB

was frequently referenced in subsequent research. As such, it is important to note that researchers often spoke of the contents of the TWB in thousands: reference was often made to whether words occur in the first 5,000 or in the second 5,000 of the list. Later researchers referred to the frequency of a word based on whether it occurred in the first thousand, the second thousand, and so forth. One interesting finding was that the first 100 words in the Thorndike list made up nearly half of all written material and that the first 300 words of the list accounted for nearly 65% of all written material (Dubay, 2007a). The purpose of TWB was to assist teachers in determining what words were appropriate for students at a particular age as well as to help teachers determine the importance of a word based on its ranking in TWB (Thorndike, 1921). His work was important because it was the first time in the United States that words were listed by frequency (Dubay, 2007a); it is also important because this list laid the groundwork for other researchers to examine readability quantitatively, because they now had a point of reference that demarcated commonly-used words from those that were uncommonly used (and, presumably, more difficult). Thorndike suggested that the word lists in TWB would also be helpful in teaching immigrants to learn English; thus, the TWB might be considered one of the earliest resources to help English Learners (ELs) (Dubay, 2007b).

The TWB of Thorndike did, indeed, serve as an impetus for readability research. Only two years later, Lively and Pressey (1923) developed a method for measuring what they called the "vocabulary burden" of textbooks. The researchers were interested in measuring difficulty of a textbook sampling as well as in finding an appropriate number of words sampled that might yield a reliable indication about reading vocabulary (Lively & Pressey, 1923). The researchers used techniques similar to those employed by Thorndike to determine a frequency, but they also assigned a "zero value" to words: words that appeared in the samplings yet did not appear in the

Thorndike list were listed as zero value (ZV). These words constituted the "technical vocabulary" size. Consequently, Lively & Pressey calculated a weighted median index and determined that the higher the median index, the easier the word. They sampled passages from elementary, junior high, and high school textbooks, English literary classics, college textbooks, and newspapers. While the researchers were not surprised to find ZV words in the upper level and adult reading material, they were rather surprised to find that ZV words appeared even in second-grade textbooks: this suggested that children even as young as eight years old would encounter "difficult" words interspersed throughout relatively simple reading material. The Lively-Pressey formula is believed to be the first measure of readability in American history (Klare, 1963; Huldén, 2004).

Researchers McCall and Crabbs published *Standard Test Lessons in Reading* (STL) in 1925. Although their test lessons are often included in discussions about readability, this writer wishes to convey their work as a reluctant inclusion. This reluctance is not because of the quality or composition of their work, but rather because the STL –initially grade-level, multiple choice reading comprehension assessments given to New York City public school students– were never intended to be used as a standard for readability measurement calibration (Badgett, 2010). Critics included Dale & Chall, who noted "serious deficiencies" in using STL (Dale & Chall, 1948, p. 15). In the latter half of the 20th century, researchers argued that the norms for STL were seriously outdated and in need of re-norming (Jacobson et al., 1978), while other critics stated that the STLs were never sufficiently standardized (Stevens, 1980). The inclusion of STL should serve as a cautionary reminder that work intended for one purpose may eventually be repurposed for other means not originally intended, and that it is wise to understand the history and purpose of work rather than to simply criticize.

Readability continued to be a topic of focus in the late 1920's: in 1927, F. D. Keboch evaluated the difficulty of five seventh-grade American History textbooks. His study determined that the number of words listed in the second 5000 words of Thorndike's TWB related to reading difficulty (Keboch, 1927; Tribe, 1956). In 1928, Mabel Vogel and Carleton Washburne of Winnetka Public Schools published an article called "An Objective Method of Determining Grade Placement of Children's Reading Material" in the *Elementary School Journal*. The justification for the study and development of the method was, according to the authors, the need for an objective method for authors to gauge their writing to students' abilities and interests (Vogel & Washburne, 1928). The researchers turned to their earlier project, the *Winnetka Graded Book List*, in which 700 books were listed based on collective favorable recommendations from a group of 36,750 schoolchildren. From the 700, 152 books were selected for the Winnetka study. A cadre of 20 volunteer teachers examined the 152 books and evaluated each based on vocabulary difficulty, sentence structure, parts of speech, paragraph structure, general structure, and even "physical makeup" – the actual weight of the book, length of line, distance between lines, and font size (Vogel & Washburne, 1928), and ran a series of regression equations to determine correlation. The researchers found that the best multiple correlation was a combination of four of the elements evaluated: within a 1,000 word sampling, elements included the number of different words, the number of prepositions, the number of words, the number of words *not* occurring in Thorndike's list, and the number of simple sentences taken from 75 sample sentences. This combination predicted a high degree of reliability and would provide a reading score that students would need in order to read the text successfully and without great difficulty (Vogel & Washburne, 1928). This regression equation suggested that a prospective author could check his writing against the equation to determine

whether the level was suitable for children at a given age and grade level.

In 1928, William E. Dolch of the University of Illinois published a paper in *The Journal of Educational Research*. Dolch stated that for reading to be of the greatest benefit to schoolchildren, students must be able to understand and read the material easily. He also acknowledged the affective factor (interest, or motivation) in reading comprehension. He stated that reading, a unified process, must involve all of these factors, and that a text that is too difficult may preclude not only understanding, but also interest. In sampling reading passages as had been traditionally performed, Dolch argued that there was room for error in that a sample may not adequately represent the whole of the text, because a text is not homogenous and, therefore, cannot be reduced to the content of its samples. He further argued that earlier methods of measuring vocabulary burden and readability using Thorndike's *Teacher's Word Book* (TWB) were inaccurate because many of the words selected for inclusion in the TWB came from adult reading material and, thus, were not necessarily appropriate for a given grade; thirdly, Dolch stated, the Thorndike method presumed a close correlation between word frequency and difficulty that could not be proven (Dolch, 1928b). Dolch published an earlier article, "Combined Word Studies", which included a list of words that could serve as a basis for determining appropriate vocabulary words for students based on grade levels 1-8 (Dolch, 1928a).

In 1929, Alfred Lewerenz published research findings of difficulty of words based on the initial letter of each word. This paper, published in the *Educational Research Bulletin of Los Angeles Schools,* has been lost in its original form; however, it is referenced in other works including George Johnson's article, "An Objective Method of Determining Reading Difficulty" which appeared in the April 1930 edition of *The Journal of Educational Research*. Lewerenz compared the words appearing in Webster's Elementary School Dictionary and reported that

words beginning with W, H, and B were found in easy material; he also suggested that words starting with E and I were not as commonly found in the student dictionary and that these words were, as such, more difficult. In more complex reading material, Lewerenz found that this balance was reversed (Johnson, 1930). Johnson's own research offered a different measure to determine reading difficulty, stating that his own method was both reliable and easy to administer: "the percent of polysyllabic words is a measure of the difficulty which children will have in reading the book" (Johnson, p. 284, 1930). His study suggested that a random sampling of 3000 words from any given text would yield reliable results. His experiment examined the difference in comprehension between two student groups. Both groups were given similar reading passages of approximately 800 words, but one version of the passage contained 83 words that were monosyllabic while the other version contained polysyllabic synonyms for the monosyllabic counterparts in the first version. Johnson found that students given the "short word" version outperformed those given the "long word" version of the passage, which suggested that monosyllabic words were easier to read and understand. This finding corresponded to the frequency and occurrence of polysyllabic words in the TWB in that the percentage of polysyllabic words increased in the first five of the 500 word groups in the TWB. This finding suggested that the longer the word, the less frequently it appeared in text (Johnson, 1930).

Readability Research during the Depression and Prior to World War II

In 1931, researchers W. W. Patty and W. I. Painter of Indiana University developed a method for measuring the vocabulary burden of textbooks (Dubay, 2007a). Their study criticized Lively & Pressey's approach because they claimed that the Lively & Pressey method failed to account for difficulty of long sentences; likewise, they claimed that Dolch's method of

calculating word difficulty could not be regarded as absolute fact (Patty & Painter, 1931). They proposed a different method for evaluating the vocabulary burden of textbooks by arriving at an *average word-weighted value* (described later). They sampled all state-adopted high school textbooks with the exception of foreign language textbooks, taking the third line of each fifth page of the text; they then arranged each word alphabetically and tabulated the frequency of each word as it appeared in the sampling. The total number of words recorded was called the range of the sample. They compared the words encountered in their sampling to the words in Thorndike's TWB. Their tabulations resulted in a weighted value (WV). This weighted value allowed the word to be considered in proportion to its frequency of use. The average word-weighted value (AWWV) was calculated by dividing the total weighted value (TWV) by the total number of words in the sample (TWS). An interesting finding resulted in evaluating the state-adopted textbooks to determine the grade of highest vocabulary burden: surprisingly, the highest burden was found in tenth-grade books rather than in twelfth-grade books as might have been suspected (Dubay, 2007a; Patty & Painter, 1931).

Although Lyman Bryson was not considered one of the pioneers in readability research, he certainly merits mention. He, along with Charles Beard and M. A. Cartwright, established the Readability Laboratory at Columbia Teachers College in 1934 (Dubay, 2007b; Ohles, 1978). He served as a professor of Adult Education at Columbia until his retirement in 1953 (Ohles, 1978). While Bryson's work did not produce readability measurements, his strong beliefs about the need for clarity and simplicity in writing did significantly influence two of his students, Lorge and Flesch, who would later offer substantial contributions to the field of readability research (Dubay, 2007b).

In 1935, Gray & Leary published *What Makes a Book Readable* in which they examined

what factors impacted readability. They surveyed librarians, publishers, and others involved or interested in adult education, and arrived at four broad characteristics of readability: content, format, general features of organization, and style of expression and presentation. Of these four characteristics, content was deemed to be the most significant factor in readability. They suggested that counting the number of different words in a passage could yield a reliable estimate of the readability of the material, but that other factors such as the number of simple sentences, prepositional phrases, and personal pronouns would provide a more accurate picture of text difficulty. They ultimately arrived at eight factors that indicated difficulty: the number of easy words, the number of different hard words, the percentage of prepositional phrases, the percentage of personal pronouns, average sentence length, the percentage of different words, percentage of monosyllables, and the percentage of simple sentences. Unlike earlier studies, Gray & Leary's work focused on readability factors for adult readers rather than what makes reading easy or difficult for school-age children. Their study also refuted the work of Lewerenz (that words beginning with a certain letter were more difficult than others).

In 1937, researcher Guy Buswell of the University of Chicago began examining readability among civilians. He surveyed 1,000 adults of varying levels of education and determined that reading skills increased according to their level of education. This study was one of the first in the nation to examine reading practices among civilians, as prior practices involved the determination of literacy or illiteracy among solidiers (Dubay, 2007b).

Readability research from World War II until 1950

Dale & Chall (1948) referred to Flesch's 1943 criticism of the use of the Lorge formula in conjunction with the Dale list of words: Flesch argued that the Lorge/Dale method did not account for discrimination between materials that were written above an eighth-grade level.

Flesch's readability formula used three factors to determine readability: the relative number of personal references, sentence length, and relative number of affixed morphemes.

Irving Lorge examined the concept of readability itself and made a relatively bold statement, considering the numerous articles that implied understanding of readability: that the term *readability* was somewhat arbitrary and difficult to define. Lorge's research had built on his 1939 work that used easily applied methods of predicting the difficulty level for student texts. He examined earlier methods that involved counting words within sentences, the number of difficult words encountered, and the number of polysyllabic words encountered. He concluded that the most significant predictor of reading difficulty involved some function of vocabulary (Dale & Chall, 1948).

The publication of *Foundations of Reading Instruction, With Emphasis on Differentiated Guidance* by E. A. Betts (1946) brought about new insights in to the concept of readability and how it was regarded. His work was significant in many ways, but perhaps the most meaningful contribution to looking at reading and vocabulary burden was the determination of different levels at which the reader could function. Betts' work looked at reading from four levels: basal (or independent), instructional, frustration, and probable capacity. This system is still used today and is known as the Betts Criteria. Betts believed that children would perform best when they were intellectually challenged (Stange, 2013). The basal level corresponded to a child's ability to read material without any difficulty or assistance; the instructional level corresponded to the ability of the highest reading level at which place instruction could begin; the frustration level indicated that a child could not navigate the text easily and would likely experience difficulty or frustration from the vocabulary used or the structure of the sentences; and the capacity level corresponded to the ability of a child to comprehend material that was read to him or her (Betts,

1946). The independent level of reading corresponded to fewer than 5% errors in oral reading, instructional level was deemed to have corresponded to 5% errors, and frustration level corresponded to more than 5% errors (Fry, 2002).

Researchers Dale & Chall (1948) addressed the criticism leveled at them by Flesch's 1943 work (and added that Flesch's work was not without its own shortcomings). Their first criticism addressed the use of the affix count. Dale & Chall suggested that counting affixes was subject to human error: that if multiple reviewers attempted affix counts, each reviewer could arrive at a different number; if reviewers consulted the dictionary to obtain an accurate number or morphological affixes, the process of affix counting would then result in a cumbersome and laborious process. Their second criticism concerned Flesch's use of a "personal reference" count as a measure: they argued that personal references did not prove to be a reliable index of difficulty, citing a contemporary magazine article that described how a passage from Koffka's *Principles of Gestalt Psychology* received a score tantamount to that of an elementary psychology textbook despite the consensus that this was *not* an easy work to read. The passage contained numerous personal references that, according to Flesch's scale, should render the text easy to read and to understand; however, argued Dale & Chall, these references, did not refer to specific individuals but rather to abstractions, thus offering no true ease in reading or in comprehension (Dale & Chall, 1948).

The same year, Flesch addressed Dale & Chall's criticisms in *A New Readability Yardstick* (1948). Flesch acknowledged that the counting of affixes was prone to tedium and that references to people within a text might be perceived as arbitrary, but insisted that this was due to the misperception of those who didn't understand the underlying process. He further acknowledged that the time involved in taking the reading samples – six minutes – was too long

for practical use, and that many who applied the Flesch formula of 1946 tended to rely on sentence length (being the easiest component to measure) to the exclusion of the other elements. Flesch reanalyzed his earlier formula and arrived at a dual formula, one that would predict reading ease and a second that would give an indication of human interest in a text. The "reading ease" score ranged from zero to 100, with 0-30 deemed "very difficult", 30-50 "difficult", 50-60 "fairly difficult", 60-70 "standard", 70-80 "fairly easy", 80-90 "easy", and 90-100 "very easy". Human interest scores ranged from "dull" (0-10), "mildly interesting" (10-20), "interesting" (20-40), "highly interesting" (40-60), and "dramatic" (60-100) (Flesch, 1948).

In 1949 Dale & Chall took aim at the very meaning of *readability*. In their article, "The Concept of Readability", they discussed previous readability measures and stated that these formulas did not delve deeply in to what readability truly *was*: some formulas did not account for the conceptual difficulty of text while others used measures that were relative (such as, the number of ideas contained within a passage, or the perceived "human interest" of a text). Another limitation of previous formulas was that they failed to consider the appeal of the subject matter to the particular reader. It is noteworthy to mention that Dale & Chall discussed the importance of examining the attributes of the reader himself: the motivation, adeptness of skill, purpose, maturity, and intelligence all, they argued, impacted the readability of written material. This consideration marked one of the first occasions in readability research in which the focus of readability was shifted from text to reader.

An important work rounding out the first half of the 20th century is that of George Kingsley Zipf who, in 1949, published *Human Behavior and the Principle of Least Effort* (PLE). This principle states that animals will naturally tend toward a path of least resistance (Zipf, 1949). This principle, Zipf stated, applied to many aspects of animal behavior including word

use. He used mathematical models to determine the relationship between difficult words and easy words, and this work marked the beginnings of research of word frequency (Dubay, 2007a). Although this principle was not new to Zipf's research, his name became associated with the principle of least effort, and *Zipf's Curve* (also known as *Zipf's Law*) is still used to describe this principle (Powers, 1998). Klare ascribed characteristics of the PLE to M. V. Bear, a researcher from a quarter century earlier, who quantified the correlation between syllable length and word difficulty. The work of Bear influenced later researchers including G. R. Johnson, who used Bear's findings to develop a grading system for his materials (Klare, 1963).

Readability, War, and Profit: The Connection

At first blush, it might seem unusual to tie readability research to world wars, and it might be equally unlikely to think that readability research had any link at all to economic matters; however, if one examines interest and funding readability studies in comparison with the nation's history between the beginning and mid-20th century, this relationship may not appear to be as far-fetched.

As World War I ensued, the term "functional literacy" was coined to describe the reading skills a serviceman would need to carry out his duties (Scribner, 1984). Prior to this time, a person was either regarded as literate or illiterate (Dubay, 2007b), and there was not much examination given to a continuum of literacy. During the first World War, a fourth-grade education was considered to be sufficient to carry out one's duties and thus considered literate (Scribner, 1984); however, the technological advances with military weaponry made it necessary for servicemen to be more literate than in prior wars. New machinery brought with it an additional component: the ability to read and follow instructions regarding their use. Suddenly, an illiterate soldier was more of a liability than in past wars, and illiteracy was even considered a

threat to our national security (Schantz & Zimmer, 2003). The Army Alpha and Beta tests of 1917, which were initially designed to measure intelligence, revealed a secondary purpose of determining literacy among soldiers (Dubay, 2007b).

Readability research continued to thrive through the 1920s. Adult education efforts increased during the Depression as the federal government allocated substantial funding in this area (Dubay, 2007b). At this same period, the newspaper industry suffered devastating financial losses partly due to the advent of radio (Kirchloff, 2011). Between 1929 and 1941 newspaper advertising revenue dropped 45%, (Kirchloff, 2011). An interesting figure that would suggest that readers were not buying newspapers as in years prior to the Depression is one mentioned by Gray & Leary (1935): that major U. S. libraries experienced phenomenal increases in the number of items borrowed: the public was increasingly frugal, and borrowing reading materials from the library rather than purchasing their own was evidence of this. In fact, one of the libraries mentioned reported a 173% increase in items borrowed from 1928 to 1929 (Gray & Leary, 1935). The need to increase revenue (and to increase readership) meant that newspaper publishers reexamine the readability of news items in print (Dubay, 2007b).

World War II again marked a time when concern for literacy was paramount to our nation's interests. On the home front, citizens needed to band together to assist with the war effort, both directly and indirectly, and newspapers were the medium for reaching audiences far and wide: not only to provide news and information, but also to appeal to the public to financially support the war through the purchases of war bonds (Dale & Chall, 1948). In 1947, Melvin Lostutter published findings from research in which he applied both the Flesch and Lorge readability formulas to 150 newspapers and determined that most articles were written at a level that was approximately five years beyond the education level of most readers (Johns & Wheat,

1984; Dubay, 2007b). The lessons learned from the depression-era economic plunge served as a reminder to newspaper publishers that they could not afford to write and attempt to reach a market that either wasn't willing or wasn't capable of reading their print (Dale & Chall, 1948).

The issue of illiteracy among soldiers was a grave concern as well. N. G. McCluskey described the problem of illiteracy as something that was even more devastating than anything that the enemy could launch against the United States, adding that illiteracy precluded many otherwise able-bodied men from contributing to the effectiveness of the war effort (Schantz & Zimmer, 2003). Increasing demand for technological and communication skills among servicemen made the need for literate individuals even more imperative as weaponry and the intricacy of maneuvers became more complex (Fattu & Standlee, 1954).

Conclusion

Klare (1963) estimated that researchers had introduced nearly 30 formulas between the beginning of the 20th century and 1963. In the 1930s, reading as a separate field of study was receiving pedagogical validation (Barry, 2008), and along with this and the Readability Lab at Columbia Teachers College was a pressing need to examine literacy at the adult levels as well. War, the Depression, the decrease in newspaper readership and the increase in library borrowings made the need to closely evaluate what made text readable. These difficult days in our nation's history made two things clear: it was imperative that the consideration of the reader and his or her needs was vital for national interest and that these efforts be continued as the nation prepared to enter in to the Cold War so that our citizens might survive the dark days that would lie further ahead.

References

Adams, M. J. (2011). Advancing our students' language and literacy: The challenge of complex texts. *American Educator*, *34*(4), 3-11.

Badgett, B. A. (2010). Toward the development of a model to estimate the readability of credentialing-examination materials. *ProQuest LLC*.

Barry, A. (2008). Reading the past: Historical antecedents to contemporary reading methods and materials. *Reading Horizons*, *49*, 31-52.

Bejar, I. I., & Blew, E. O. (1981). Grade inflation and the validity of the scholastic aptitude test. *American Educational Research Journal*, *18*, 143-156. doi: 10.3102/00028312018002143

Betts, E. A. (1946). Foundations of reading instruction, with emphasis on differentiated guidance.

Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin*, 11-28.

Dale, E., & Chall, J. S. (1949). The concept of readability. *Elementary English*, *26*, 19-26.

Dolch, E. W. (1928a). Combined word studies. *The Journal of Educational Research*, 17, 11-19.

Dolch, E. W. (1928b). Vocabulary burden. *The Journal of Educational Research*, 17, 170-183.

Dowling, T. C. (2005). *Personal perspectives*. Santa Barbara, Calif: ABC-CLIO.

DuBay, W. H. (2007a). *Unlocking language: The classic readability studies*. Costa Mesa, Calif: Impact Information.

DuBay, W. H. (2007b). *Smart language: Readers, readability, and the grading of text*. Costa

Mesa, Calif: Impact Information.

Fattu, N. A., & Standlee, L. S. (1954). *Analysis of reading difficulty of selected Navy materials*. Indiana University at Bloomington Institute of Educational Research.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221.

Fry, E. (2002). Readability versus leveling. *The Reading Teacher,* 56, 286 - 291.

Gray, W. S., & Leary, B. E. (1935). What makes a book readable.

Huldén, M. (2004). *Linguistic complexity in two major American newspapers and the associated press nNewswire, 1900–2000* (Unpublished Master's Thesis). Åbo Akademi University.

Jacob, M. (2012). Encyclopedia of Tudor England. Booklist, (14). 29.

Jacobson, M. D., Kirkland, C. E., & Selden, R. W. (1978). An examination of the McCall-Crabbs standard test lessons in reading. *Journal of Reading*, 224-230.

Johns, J. L., & Wheat, T. E. (1984). Newspaper readability: Two crucial factors. *Journal of Reading*, 432-434.

Johnson, G. R. (1930). An objective method of determining reading difficulty. *The Journal of Educational Research*, 283-287.

Kalantzis, M., & Cope, B. (2012). *New learning: Elements of a science of education*. Cambridge: Cambridge University Press. doi:10.1017/cbo9781139248532.003

Keboch, F. D. (1927) Variability of word-difficulty in five American history texts. *The Journal of Educational Research,*15, 22-26.

Kirchhoff, S. M. (2010). *US newspaper industry in transition*. DIANE Publishing. doi: 10.1201/b13161-13.

Klare, G. R. (1963). *The measurement of readability*. Ames: Iowa State University Press.

Lively, B.A., & Pressey, S.L. (1923). A method of measuring vocabulary burden of textbooks. *Educational Administration and Supervision*, *9*, 389–398.

Lorge, I. (1939). Predicting reading difficulty of selections for children. *The Elementary English Review*, 229-233.

Lorge, I. (1949). Readability formulae-an evaluation. *Elementary English*, *26*(2), 86-95.

Ohles, J. F. (Ed.). (1978). *Biographical dictionary of American educators*. Greenwood Publishing Group.

Patty, W. W., & Painter, W. I. (1931). A technique for measuring the vocabulary burden of textbooks. *The Journal of Educational Research,* 127-134.

Powers, D. M. (1998). Applications and explanations of Zipf's law. *Proceedings of the joint conferences on new methods in language processing and computational natural language learning of the Association for Computational Linguistics.* 151-160.

Scribner, S. (1984). Literacy in three metaphors. *American Journal of Education*, 6-21.

Sherman, L. A. (1893). *Analytics of literature: A manual for the objective study of English prose and poetry.* Ginn.

Stange, T. V. (2013). Exploring text level difficulty and matching texts for reading achievement. *Education Matters: The Journal of Teaching and Learning, 1*(2), 111-128.

Stevens, K. C. (1980). Readability formulae and McCall-Crabbs standard test lessons in reading. *The Reading Teacher*, 413-415.

Tribe, E. B. (1957). *A readability formula for the elementary school based upon the Rinsland vocabulary* (Doctoral dissertation, The University of Oklahoma.).

Thorndike, E. L. (1921). *The teacher's word book* (Vol. 134). New York: Teachers College,

Columbia University.

Vogel, M., & Washburne, C. (1928). An objective method of determining grade placement of children's reading material. *The Elementary School Journal*, 373-381.

Zimmer, J. (2003). A Brief History of Meliorism (1900-1960). History of Reading News. *24*(1). Retrieved from http://www.historyliteracy.org/newsletters/histlit.2003.27.1.pdf

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, Mass: Addison-Wesley Press.