Reading texts in Croatian Elementary School EFL Textbooks: A Readability Analysis

Nedić, Tijana

Sveučilište J.J. Strossmayera u Osijeku
Filozofski fakultet Osijek
Diplomski studij engleskog jezika i književnosti - nastavnički smjer
i njemačkog jezika i književnosti - nastavnički smjer

Tijana Nedić

Analiza težine tekstova za čitanje u hrvatskim udžbenicima za engleski jezik za osnovnu školu

Diplomski rad

Mentor: prof. dr. sc. Višnja Pavičić Takač

Osijek, 2016

Sveučilište J.J. Strossmayera u Osijeku
Filozofski fakultet Osijek
Odsjek za engleski jezik i književnost
Diplomski studij engleskog jezika i književnosti - nastavnički smjer
i njemačkog jezika i književnosti - nastavnički smjer

Tijana Nedić

Analiza težine tekstova za čitanje u hrvatskim udžbenicima za engleski jezik za osnovnu školu

Diplomski rad

Humanističke znanosti, filologija, anglistika

Mentor: prof. dr. sc. Višnja Pavičić Takač

Osijek, 2016

University of J.J. Strossmayer in Osijek
Faculty of Humanities and social Sciences
Teaching English as a Foreign Language, MA Programme
and Teaching German as a Foreign Language, MA Programme

Tijana Nedić

Reading Texts in Croatian Elementary School EFL Textbooks: A Readability Analysis

Diploma paper

Mentor: Višnja Pavičić Takač, PhD

Osijek, 2016

University of J.J. Strossmayer in Osijek
Faculty of Humanities and social Sciences
Department of English language and literature
Teaching English as a Foreign Language, MA Programme
and Teaching German as a Foreign Language, MA Programme

Tijana Nedić

Reading Texts in Croatian Elementary School EFL Textbooks: A Readability Analysis

Diploma paper

Humanities, philology, English studies

Mentor: Višnja Pavičić Takač, PhD

Osijek, 2016

Abstract

The topic of this research is readability formulas and their application in the analysis of elementary school EFL textbooks. In the first part of the paper, a short history of readability is presented, followed by an overview of the application of those formulas in the EFL classroom. The present study was conducted on three different textbook series for elementary school, for which readability levels were calculated using the most popular tests. In addition to the differences between grades, the research investigated the differences within the series, as well as the correlation of years of learning with test scores and text variables, such as sentences and words.

Key words: readability, readability formulas, EFL textbooks

Sažetak

Ovo istraživanje bavi se formulama za procjenu težine teksta u osnovnoškolskim udžbenicima za engleski jezik u Hrvatskoj. U prvom se dijelu ukratko predstavlja povijest formula za procjenu težine teksta, nakon čega se predstavlja primjena tih formula u nastavi engleskog kao stranog jezika. Samo istraživanje provedeno je na tri različite serije osnovnoškolskih udžbenika, za koje se izračunala prosječna težina teksta koristeći se najpopularnijim formulama. Osim utvrđivanja razlika između razreda, proučavane su i varijacije unutar serija, kao i korelacije godina učenja s rezultatima različitih testova te tekstualnim varijablama poput broja riječi i rečenica.

Ključne riječi: formule za procjenu težine teksta, udžbenici engleskog za osnovnu školu

Contents

1. Introduction

This diploma paper consists of two parts. The first part brings an overview of the theoretical background of readability and readability formulas, as well as a review of relevant studies in which readability formulas were used.

The second part of the paper will focus on the present study, where three ELF textbook series were analyzed by using the main readability formulas to give a clear overview of readability scores in elementary schools. The emphasis will be on the correlations and changes within one publisher/series throughout the grades, as well as the comparison of grades across different series. The first hypothesis is that the distinction between the two locally published series is smaller when compared to the foreign textbook (*Project*). The second hypothesis is that the readability scores will get higher, which will indicate that reading texts do get harder with every following grade. The last hypothesis is that there is a significant correlation between the readability tests scores and the characteristics of the text, such as the number of words in total and per sentence.

2. Theoretical Background

2.1. The Term Readability

From the point of psycholinguistics, reading is defined "as a multi-component skill operating a number of different levels of processing: lexical, syntactic, semantic, and discoursal" (Just and Carpenter, cited in Crossley et al., 2008:477). The field of psycholinguistics explains how readers make connections between text and word representation, recalling their previous reading experience, where readers transfer their reading strategies from their native first language (L1) reading experience to second language (L2) reading. Alderson and Banerjee (2001:79) explain reading as "an interaction between a reader with all that the reader brings with him/her: background knowledge, affect, reading purpose, intelligence, first language abilities and more – and the text – whose characteristics include topic, genre, structure, language (organization, syntax, vocabulary, cohesion)". The readability definition by Prins and Ulijn (cited in Kasule, 2011:63) also includes the writer in the text-reader relationship. They defined readability as "the ability of the text to communicate the intention of the writer to the intended reader".

Klare (cited in DuBay, 2004:3) defined readability as "the ease of understanding or comprehension to the style of writing, focusing on the separation of the writing style from other issues such as content, coherence, and organization". McLaughlin (cited inBrangan, 2014:37), the author of the SMOG readability formula, emphasizes the importance of the reader-readability relationship, focusing on the connection between the text itself as well as the characteristics of the reader such as the reader's reading skill, prior knowledge, and motivation. DuBay (2007) stressed that most people confuse readability with legibility, which is the visual perception of the layout (background, font style and size, spacing, etc.).

Dale and Chall, the authors of one of the most popular readability formulas, classify readability in its "broader sense" as “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 readers has with it. The success is the extent to which they understand it, read it at an optimal speed, and find it interesting” (Dale and Chall, 1949:5).Vogel and Washburne (cited in Lorge, 1949:9) interpreted readability as a text index that displays the average amount of reading ability a person needs to understands the text.Zamanian and Heydari (2012) stated that readability studies focus on measuring comprehension of a piece of writing, concentrating on linguistic factors.

We can define readability as the needed reading skill level (or education level if we are talking about English as a foreign language (EFL) reading) to understand a text. Each text has its readability level, which can be calculated by using different readability formulas, which shall be explained in the next chapter. Reading comprehension is always tested by comprehension tests; from multiple choice exercises to cloze tests (gap filling). At the same time, there is criticism directed toward that method of reading comprehension testing, because it is still unclear whether the answers depend on just the understanding of the read text (meaning, the gathered information), or a good/bad memory and prior knowledge (DuBay, 2007). Readability studies are usually conducted from the aspect of a native English user and their reading abilities. Alderson (cited in Alderson and Banerjee, 2001) questions if the readability problems in the second language should be observed as L2 problems or readability problems.

DuBay (2007) explains that the reading level does not have to match the assigned grade or education level of a person. He reports that the average reading skill level in the USA is around the 8th grade. Furthermore, he claims that within one class a teacher can find a variation in the reading skill over five grades between the students. He adds that the general audience, especially in the field of healthcare, has trouble understanding the complex and too advanced reading level of some texts, leading to all kinds of misunderstandings.

2.2. Readability Formulas

2.2.1. Definitions, Application and Common Variables

According to DuBay (2004), readability has been an inspiration and a puzzle to educators, who struggled to discover new ways of predicting text difficulty. Throughout the years, they invented hundreds of readability formulas, which have been used word wide. By the 1980s, there were 200 formulas, with strong theoretical and statistical validity to back them up. According to both DuBay (2007) and Brangan (2014), there are readability formulas for other languages, such as French, German, Swedish and more. Carell (1987) explained that it was very easy creating readability formulas for European languages, since they have a similar syntax to English, whereas the formulas for non-European languages required more adjustment and more variables.

Kirkwood and Wolfe (cited in Zamanian and Heydari, 2012) defined readability formulas as an analytical way of predicting readability. They explained that readability formulas are much more objective, but a good readability formula will correlate well with the results of expert judges. Crossley et al. (2011:87) define the traditional readability formulas as "simple algorithms that measure text readability based on sentence length and word length". Kondru (cited in Zamanian and Heydari, 2012:43) defined readability formulas as "an equation that gives an estimate of a readability of a text... in terms of number of years of education one needs to comprehend that text". Fundamentally, we could state that readability formulas are equations which are used to predict the comprehensibility of written material by counting and measuring structural elements of a text (Dale and Chall, 1949). Brangan (2014) mentioned that the most popular formulas were written during the golden age of the readability formulas – the 1950s when famous writers and scholars such as Rudolf Flesch, George Klare, Edgar Dale, and Jeanne Chall developed the first versions of their formulas.

Readability research was widely conducted in the United Kingdom and the United States (Johnson, 1998). Most of the formulas which contain grade levels as the final measuring item correspond with the US grade level. Johnson adds that one of the most problematic aspects of assessing readability on a subjective level is that teachers usually underestimate the difficulty of the text. Zamanian and Heydari (2012) explained that the biggest advantage of readability formulas is that a person gets a numeric score that can be interpreted from an index, which

means that it is very universal, where in the past a person could just focus on the text comments from judges, where there could have been problems in the interpretation.

Over the years, the field of applications of the readability formulas and readability research flourished with the increase of the readability formulas. Some of the readability studies through the years were conducted in fields such as political literature, all kinds of manuals, different consent forms, healthcare information, court and legislation documents (where even President Clinton directed that all federal agencies had to issue documents and regulations in plain language), journalism, and textbook publishing (DuBay, 2004:55).

Readability is a popular field of study even today. Scholars are testing different corpora, trying to alter and adapt the most popular formulas, translating and adjusting the formulas in other languages. One of the latter mentioned studies took place in Croatia, where Sanja Brangan, from the Medical Department of the University of Zagreb, tested the readability in the healthcare communication with the altered Flesch Reading Ease formula. Her other study was the quantitative text difficulty level research where she compared English and Croatian paragraphs from 4 different fields – literature, SETimes, pop-science articles and scientific papers from the medical journal JAMA (Brangan, 2014:48).

Greenfield (2004) explained how the validity of readability formulas is tested. In a native English context, the readability formulas' validity is the accuracy of prediction (meaning, if the formulas managed to match the assigned text to a reader who will have no problems in understanding the text). However, in an EFL context, it is matching the students' performance on the follow-up reading comprehension tests (cloze test, multiple question tests) with the predicted results. Dale and Chall (1949) emphasize that the success of the reader (the sheer understanding of the text) can also be influenced by the reader's reading skills, intelligence, experience, maturity, interest and purpose in reading. That is why Heydari (2012) expressed the difference and importance of the reader variables and text variables. The variables that are present in all readability formulas come from the field of text based factors. As in Figure 1 bellow, Gray and Leary (cited in DuBay, 2006:40) "identify four factors affecting readability, namely: content (including prepositions, organization, coherence), style (including semantic and syntactic elements), design (including typography, format, and illustrations), and structure (including chapters, headings, and navigation)". The two main variables come from the two most influential fields; the first one is semantic content (vocabulary) and the other one is the syntactic structure (visible in sentence length) (DuBay, 2004).

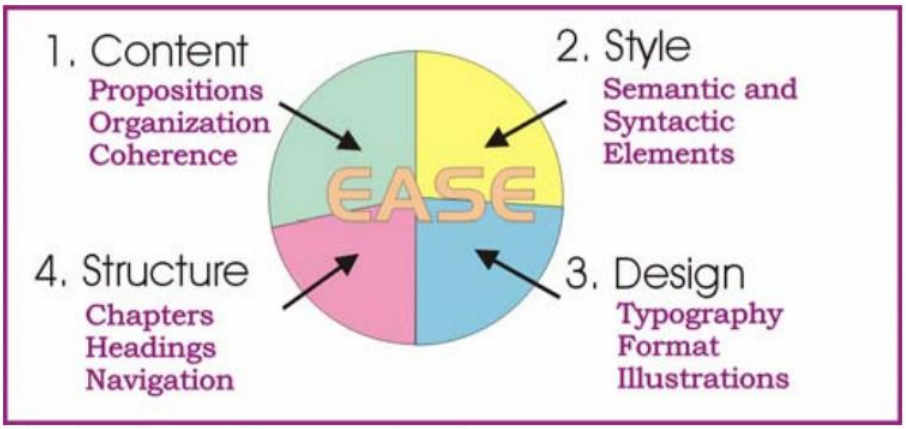

Figure 1: Four elements of reading ease by Grey and Leary (cited in DuBay, 2006:40)

Lorge (1949:12) listed his version of the four elements of readability formulas, which are: vocabulary load, sentence structure, idea density, and human interest. Brown (1998) and many more scholars (such as Greenfield and Crossley) argue and emphasize the need for new variables to be taken into account for more modern readability formulas. They want to introduce variables focusing on the reader, such as readers' age, education and cultural background (highlighting the difference between English and their native language). During his EFL study in Japan, he introduced his own variables for the new and improved readability formula, which he designed for EFL use. Those variables are: syllables per sentence, frequency of a word in a passage, percentage of long words (more than seven letters), and percentage of function words (Brown, 1998). It is noticeable that those variables require a more thorough examination of the text and sentence structure when compared to one of the first and original variables used for readability formula calculation.

However, while measuring readability, one must also think about the reader and his features. Bensoussan (cited in Kasule, 2011:64) identified seven factors of the reader that are not measurable with readability formulas but can have a great influence on the reading outcome. Those factors are: faulty top-down processing, faulty bottom-up processing, linguistic proficiency, lack of motivation, over-motivation, familiarity with the topic, and misleading or unfocused questions. Kasule (2011) explains that, because it is hard to measure those factors, today it is common to use readability formulas to get the assessment of the linguistic and semantic hardness of the text, and then do a comprehension test, such as the cloze test, to double check the reading outcome and readability, as well as appropriateness of the text.

Even though readability formulas have been criticized throughout the years of their usage (that shall be found in the chapter 2.2.4.), they still have a wide range of usage because they measure reading material without being limited by subjective features of personal judging or text leveling (Zamanian and Heydari, 2012). McClare (cited in Zamanian and Heydari, 2012) warned that writers, teachers, editors, and all the people dealing with text creation, should use readability formulas just as an evaluation tool, but not as a writing tool.

2.2.2. Early Studies

"The first aim of the classic readability studies was to develop practical methods to match reading materials with the abilities of students and adults" (DuBay, 2004:13). During the 1930s, it was necessary to find a way to provide appropriate reading texts for a target audience. Throughout the whole century publishers, librarians, teachers, scholars and investigators searched for a method to determine the reading level of a text (DuBay, 2004).

One of the first investigators in the field of readability was Sherman in the 1880s. His scientific breakthrough was marked by investigating literature through statistics. In his study, he examined older books and made a discovery how sentences become shorter over time. His discovery is significant because he was the first one to use statistical analysis as a method of objective literary criticism, which became the milestone of all the future readability research.

During the 1920s, a new major trend in the field of readability emerged: vocabulary frequency lists. They were considered as an objective measuring tool, which was very useful to the teachers, especially those who were facing adaptation of reading material for the first or second generation of immigrant students. Thorndike published the *Teacher's Word Book* in 1921, where he listed 10,000 words by frequency of use. Thorndike (cited in DuBay, 2004) claimed that the more frequently a word is used, the students will get more familiar with it, which will lead to their more frequent usage of that word. He also concluded that as we grow, our vocabulary grows, which facilitates the possibility of using more complex grammar structures later on. During that era, the vocabulary of a person was often used as the sole measurement of their knowledge. In the upcoming years, a lot of readability studies were alluding to the importance of this frequency factor when reading and teaching words in general. During the 60s, Klare (cited in DuBay, 2004) supported his theory of faster acquisition of more frequent words. Even in the

modern readability studies at the turn of the century, scholars still allude to the factor of word frequency and its importance in language teaching.

The trend of using statistics in the calculation of text difficulty continued to develop. Zipf investigated the mathematical relationship between hard and easy words, composing the so-called Zipf's curve (cited in Brangan, 2011). Dale and O'Rourke continued the work on word frequency lists, creating *The Living Word Vocabulary: A National Vocabulary Inventory* in 1981, thus creating a new approach, in which they do not only match the text and the readers' reading abilities, but also take into consideration the experience and background of the reader (cited in DuBay, 2004).

The first readability formula was developed by Lively and Pressey in 1923. Their aim was to "reduce the 'vocabulary burden' of textbooks" (DuBay,2006:6) since their field of study were the selection and adaptation of science textbooks for junior high school. The problem of those textbooks was that they were loaded with hard and complex technical words, which led to inadequate teaching because the teachers had to spend most of their lessons explaining the terms, instead of teaching the subject (Lively and Pressey, 1923, cited in DuBay, 2006). They were using several methods to determine the readability of the textbooks, like counting the general vocabulary load (the words in the text sample), the zero value words (technical terms and words that are not on Thorndike's Vocabulary List), as well as finding out the index number of the words which were on Thorndike's list (Lively and Pressey, 1923, in DuBay, 2006:9). However, the authors noted one shortcoming of their study: they only used a thousand word sample. It was questionable whether the sample represented the whole book in a just way, or if another 1000 words would have made a difference. Nevertheless, they gave an interesting method of readability calculation.

Vogel and Washburne created the **Winnetka formula** in 1928. They expanded their study to the structural characteristics of the text, such as sentence types and prepositional phrases (DuBay, 2006). Vogel and Washburne created the *Winnetka Graded Book List*, where they listed children's books, according to the reading ability grade (Vogel and Washburne, 1928, cited in DuBay, 2006:18). For their new study, they took 152 books from the Winnetka list, and with the help of 20 teachers, they established categories, which they considered to be the cause of the book placement in a selected grade. They got four categories which they named the key readability factors. Those factors are: vocabulary difficulty (using Pressey's technique), sentence structure, parts of speech, paragraph construction, and general text construction (Vogel and

Washburne, 1928, cited in DuBay, 2006:19-20). Dale and Chall (1949) also stressed the importance of selecting appropriate books for children. That kind of careful selection is necessary because the right book helps develop the reading skills. The Winnetka formula, or, as the authors said, the formula to determine any book grade placement, is a 7 step guideline, with detailed descriptions how to calculate all the needed components (labeled X2 to X5). When all the components are calculated and have a numeric value, the formula below gives a value that can be compared to the Stanford Achievement Test (Figure 2) to see the grade value of that book.

$X1 = .085X2 + .101X3 + .604X4 - .411X5 + 17.43$[1]

GRADE STANDARDS—PARAGRAPH-MEANING SECTION OF THE STANFORD ACHIEVEMENT TEST

Score	Grade
4–16	II
18–34	III
36–52	IV
54–62	V
64–70	VI
72–78	VII
80–86	VIII
88–94	IX
96–102	X
104–112	XI

Figure 2: Stanford Achievement Test grade standards (cited in DuBay, 2006:24)

Patty and Painter (1931, cited in DuBay, 2006) tried to create a new measurement tool for textbooks. They criticized Lively and Pressey, claiming their 1 000 word samples are not an appropriate way to generalize the readability of the whole book, so they came up with the idea to take one line from every 5th page in the textbook. In 1935, Grey and Leary published their *What Makes a Book Readable*, where they investigated the correlation of different readability factors, henceforth coming up to the conclusion of four basic elements of reading ease (Figure 1). From the initial 228 elements that affect readability, they gained 64 variables that had correlations higher than .35 (Grey and Leary, 1935, cited in DuBay, 2006:41), leading to their creation of the

[1]The Winnetka formula by Vogel and Washburne, 1928, cited in DuBay, 2006:24.

two most important readability variable categories: semantic vocabulary difficulty and syntactic sentence length (Grey and Leary, 1935, cited in DuBay, 2006:42).

The era between the two world wars is called the era of the classic readability formulas. One of the pioneers in that field, following the footsteps of Sherman and Zipf, was Kitson. His field of research was the readability of newspapers and magazines, focusing on the readability levels of the printed materials with the target audience of the publication. Kitson's conclusion (cited in DuBay, 2004:13) was that sentence length and word length in syllables were an important factor in measuring readability. His two variables became the main foundation of all the future readability formulas.

The most famous and abiding readability formula created in that time is the **Flesch Reading Ease**. The first version of the formula was published in the dissertation *Marks of a Readable Style*, by the Austrian refugee Rudolf Flesh in 1943. In 1948 he created the Reading Ease formula, using the number of syllables and the number of sentences for each 100-word sample.

Reading Ease score = $206.835 - (1.015 \times \text{ASL}) - (84.6 \times \text{ASW})$

Where:
ASL = average sentence length (number of words divided by number of sentences)
ASW = average word length in syllables (number of syllables divided by number of words)[2]

The reading ease formula has a scale from 1 to 100, where "100 indicates reading matter understood by readers who have completed the fourth grade and are ... functionally literate" (DuBay, 2004:20). The rest of the scale is visible in Table 1 below:

Table 1: FRE index and interpretation (cited in DuBay, 2006:97)

FRE Score	Explanation	Estimated grade
90-100	Very easy	4th grade
80-90	Easy	5th grade
70-80	Fairly easy	6th grade
60-70	Standard	7th or 8th grade
50-60	Fairly difficult	High school
30-40	Difficult	High school/college
0-30	Very difficult	College graduate

Flesch's Reading Ease formula is the most widely used formula, as well as one of the most tested and most reliable formulas, even today. According to DuBay (2004), in 1976, as a request from

[2]This and all the following formulas were taken from the page: http://www.readabilityformulas.com/

the military, the Reading Ease formula was modified so the results could be presented as a grade-level index. The new formula became known as the **Flesh Grade-Scale formula** or the **Flesch-Kincaid Grade formula** (or Grade Level (FKGL), which is used in the abbreviation in this study).

FKRA = (0.39 × ASL) + (11.8 × ASW) - 15.59
Where,
ASL = Average Sentence Length (i.e., the number of words divided by the number of sentences)
ASW = Average number of Syllable per Word (i.e., the number of syllables divided by the number of words)

This formula was predominantly used in journalism (by Flesch himself who conducted several research studies about readability in newspapers) and by librarians, teachers, and publishers, who could easily match books and other reading materials with the needs and abilities of their students (Brangan, 2011). Today it is used in most of the computer programs dealing with readability, which helps its long-lasting popularity.

The other longstanding formula from that era is the **Dale-Chall original formula**. Edgar Dale, a university professor, specialized in the field of communications, wanted to "improve the readability of books, pamphlets and newsletters – the stuff of everyday reading" (DuBay, 2004:22). The first version of the formula came out in 1948 and was used for adults and children above the 4th grade. The formula goes:

Raw score = 0.1579×(PDW) + 0.0496×(ASL) + 3.6365
Where:

Raw score = uncorrected reading grade of a student who can answer one-half of the test questions on a passage.
PDW = Percentage of difficult words not on the Dale-Chall word list.
ASL = Average sentence length

The distinctiveness of the formula is that, besides the common sentence length variable, it uses a percentage of "hard words". Those "hard words" are words which are not on the Dale-Chall list of 3,000 words, which are known to fourth-grade readers (DuBay, 2004). The score is numerical, with a table that matches the score with the corresponding grade, as visible in Table 2.

Table 2: Dale- Chall readability scores and grade level[3]

Adjusted score	**Grade level**
4.9 and below	Grade 4 and below
5.0-5.9	Grades 5-6
6.0-6.9	Grades 7-8
7.0-7.9	Grades 9-10
8.0-8.9	Grades 11-12
9.0-9.9	Grades 13-15
10 and above	Grades 16 and above (college graduates)

Dale and Chall published their first draft of the formula in 1948. In the same year they published an article, *A Formula for Predicting Readability: Instructions*, where they wrote a step by step guide how to calculate the readability of a text, from selecting the right text sample, to labeling and counting words, up to the worksheet samples, followed by the interpretation table and their own word list (Dale and Chall, 1948, cited in DuBay, 2006:75-94).

The last notable formula of the classic era is the **Fog Index**. Robert Gunning was exploring the field of textbook publishing, when he came to the conclusion that high school graduates have problems reading, due to the "fog" within texts, as well as the problematical writing which becomes problematic reading. In *The Technique of Clear Writing*, he published his Fog Index. The formula was special because he used the average sentence length (a common variable) and the number of words with more than two syllables for each 100 words (DuBay, 2004).

Grade level= 0.4 × ((average sentence length) + (percentage of Hard Words))
Where: Hard Words = words with more than two syllables

[3]Taken from Dale, E. and J. S. Chall. 1948, *A formula for predicting readability*

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

Figure 3: Gunning Fog Index Interpretation (cited in Zamanian and Heydari,2012:45)

The most significant discoveries of the first readability studies were the variables, as well as the correlations with the grades, years of education, and frequency of words. The authors set a solid foundation for the further readability research that occurred in the other half of the 20th century, stimulating new studies with the main goal to improve the formulas, as well to test and discover other factors affecting readability. However, by the end of this era, the first waves of criticism regarding the overuse of readability formulas occurred. But, nothing could diminish the main achievement of readability research: making the community aware of the problem of inadequate readability.

2.2.3. Recent Studies

Since the readability formulas were often criticized because of their lack of inner text understanding, the new era of readability formulas was marked with the need of bigger inner structure perspective and comprehension testing. In the 1980s, the New Literacy Studies appeared. The focus was on literacy, and whether it is a skill independent of one's interests, social, and economic environment (like it was thought before), or if there are multiple kinds of literacy a person can possess and acquire, as well as including the concept of identity of the reader in his own literacy. DuBay (2007) exemplifies that claim, by listing the new ideas, such as developing new criteria to evaluate the passages, continuing the introduction of new formula variables, as well as, the introduction of the first computerized formulas. The most significant novelty of this era was the combining the revised readability formulas with the cloze test, to give

the researchers a better understanding of the readers' understanding of the text. The formulas would calculate the lexical and syntactical difficulty level of the text, and in the next step, a group of the right grade level would solve a cloze test to double check the appropriateness of a text.

Some of the most significant and commonly used methods of this era are the cloze test and new readability formulas. It is also important to mention Klare and his numerous studies in the 1960s, where he was working in the community of scholars, whose goal was to understand more how the readability formulas and variables work and to perfect them for future usage. Some scholars were also investigating the influence of prior knowledge and retention, as well as the mutual influence of readability and the readership (DuBay, 2007).

Another important method was text leveling, which is the name for the "subjective analysis of reading level that examines vocabulary, format, content, length, illustrations, repetition of words, and curriculum" (DuBay, 2007:38). Even though text leveling was used in the era prior to readability formulas, some scholars felt the need to revive it, in a movement called Reading Recovery System. They claimed that text leveling is more sensitive to the needs of the readers than readability formulas. Fry (2002:286) said that the main difference between readability formulas and text leveling is that the formulas give a numeric score to rank books according to difficulty, while leveling is a subjective system of determining the difficulty of texts or books, mostly used for beginning reading levels. He claims that readability formulas should be used for higher grades because they have a wider range of scores, while the leveling technique is most suitable for kindergarten and beginning elementary school grades. This way of subjective grade assigning was used also before the invention of readability formulas, and now it is used together with them, as a kind of post-testing.

Taylor (cited in DuBay, 2004) invented the cloze test, a type of test where every 5th word would be erased, and the reader had to fill in the gap with a word that suits the gap. The advantage of this testing method was that it did not just measure words, but their connection and relationships with other words within a sentence. It became very popular and mostly replaced multiple choice questions in the post readability formulas reviews. When calculating the score of a cloze test, only the correct form (the right person, number, tense, voice) of the word is the right answer, meaning no synonyms can be acknowledged. A low score usually means a hard text, especially for the level of the tested reader. That is the main reason cloze tests are mostly for intermediate and advanced readers. However, some scholars, like Carver (cited in Brown, 1998:9), argued

that the "cloze difficulty estimate depends both on the ability level of the particular group which was administered the cloze test, as well as the difficulty level of the material".

Bormuth (cited in DuBay, 2004:43) focused his studies on the changes of readability scores when the readability variables were changed, and how much all of that affects the later comprehension. He came to the conclusion that "cloze testing made it possible to measure the effects of those variables not just on the difficulty of the whole passages but also on individual words, phrases, and clauses". He also established testing batteries (like the ones Greenfield used in his study). In one of his researches he used 20 different passages and came to the conclusion that different readers (meaning, readers with different readability levels) have a different word choice from those who do not have such a developed reading skill and vocabulary. All his research led to the creation of 24 different readability formulas with different variables.

During this era, the readability formulas focusing on primary grades were developed. Such formulas are: the Sprache formula (1953), the Harris-Jacobson primary readability formula (1973), as well as the Powers-Sumner-Kearl formula (http://www.readabilityformulas.com/free-readability-formula-tests.php). They will not be used in this study because they are designed to fit the needs for just the first three to four grades of primary school, which makes them unsuitable for the higher elementary school grades. The latter mentioned formula is often combined with the new Dale-Chall formula for higher grades.

Edward Fry made an important contribution to the field of readability research with his **Fry Readability Graph**. According to DuBay (2004), Fry's original work was focused on the readability in high school, but in 1969 he extended and adapted the graph to answer the needs of readability measuring for both primary school and college. The formula works on a sample of 100 words. The score is a graph, where Y –the vertical line, denotes the number of sentences per a 100 word passage, and the line horizontal line X indicates the average number of syllables per a 100 word sample. Where those two lines meet, we read the approximate readability score.

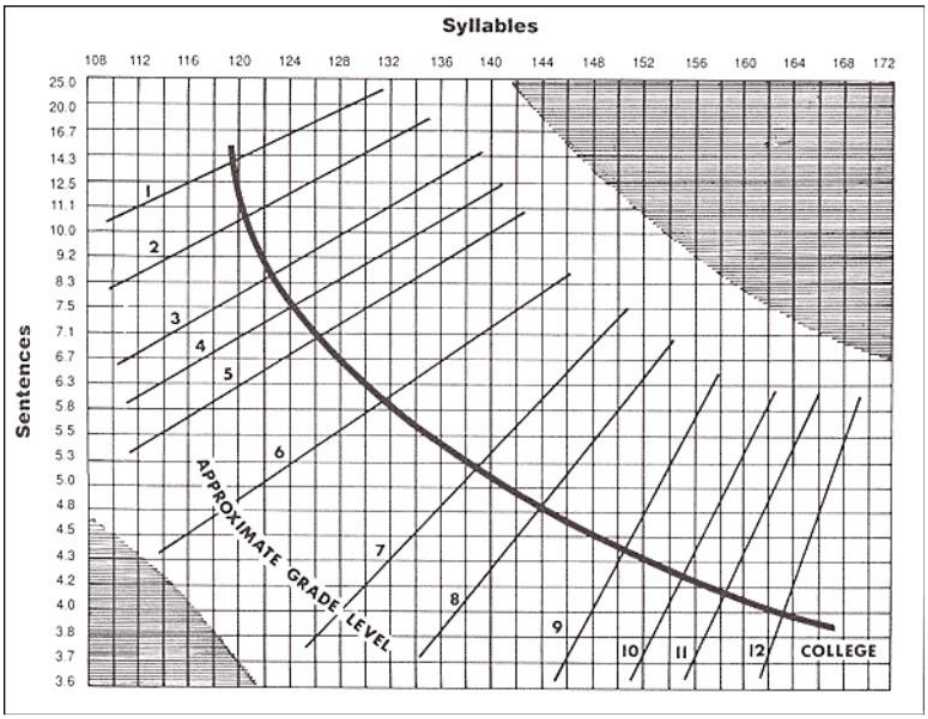

Figure 4: Fry readability Graph example (cited in Fry, 2002:5)

In 1969 McLaughlin published his **SMOG formula**. The abbreviation SMOG stands for "Simple Measure of Gobbledygook", which indicates all confusing and "unclear" words in a text (Brangan, 2011:34). The index of this formula uses years of learning required to understand a text. The SMOG formula is mostly used in the field of healthcare.

What follows is a description of further developments in the field of readability studies as outlined by DuBay (2004). Another important event in the field of readability studies that took place in this era is the revision of the **Dale-Chall formula**. In 1995 they updated their list of 3 000 words, because of the obsolescence of the old word list. Today this formula is available on the Internet, which shortens the time to calculate the words and other factors, without working simultaneously on the text and the word list.

The FORCAST formula was published in 1973, as the work of Caylor, Sticht, Fox, and Ford. It was designed to test the readability levels for adults, especially for the US military. The formula was also combined with the cloze test to increase the credibility of the results. The FORCAST formula was also used to determine the percentage of reading materials for different occupations. One of the features of the formula is that it does not use full sentence length, which makes it perfect for shorter text forms, such as Web sites, applications, and forms.

In 1975 Coleman and Liau developed the Coleman-Liau Index. The distinguishing feature of this formula is that it uses the number of characters in a text, instead of the word number.

Another commonly used readability test is **ARI – Automated Readability Index**.

ARI = 4.71 × (characters / words) + 0.5 × (words / sentences) - 21.43

ARI represents a score, which can be defined as the needed year of education to understand a text. In the table below, we can see the years of learning (age) and the prescribed grade in the US grade system.

Table 3: List of ARI scores/age and the modification to the grade level[4]

Score	Years	Grade
1	5-6	Kindergarten
2	6-7	First Grade
3	7-8	Second Grade
4	8-9	Third Grade
5	9-10	Fourth Grade
6	10-11	Fifth Grade
7	11-12	Sixth Grade
8	12-13	Seventh Grade
9	13-14	Eighth Grade
10	14-15	Ninth Grade
11	15-16	Tenth Grade
12	16-17	Eleventh grade
13	17-18	Twelfth grade
14	18-22	College

ARI also uses the variable character per word instead of the common syllable per word, which can be found in other formulas, just as the Coleman-Liau Index. For that reason, ARI is commonly used in computer readability programs.

Crossley et al. (2008) tried to show how important are the psycholinguistic factors of L2 reading, saying that readers are making a connection between text and mental representation and

[4]Table taken from http://www.readabilityformulas.com/automated-readability-index.php, April 2016

visualization of the read material. As a result, they concluded that the readers often use the previous reading experience (acquired in L1 reading), and then they transfer it to L2 reading, which brings us to the last notable trend in readability research: the Coh-Metrix Index. Crossley et al. (2008:475) defined **Coh-Metrix** as a "computational tool that measures cohesion and text difficulty at various levels of language, discourse, and conceptual analysis". They speak in favor of the Coh-Metrix index because it takes into account the psycholinguistic factors of reading comprehension, such as decoding and meaning construction within a text. They continue by explaining that the readability formulas neglect the deeper levels of text processing (cohesion, syntactic complexity, rhetorical organization, and prepositional density), which play a role in the Coh-Metrix score. Moreover, they claim that Coh-Metrix can be used for constructing simplified texts or adapting authentic texts for L2 readers, where the readability formulas are just prescriptive guides, rather than measuring tools.

As mentioned before, one of the important features of the recent readability studies is that readability formula scores are often combined with other objective testing methods (like the cloze tests) or more subjective methods (such as text leveling). The modification can be explained in terms of the modernization and a greater need for workers with high reading proficiency (DuBay, 2004:43).

In the last two decades, the need for specially designed readability formulas for ELF usage emerged. Since English became a lingua franca and a common school subject all over the world, readability researches started to widen their research subjects to foreign language students. One of the perplexing features in these studies is the wide range of conflicting results. Hamsik (cited in Greenfield, 2004) came to the conclusion that the common L1 readability formulas are suitable for EFL usage, whereas Brown's studies (1998 in Japan and 2012 in Russia) gave opposite results. In both studies Brown gathered students from different universities who could be considered as quite similar test subjects, since they share the same age, education level, and language background. These are the factors that Brown considers important for a good readability formula, which had always been neglected in the traditional readability formulas. In his study, the traditional readability formulas had a 20-30% grade appropriate correlation with the actual cloze test performance of the students, whereas his personal formula for EFL usage showed a better and larger correlation with the actual performance of the students. Below in Figure 5, his EFL formula called Brown’s EFL Difficulty Estimate is displayed. Even though this is a readability formula created for EFL usage, it was not used in this study.

Brown EFL Difficulty Estimate

EFL Difficulty = 38.7469 + (.7823 x Syllables per Sentence)
+ (-126.1770 x Passage Frequency)
+ (1.2878 x % Long Words)
+ (.7596 x % Function Words)

Figure 5: Brown's readability formula (cited in Greenfield, 2004:6)

A few years later, another important step in the EFL studies was made by Greenfield in his Miyazaki study. The study was also conducted in Japan, but only at Miyazaki International College. Greenfield (2004) criticized Brown's approach and study because his research subjects were from different colleges from all over Japan, whereas Greenfield was focusing just on this one. This is also a limitation to the study and application of its results since it can only be applied to Japanese EFL students. He also created his own EFL formula, which goes:

Miyazaki EFL Readability Index EFL Difficulty = 164.935 – (18.792×Letters per Word)
– (1.916×Words per Sentence)

Greenfield (2003) also introduced a look-up table (Figure 6) to make the application of the formula easier. To avoid the multiplication, one can easily determine the Miyazaki EFL Index score (shortened MEI in the further text and the study) by counting the average letters per word and words per sentence, and follow where those two values meet. However, the problem with this study is that it is questionable how much we can apply it in the context of this current study (or any other study) since it was designed for Japanese EFL university students. It remains uncertain if the set difficulty line with the score 50 in the MEI table can be used for another language background or another education level (elementary school in this particular study). Nevertheless, this formula is an innovative new approach to EFL readability and should be further investigated in different settings with different language backgrounds and education levels, especially when combined with reading comprehension follow-up testing to get an even better understanding of both the readers and the formula.

Figure 1. Miyazaki EFL Readability Index Look-up Table

Words per Sentence	Letters per Word																								
	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0	5.1	5.2	5.3	5.4	5.5	5.6
2		99	97	95	93	92	90	88	86	84	82	80	78	77	75	73	71	69	67	65	63	62	60	58	56
3	99	97	95	93	92	90	88	86	84	82	80	78	77	75	73	71	69	67	65	63	61	60	58	56	54
4	97	95	93	91	90	88	86	84	82	80	78	76	75	73	71	69	67	65	63	61	60	58	56	54	52
5	95	93	91	90	88	86	84	82	80	78	76	75	73	71	69	67	65	63	61	60	58	56	54	52	50
6	93	91	90	88	86	84	82	80	78	76	75	73	71	69	67	65	63	61	59	58	56	54	52	50	48
7	91	90	88	86	84	82	80	78	76	74	73	71	69	67	65	63	61	59	58	56	54	52	50	48	46
8	89	88	86	84	82	80	78	76	74	73	71	69	67	65	63	61	59	58	56	54	52	50	48	46	44
9	88	86	84	82	80	78	76	74	73	71	69	67	65	63	61	59	57	56	54	52	50	48	46	44	42
10	86	84	82	80	78	76	74	72	71	69	67	65	63	61	59	57	56	54	52	50	48	46	44	42	41
11	84	82	80	78	76	74	72	71	69	67	65	63	61	59	57	56	54	52	50	48	46	44	42	41	39
12	82	80	78	76	74	72	71	69	67	65	63	61	59	57	56	54	52	50	48	46	44	42	40	39	37
13	80	78	76	74	72	70	69	67	65	63	61	59	57	55	54	52	50	48	46	44	42	40	39	37	35
14	78	76	74	72	70	69	67	65	63	61	59	57	55	54	52	50	48	46	44	42	40	39	37	35	33
15	76	74	72	70	69	67	65	63	61	59	57	55	54	52	50	48	46	44	42	40	38	37	35	33	31
16	74	72	70	69	67	65	63	61	59	57	55	53	52	50	48	46	44	42	40	38	37	35	33	31	29
17	72	70	68	67	65	63	61	59	57	55	53	52	50	48	46	44	42	40	38	37	35	33	31	29	27
18	70	68	67	65	63	61	59	57	55	53	52	50	48	46	44	42	40	38	36	35	33	31	29	27	25
19	68	67	65	63	61	59	57	55	53	51	50	48	46	44	42	40	38	36	35	33	31	29	27	25	23
20	66	65	63	61	59	57	55	53	51	50	48	46	44	42	40	38	36	35	33	31	29	27	25	23	21
21	65	63	61	59	57	55	53	51	50	48	46	44	42	40	38	36	34	33	31	29	27	25	23	21	19
22	63	61	59	57	55	53	51	49	48	46	44	42	40	38	36	34	33	31	29	27	25	23	21	19	18
23	61	59	57	55	53	51	49	48	46	44	42	40	38	36	34	33	31	29	27	25	23	21	19	18	16
24	59	57	55	53	51	49	48	46	44	42	40	38	36	34	33	31	29	27	25	23	21	19	17	16	14
25	57	55	53	51	49	48	46	44	42	40	38	36	34	32	31	29	27	25	23	21	19	17	16	14	12
26	55	53	51	49	47	46	44	42	40	38	36	34	32	31	29	27	25	23	21	19	17	16	14	12	10
27	53	51	49	47	46	44	42	40	38	36	34	32	31	29	27	25	23	21	19	17	15	14	12	10	
28	51	49	47	46	44	42	40	38	36	34	32	30	29	27	25	23	21	19	17	15	14	12	10		
29	49	47	45	44	42	40	38	36	34	32	30	29	27	25	23	21	19	17	15	14	12	10			
30	47	45	44	42	40	38	36	34	32	30	29	27	25	23	21	19	17	15	13	12	10				

Note: 50 = average difficulty (gray boundary). Above the diagonal is easier, below more difficult.

Figure 6: Greenfield's Readability index table (Greenfield, 2003:43)

Even though a lot of people still use readability formulas (in textbook publishing, healthcare), the new and improved readability tools, such as Coh-Metrix, are creating a clear path to the new era of readability.

2.2.4. Problems with Readability Formulas

Not long after the readability formulas prospered, did the first waves of criticism come. The main complaint was that textbook publishers only used readability formulas as the sole criterion when writing textbooks, neglecting cohesion in order to get the desired readability score for the target grade. Armbruster et al. (1985) complained that the readability formulas failed to measure the right grade level because they just stay focused on word difficulty and sentence length, and do not consider the following characteristics of a text, such as content difficulty and familiarity, the organization of ideas, author style, and page layout.

Even the creators of one of the most popular formulas, Dale and Chall (1949), noticed a shortcoming of their formula: the lack of consideration for the reader and the appeal of the subject. Lorge (1949) pointed out that, not only do the readability formulas fail to evaluate the conceptual difficulty directly, but they also neglect text organization. Zamanian and Heydari (2012) also criticize the readability formulas for only focusing on the surface structure (text variables such as the numbers of sentences, words within sentences, etc), neglecting the deeper syntactic and semantic structures in the text itself, such as cohesion, complexity of ideas and a general schemata, as well as "the positioning and organization of sentences and paragraphs in the text, and information flow through the text" (Dreyer, cited in Zamanian and Heydari, 2012:47).

Fulcher (1997) had similar remarks. He complains about neglecting the reader as a very important variable when calculating readability, especially focusing on the personality of the reader as well as their motivation to understand the text and their ability to comprehend the text. He is also concerned that the readability formulas do not take into account the factor of the layout and illustrations because mostly they do help either to activate pre-knowledge or to help predict or understand the text. Legibility itself is also an important factor in determining text difficulty, and is manifested in font size and font type. Other text aspects commonly dismissed by readability formulas include the use of conceptual complexity, textual organization, word choice and syntax (Fulcher, 1997:501). Crossley et al. (2011) criticized the readability formulas by claiming that they forget to test the comprehension and lexical decoding.

One of the main problems was because textbook publishers used readability formulas as the only criterion in creating their textbooks. This usually led to a syntactically acceptable level, but without any comprehension. The second big disadvantage of readability formulas, according to Armbruster et al. (1985), was that the formulas neglect the reader, without a clear focus of the reader's motivation and interests. Armbruster et al. disapproved the readability formulas because different formulas give different readability scores (which is also the case in this study). However, they do make a point when talking about the oversimplification of the sentences and terms, especially in science textbooks. Such shortening and replacing of specialized vocabulary with an everyday and common expressions can indeed lead to even bigger problems in understanding the text. Several readability calculator pages also advise teachers and others using the calculators, not to take the readability score as the only criterion in making or editing a text, because such simplification can cause even more trouble in understanding and comprehension. Davison and Kantor (cited in Zamanian and Heydari, 2012) also criticized the abuse of readability formulas, because even if the rewritten materials are considered age or grade

appropriate, they were often more difficult, because the path to a lower readability produced materials which were harder on the deeper text levels. Crossley et al. (2008) speak in favor of Coh-Metrix, as an ideal replacement for the classic readability formulas, because it takes into account the deeper levels of text processing, which the readability formulas often neglect. Greenfield (2004) gave an illustration of the limitation of the readability formulas, by explaining they only predict the statistical correlations and difficulty, based on bare textual factors and variables, but they do not provide or explain the causes of those results. Brown (1998) also criticizes the readability interpretation scales by saying that they are very hard to apply to an L2 research environment since they all are created to fit the native speakers' school grades.

To create better and more advanced readability formulas, Brown (1998) suggested creating EFL readability formulas for each language background. Fuchs et al. (1983) gave support to such an idea of the importance of the language background since it was proven that the students' background does affect their readability level (even though the study was conducted in Minnesota and New York, meaning both native speakers).

Brown (1998) and Greenfield (2004) state that another shortcoming of readability research, especially in the EFL context, is the exclusiveness of the results, meaning that the results gathered in their two studies in Japan cannot be generalized for other countries, like Croatia. Another critical remark regarding the readability formulas was directed toward the diversity of the interpretation scales and indexes. The thought behind this is that some indexes (like FRE and MEI, as shown in the study), have a decreasing scale, whereas all the others have an increasing scale, focused on grades, which often gives reversible and negative results and correlation during the statistical interpretation and calculation. Fuchs et al. (1983) also addressed the shortcomings of the readability indexes, meaning that the formulas themselves do not agree upon the 'same' grade for a given text, causing confusion, both in the application of formulas, as well as the interpretation and text-reader matching process. Bertram and Newman (cited in Zamanian and Heydari, 2012), as well as Hamsik (cited in Greenfield, 2004), exemplify that argument of exclusiveness by claiming that some formulas do not have enough statistical background, thereupon, no direct validity.

Another big controversy dealing with readability formulas happened at the end of the 1960s, during the birth of the Plain English movement. As Crystal (1995) states, the language used by government departments was unnecessarily complicated in order to sound more formal. Crystal (1995) claims that the main argument of the government officials was the genuine nature of the

official language, which has to be used in such official affairs. DuBay (2007) explained that even simple instruction manuals (such as for car seats) were unnecessarily complicated, which lead to accidents and misuse of items, caused by the overcomplicated language. Both Crystal and DuBay give examples of other fields in which the plain language movement flourished, such as insurance policies, hire-purchase documents, licenses, contracts, guarantees, safety instructions, and many other documents which define our rights and responsibilities (Crystal, 1995:377). The government embraced the readability formulas, because they realized, that by using a simpler language in legal documents, they could avoid many misunderstandings and complication later on. However, many people, especially scholars, attacked that overuse of readability formulas by writing negative articles about the formulas and warning about their excessive usage, which led to a conflict between writers and publishers, claiming that readability formulas demolish all the style of an individual writer (DuBay, 2004).

2.3. Application of Readability Formulas in EFL

Johnson (1998) gave his definition of readability, focusing on the writer and his intent of transmitting selected information toward the reader, for the author's success in delivering the intended message induces readability. The importance of conveying the message to the reader is even more important when it comes to an EFL classroom, where the readers have to decode the messages and the meaning of the text in a foreign language.

According to DuBay (2004), textbook publishers struggled for years, especially in the 19th and 20th century, to match the readability level to their audience – the students. Textbooks filled with hard technical terms lead to inefficient teaching, which in turn lead to their downfall and replacement by more suitable materials. The first tools for measuring the appropriateness of a text were word frequency lists. Some readability formulas, like the Dale-Chall formula, still use such word frequency lists. However, it is important to stress that those lists are updated and supplemented with new and more used words. Those ideas all date back to Thorndike, his *Teachers' Word Book*, and his idea about the easier and faster acquisition of more common and simpler words.

Fulcher (1997) said that most teachers have trouble with assigning the right text for their students. However, when they give their learners a far more demanding text, that is inappropriate and too challenging for their current level, the learning process suffers. Most commonly this results in students losing their motivation to continue reading the text. Fulcher stresses the importance of reader based factors, especially readers' motivation, background knowledge, and general interest in the subject, followed by his or her previous reading experience and general reading skills, acquired throughout his/her education. Those factors are important because sometimes the readability formula finds a "perfect match" with an appropriate text level, but the reading comprehension could fail, based on the readers' lack of motivation or interest in the subject of the text. Fulcher's study, where he used both readability formulas and a group of judges, resulted in a clear disagreement between the objective and subjective scores. Judges mostly used the variables more important to the EFL classroom, such as the motivation of the reader, while the formulas just focused on the text itself.

One common phenomenon in EFL reading is the intuitive text simplification (Allen, cited in Crossley et al., 2011). Teachers often shorten or rewrite the text to make it easier for their students. Like we mentioned before, that (over)simplification can lead to acceptable readability

levels, but can damage the cohesion of the text. Crossley et al. (2011) explain that for EFL learners it is better when teachers use the intuitive approach, meaning, they re-edit or rewrite the text guided by subjective parameters. The second type of re-editing a text is the structural approach, where the writers or editors, even the teachers, focus on tools such as readability formulas or the Coh-Metrix index. The authors often criticize this approach, because it focuses just on the text characteristics, while completely neglecting the reader based variables, which the teachers have in mind when adapting an authentic text for an L2 classroom.

Yano et al. (1994) focus on the linguistic aspects of a reading text and how it can influence the reading achievements of EFL learners. They claim that the right text can increase the comprehensibility of non-native speakers, but that the removal of linguistic elements, which they call linguistic simplification, damages the learning process. They claim that readability formulas usually lead to such linguistic simplification. If a text has a high score on a readability test, the editors or writers will try to modify the text, usually by using simpler and more frequent words, sometimes even damaging the overall comprehensibility of the text. That oversimplification damages the learning process because the learners will not encounter enough new and more challenging materials that stimulate the learning, and prevent them from developing both their reading skills and vocabulary. However, we can exclude that case from the Croatian classrooms. Our textbooks are usually written with target vocabulary and syntax structures, as specified in the national curriculum, so the textbook authors will focus on covering those during the writing of texts.

Carrell (1987) was investigating both reading and readability in the EFL classroom. Since most of the readability formulas are primarily used in publications where English is the mother tongue, one of her remarks is that because the formulas just focus on the text itself, it is hard to transfer them in other languages. Furthermore, if the readability level of a text in English is appropriate (or even low), the text could still cause problems, especially in the domain of comprehension, because of the influence of the native tongue of the student. Another significant remark is that the readability formulas in foreign language grading are often misused, mostly causing oversimplification of texts. She explained the danger of the vicious circle of oversimplification; when teachers use a textbook with simple language, the reading skills of the students will deteriorate, which will lead to another simplification of the textbooks and so on. She claimed that the best way to adapt a text is the intuitive approach of the author, teacher or editor, especially if they have experience with students to rely on. To prevent misuse of readability formulas for adapting texts, Carrell (1987) listed different guidelines for teachers

when deciding on the reading materials for their class, especially an EFL class. Some of the guidelines include the teacher as a supporter, who provides the student with reading advice, considering the comprehension domain, as well as different lexical and syntactical factors. It is also important to note that one should not always rely just on readability formulas, especially for more advanced learners, because they could have an even richer vocabulary span than needed for the text, and as a result could find the materials boring and not challenging enough.

O'Donnell (2009) had a similar study, where she investigated the role of text modification in an L2 Spanish class. She addressed the problem of authenticity of reading materials, whether it is better to use authentic or pedagogically modified materials. The best solution for the EFL class is the usage of authentic texts, followed by slight modifications. We could discuss that it is more appropriate to use text leveling than just bare readability formulas to adapt such texts, mostly because of the latter mentioned role of the curriculum which prescribes the needed content which needs to be acquired. Crossley et al. (2011) were also researching that field, criticizing the overdependence of the readability formulas when creating or adapting texts, mostly because the formulas damage the comprehensibility, while creating a false sense of a linguistically simplified text.

When conducting research with EFL learners, most researchers focus on the correlation between readers' evaluation and the formula scores, meaning the reality (readers' performance on the cloze tests, multiple choice, or any other comprehension check) and the predictions (the estimated score from the chosen readability formula) (Brown,1998). Brown et al. (2012) conducted a study by testing students from several Russian universities. After the study, they came to the conclusion (supported by statistical backup) that the L1 readability formulas and readability indexes are more related with each other than with the cloze test results of the Russian students. They criticize such testing methods which focus on cloze test comprehension to double check the readability formulas, because they say that the readability formulas and cloze tests focus on entirely different text aspects: readability formulas focus on the text factors and word frequency, whereas the cloze tests focus on the students' proficiency and semantic levels of a text, like the appropriateness of a word in a given sentence.

Regarding the field of readability in EFL, the two main studies were both conducted in Japan. Brown's study involved Japanese students from 18 colleges in Japan and had shown that the L1 readability formulas are well connected with each other, but the cloze test performance of the students did not have such a great correlation with the predicted readability results. Brown's

formula and his variables created to fit the needs of EFL students gave better results than the traditional formulas He claimed the readability formulas had 20-30% accuracy in predicting text difficulty for EFL students, where his formula had 55% accuracy.

Jerry Greenfield introduced his EFL readability formula and following look up index in 2003. He named it Miyazaki EFL Readability Index (MEI), according to the college where he was teaching and conducting his research. The variables of his formula are very similar to Flesch Reading Ease (FRE) and so is the interpretation scale. In his look up table, he said that the average line of difficulty is the score 50. Everything above is considered from challenging enough to easy, and everything bellow is considered to be too difficult. Greenfield (2004) used Bormuth's academic text passages to cloze tests the students. He got results which are the opposite of Brown's study: readability formulas have a high correlation with academic texts for both native and EFL readers. He explained that the results were so different, not only because they were both focusing on different aspects of readability formulas in EFL usage (Greenfield was testing the older results from native speakers and the new ones for Japanese students, while Brown used random passages with no L1 background research to see just the level of application for the L2 context), but their number of students used in the studies also affected the outcome (Browns results are lower because he used a larger group of students, which dilute the final results). The main difference between the texts is that Brown used general texts in English so that the results could be generalized for almost any type of English written material, while Greenfield was focusing on academic texts. Even though we could say that the two studies are similar, we see in the different results that some factors, like the examinees and texts used in the studies, which could change the overall result. Greenfield also posed a question regarding cloze testing which needs to be investigated: whether the non-English speaking students try harder to solve the tasks than natives.

3. The present study

3.1. Aim

The main aim of this research is to offer an overview of the Croatian EFL textbook texts, by presenting and comparing their scores on different readability tests. The tests used in this study are: Flesch Readability Ease (**FRE**), Miyazaki EFL Index (**MEI**), Flesch-Kincaid Grade formula (**FKGL**), Dale-Chall formula (**DCF**), Gunning Fog formula (**FOG**), SMOG formula (**SMOG**), and Automatic Readability Index (**ARI**).

The following research questions were posed:

1) What are the mean scores of different tests for different textbooks? By answering this question, we will also get a ranking list of the textbook series according to their difficulty. The focus will be directed towards the ranking of the textbooks, followed by comparing the differences between scores. It is expected that the two Croatian textbooks will have similar results when compared to the foreign *Project* series.
2) How do the same grades compare across different publishers? How great are the differences between the scores for the same grade from another series? The expected results are that there will not be great differences between the publishers, but a certain degree of difference between the grades is expected. This step is very useful for the field of EFL studies because it will give a clear overview of textbook difficulty for a specific grade, which could help a teacher to decide which textbook he or she will use, according to the needs and reading skill of their students.
3) Are there any irregularities between the grades themselves within a series? Do grades within a series significantly differ from the preceding and succeeding grades from the same series? The initial hypothesis for this question is that there will be a significant increase in difficulty with each succeeding grade.
4) Is there any correlation between years of learning (**YOL**) and the (FRE, MEI, FKGL, DCF, FOG, SMOG, ARI) test scores? It is expected that the grade scale tests (FKGL, DCF, FOG, SMOG, and ARI) will have a positive correlation since both the test grades and YOL will increase.

5) How are the average text factors distributed in each series and grade? Which textbook has the longest texts? Based on subjective judgment, the *Project* series will dominate in the text factors categories.
6) Is there a correlation between text factors (words, sentences, words per sentence, syllables per word, and characters/letters per word) and YOL? How big is the increase of the text factors with each following grade? We shall focus on finding significant differences in text factors categories and presenting them.

3.2. Sample/Corpus

Text samples used in this research are reading texts from Croatian EFL elementary school textbooks. Three different textbook series were used: the *Project* series by Tom Hutchinson, a foreign textbook author working for Oxford University Press, and two national series: *Way to go plus* and *Dip in*, both by the Croatian textbook publisher *Školska knjiga*.

Textbooks from grade 4-8 from each series were used as a corpus sample, which gives us a total of 15 textbooks used in this study. Three subtypes of reading texts used in this study are:

1) A story (with plot, characters, etc.)

2) A plain text about some historical, geographical or social events

3) Narrations and descriptions, such as simple personal stories about an event or personal descriptions for both people and places

The latter mentioned category was mainly used for the 4th and 5th grade to get the necessary number of texts for each textbook, mostly because in those textbooks most of the reading materials are comics or dialogues, all followed by tape recordings to support the reading with listening. For the upper grades (6-8), a personal selection was used, mostly choosing texts that had the headline *reading practice*.

All those categories lead us to a corpus of 141 text samples in total. The number of texts per grade is displayed in Table 4 bellow:

Table 4: Number of text samples per grade

Grade	Number of texts
4th	6
5th	8
6th	10
7th	10
8th	13

3.3. Instruments and Procedure

The texts were typed and stored in Microsoft Word. The readability tests were conducted via the page Readability Score (https://readability-score.com/) and Readability Formulas (http://www.readabilityformulas.com/free-readability-formula-tests.php). The Miyazaki EFL Index (MEI) was calculated in Microsoft Excel.

The readability test (and their abbreviations, used in the further text) used in this research were: Flesch Readability Ease (**FRE**),Miyazaki EFL Index (**MEI**), Flesch-Kincaid Grade Formula (**FKGL**), Dale-Chall formula (**DCF**), Gunning-Fog Formula(**FOG**), SMOG formula (**SMOG**), and Automatic Readability Index (**ARI**).

The other data about the texts, such as the total number of characters, syllables, sentences and words in the whole text, as well as the number of characters and syllables per word, or words per sentence, were calculated by using Readability Score. Most of the means, as well as the ANOVA and Pearson correlation, were calculated using SPSS. For the first three research questions, we focused mostly on the FRE and MEI results, and FKGL and DCF results (two grade level scales). The first two tests were chosen to be grouped together because they both are decreasing numeric scales. We were interested in analyzing the MEI scores since it is a formula created for Japanese university students and we applied it using elementary school reading material. We also ran an ANOVA and Post hoc Tukey test to determine whether there was a significant difference between the grades across and within a series, as well as to determine the degree of difference between the text factors based on textbooks and YOL.

3.4. Results

To determine which is the easiest and hardest series in terms of text readability, the means of different textbooks were calculated. Table 5shows the three textbook series and their final means of all the tests used in this study. The first two tests, FRE and MEI, have a decreasing numeric scale, which means that the lower the score, the more difficult the book. The remaining tests have a grade level scale and are easy and transparent to interpret: the higher the score, the more difficult the textbook.

In all readability tests (FRE, MEI, FKGL, DCF, FOG, SMOG, and ARI) the *Dip in* series is the hardest and *Way to go plus* is the easiest. The initial expectation, that the *Project* series will be the hardest, is hereby dismissed. One of the hypotheses that the Croatian national series will have similar scores to the foreign *Project* series is also dismissed because they happened to have the biggest difference in scores.

The FRE scores indicate that the textbooks could be appointed to the categories fairly easy, with the average text difficulty appropriate for 7th graders. The MEI scores could be reported as borderline easy since they clearly make it above the average difficulty line, but it is questionable if we can draw such a conclusion since we are talking about elementary school students. Even though MEI is the only readability formula for EFL use used in this study, it remains unclear the extent to which it can be applied and used for different age groups.

Table 5: Textbook means in all readability formulas

Textbook	FRE	MEI	FKGL	DCF	FOG	SMOG	ARI
WTG	82.88	68.67	4.24	5.43	3.23	5.03	2.89
DIP	75.24	63.61	5.5	6.03	7.37	6.12	5.24
PRO	80.54	65.46	4.89	5.81	6.71	5.41	3.73

By using descriptive statistics in SPSS, we got average FRE and MEI values for each textbook of the three series. When examining the FRE results, we detect a regular pattern of decreasing FRE scores with each following grade. However, in two cases (*Dip in* and *Project*) we notice that the 5th grade textbooks are more difficult than the 6th grade textbooks. This "6th grade case" could be just a coincidence, since we do not know whether the textbook publishers in Croatia, in this case, *Školska knjiga*, use readability formulas for their EFL textbooks.

When we look at the MEI results, we see that they are considerably lower than the FRE results. Generally speaking, the MEI results have a clear decreasing line, meaning that each grade becomes more difficult. We do not have the 6th grade case as in the MEI results. Further research in the field of FRE and MEI is required so that we can discover the degree of application of the two formulas together. It is also questionable whether we should call the 8th grade textbooks borderline hard (based on the original Japanese university students who were the measuring standard for creating MEI) or we could say that our elementary school EFL textbooks are at the level of some EFL materials designed for higher education in some other countries.

Table 6: FRE and MEI scores for each series

TB	FRE	MEI
WTG1+	93.72	82.89
WTG2+	92.38	76.62
WTG3+	83.56	66.83
WTG4+	76.49	62.74
WTG5+	68.24	54.28
DIP4	85.08	76.5
DIP5	72.44	64.68
DIP6	78.84	64.49
DIP7	70.09	57.7
DIP8	67.78	54.7
PRO4	87.22	74.7
PRO5	80.08	65.23
PRO6	80.86	63.18
PRO7	77.4	62.14
PRO8	77.12	60.06

The next table shows the mean values of the textbooks in the FKGL and DCF tests. The first thing visible in the FKGL results is that they correspond to the FRE listing of the hardest and easiest textbooks. The *Way to go* series has lowest scores from grades 4-6, and the *Project* series has the lowest scores for the 7thand 8thgrade. When comparing the given US grades, we notice that most of the FKLG US grades vaguely correspond with the intended Croatian grade, where the DCF grades have a higher accuracy, but only for the 4th, 5th, and 6th grade. The results could also be linked to the year of learning English since this is an EFL study.

Table 7: The textbooks with their mean FKGL and DCF scores

Grade	TB	FKGL	DCF
4th	WTG1+	1.82	4.68
	DIP4	3.23	4.37
	PRO 4	3.25	4.67

5th	WTG 2+	2.54	4.49
	DIP5	5.48	5.71
	PRO 5	4.90	5.56
6th	WTG 3+	4.24	5.11
	DIP 6	4.96	6.19
	PRO 6	5.10	6.22
7th	WTG 4+	5.52	6.27
	DIP 7	6.81	6.98
	PRO 7	5.38	6.40
8th	WTG 5+	7.07	6.59
	DIP 8	7.01	6.88
	PRO 8	5.82	6.18

For the next stage of this research, we shall focus on the series and the relationships between the grades. To discover if there is any significant increase in the readability difficulty with each succeeding grade, we ran One-way ANOVA, followed by Post hoc Tukey. The textbooks were used as the main variable, where all 15 textbooks were compared among themselves, and we simultaneously obtained the statistical data for the differences between grades within a series and grades across the series. The ANOVA results, displayed in Table 8, are significant with differences across the groups. However, at this stage of the study, we are just interested in the differences between grades within a series, whose Post hoc Tukey results will be presented in the following tables, to find the exact cases with significant differences. In the Post hoc tables, the FRE results will be reported.

Table 8: One way ANOVA results for all the tests with textbooks as the key factor

Variable		**df**	**Mean Square**	**F**	**Significance**
FRE	Between Groups	14	563.25	8.56	.00
	Within Groups	126	65.84		
MEI	Between Groups	14	581.99	11.35	.00
	Within Groups	126	51.29		
FKGL	Between Groups	14	21.62	10.75	.00
	Within Groups	126	2.01		
DCF	Between Groups	14	6.61	5.75	.00
	Within Groups	126	1.15		
FOG	Between Groups	14	21.21	12.34	.00
	Within Groups	126	1.72		
SMOG	Between Groups	14	13.95	10.02	.00
	Within Groups	126	1.39		
ARI	Between Groups	14	38.25	11.57	.00
	Within Groups	126	3.31		

The results for the *Way to go* series Post hoc test are presented in Table 9. By observing the scores, we detect a few cases with significant results. Most of the cases are connected with

WTG4+ (7th grade) and WTG5+ (8th grade). However, we see that the scores only have a significant correlation with the lower or higher grades, for example, the 4th and 8th grade, but not between the immediate grades, such as the 4th and 5th grade. There are also many negative results, caused by the reversibility of the results, where the negative scores mean that in this case the textbook is easier, and in the positive case the textbook is harder.

Table 9: *Way to go* Post hoc Tukey scores

Textbook	**Correlation**	**Mean Diff.**	**Sig.**
WTG1+	WTG2+	1.34	1.00
	WTG3+	10.6	.50
	WTG4+	17.23*	.01
	WTG5+	25.48*	.00
WTG2+	WTG1+	-1.31	1.00
	WTG3+	8.82	.60
	WTG4+	15.89*	.01
	WTG5+	24.14*	.00
WTG3+	WTG1+	-10.57	.50
	WTG2+	-8.82	.60
	WTG4+	7.07	.82
	WTG5+	15.32*	.00
WTG4+	WTG1+	-17.23*	.01
	WTG2+	-15.89*	.01
	WTG3+	-7.07	.82
	WTG5+	8.25	.51
WTG5+	WTG1+	-25.48*	.00
	WTG2+	-24.14*	.00
	WTG3+	-15.32*	.00
	WTG4+	-8.25	.51

The same approach regarding difficulty within the series was followed for the next two cases as well. For the *Dip in* series, we notice the same starting trend: the textbooks for the 7th and 8th grade are the only ones with a significant correlation to the textbook for the 4th grade. One thing that differs from the rest is the fact that DIP6 is easier than DIP5, which means that the texts for the 5th grade have a higher readability score in the tests than those for the 6th grade. Just as the previous case with the *Way to go* series, results are reversed and presented in negative values for the last two textbooks.

Table 10: *Dip in* Post hoc Tukey scores

Textbook	**Correlation**	**Mean Diff.**	**Sig.**
DIP4	DIP5	12.65	.22
	DIP6	6.24	.98
	DIP7	14.99*	.04
	DIP8	17.31*	.00

DIP5	DIP4	-12.65	.22
	DIP6	-6.40	.94
	DIP7	2.34	1.00
	DIP8	4.66	.99
DIP6	DIP4	-6.24	.98
	DIP5	6.40	.94
	DIP7	8.75	.51
	DIP8	11.06	.09
DIP7	DIP4	-14.99*	.04
	DIP5	-2.35	1.00
	DIP6	-8.75	.51
	DIP8	2.3	1.00
DIP8	DIP4	-17.31*	.00
	DIP5	-4.66	.99
	DIP6	-11.06	.09
	DIP7	-2.31	1.00

Table 11 contains the Post hoc results for the *Project* series. We do not have any significant correlations in this series. An interesting factor is that PRO5 is harder than PRO6, the same occurance that happened in the *Dip in* series. That phenomenon shall be referred to as "the 6th grade case".

Table 11: *Project* post hoc Tukey scores

Textbook	**Correlation**	**Mean Diff.**	**Sig.**
PRO4	PRO5	7.14	.95
	PRO6	6.36	.97
	PRO7	9.82	.56
	PRO8	10.10	.43
PRO5	PRO4	-7.14	.95
	PRO6	-0.79	1.00
	PRO7	2.68	1.00
	PRO8	2.96	1.00
PRO6	PRO4	-6.36	.97
	PRO5	0.79	1.00
	PRO7	3.46	1.00
	PRO8	3.74	.99
PRO7	PRO4	-9.82	.56
	PRO5	-2.68	1.00
	PRO6	-3.46	1.00
	PRO8	0.28	1.00
PRO8	PRO4	-10.10	.43
	PRO5	-2.96	1.00
	PRO6	-3.74	.99
	PRO7	-0.28	1.00

To determine the difficulty difference between the grades, based on the mean FRE scores, we ran the ANOVA test (Table 8), followed by Post hoc Tukey. For the 4th grade, we see that WTG1+ is still the easiest textbook, while DIP4 is the hardest one, corresponding with the FRE means. In the 5th grade, we detect a statistically significant result: DIP5 is significantly harder than WTG2+. The remaining grades do not have any significant results. On the contrary, we even find negative values for PRO7 (-0.91) and PRO8 (-8.88), which means that those textbooks are easier than their *Way to go* equals, supporting the results of the FRE mean scores.

Table 12: Post hoc Tukey test with correlation between different textbooks

WTG textbook	**Correlated textbook**	**Tukey score**
WTG1+	DIP 4	8.63
	PRO 4	6.50
WTG2+	DIP 5	19.94*
	PRO 5	12.30
WTG3+	DIP 6	4.72
	PRO 6	2.70
WTG4+	DIP 7	6.40
	PRO 7	-0.91
WTG5+	DIP 8	0.46
	PRO 8	-8.88

In the next stage of this research, we focused on the variable **YOL– years of learning**. Since the textbooks used in this research are written and designed for EFL students, it is clear that they must correspond with a given year of learning English, since it is not a native language in this country. Textbooks for EFL usage are written in a specific way, to cover the topics, vocabulary and grammar prescribed by the national curriculum. With that being the first guideline when creating EFL textbooks, it is easy to assume that the second guideline is to make the textbook appropriate for the given grade, de facto the given year of learning. Since it would be impossible to determine the YOL variable for native speakers, because it would correspond to their chronological age, here we have the unique opportunity to see the relationships between YOL and the readability tests. Table 13 is a display of the average FRE, MEI, FKGL, and DCF scores for each year of learning. The focus in this part of the research are not the series themselves, but their average scores on different test and their placement in the average year of learning. Like the average results of the series in Table 5, we see that the MEI results are lower than the FRE results. In FRE we almost have identical scores for the 5th and 6th grade, caused by the 6th grade

case. When comparing the FRE and FKGL mean scores to the expected US and Croatian grades (where the Croatian grades are corresponding to YOL), we notice that most average FRE scores belong in the categories "easy" and "fairly easy" and are on the readability level meant for 6th and 7th graders, relatively corresponding to the Croatian grades.

Table 13: YOL and average tests scores

YOL	FRE	MEI	FKGL	DCF
4	88.67	78.03	2.77	4.57
5	81.63	68.84	4.30	5.25
6	81.09	65.50	4.77	5.84
7	74.66	60.86	5.90	6.55
8	71.04	56.35	6.63	6.55

Table 14 shows us the Pearson Correlation between the tests themselves and each of the tests and YOL. All the results are significant. The YOL correlates with all the tests, but based on the moderate results, we could argue that there are significant correlations, indicating that when the YOL increases, the test results (meaning, the difficulty of the reading material) will also increase. The results for FRE and MEI are negative because they are reverted indexes and have larger average scores than the rest of the tests, basically, where both YOL and all the test scores increase indicating harder material, the FRE and MEI scores decrease. However, a higher correlation and significance was expected. We see that YOL correlates best with MEI, ARI, and FKGL. A surprising result is that YOL has a higher correlation with the MEI results than with the FRE results. By observing the table, we can see that ARI has the highest correlations with all the tests.

Table 14: The Pearson Correlation between YOL and all the readability tests

		FRE	MEI	FKGL	DCF	FOG	SMOG	ARI
YOL	Pearson Correlation	-.52**	-.67**	.63**	.53**	.64**	.59**	.67**
FRE	Pearson Correlation		.88**	-.94**	-.67**	-.84**	-.91**	-.88**
MEI	Pearson Correlation			-.96**	-.65**	-.87**	-.86**	-.99*
FKGL	Pearson Correlation				.68**	.91**	.93**	.96**
DCF	Pearson Correlation					.53**	.64**	.65**
FOG	Pearson Correlation						.91**	.88**
SMOG	Pearson Correlation							.87**

In the last stage of this study, we investigated the common variables of all the readability formulas: words, sentences, words per sentence, syllables per word, and characters/letters per word. Table 15 presents the mean values of each of these text factors for each of the three series. The *Project* series has the longest texts (it was anticipated in the initial stages of the study while gathering texts). Even though *Project* has the longest texts, we could argue that the *Dip in* series has more complex texts, since the values for WpS (words per sentence), SypW (syllables per word), and ChpW (characters per word) are partially higher than those for the *Project* series.

Table 15: Text factor means per series

	Words	**Sents**	**WpS**	**SypW**	**ChpW**
WTG	237.51	22.02	10.8	1.36	4.15
DIP	308.19	27.38	13.88	1.43	4.29
PRO	320.17	27.45	11.89	1.36	4.14

Table 16 contains the results for each textbook. Just like in the previous table, the *Project* series dominates in the categories words and sentence. We see clear increases with each succeeding grade, making the texts more demanding for the readers. Here we can observe the differences between grades and publishers, concluding that the *Project* series has the longest text with most words and the largest average of sentences per text. The *Dip in* series has sometimes, even more sentences per text, but has the largest scores in all the other categories. The *Way to go* scores are around the average scores for some factors, but we still could consider this series the easiest of them all, regarding the text factors.

Table 16: Textbooks and their mean values in the text factors categories

Grade	**TB**	**Words**	**Sents**	**W p S**	**Syl p W**	**Ch p W**
4th	WTG1+	135.83	21.5	6.53	1.27	3.7
	DIP4	257.83	24.83	7.42	1.27	3.7
	PRO4	190.83	23.5	8.68	1.31	3.92
5th	WTG2+	155.63	18	8.7	1.25	3.81
	DIP5	295.75	26.38	11.5	1.25	3.81
	PRO5	223.38	19.75	11.34	1.35	4.15
6th	WTG3+	254	23.2	10.6	1.33	4.14
	DIP6	279.6	26	10.84	1.33	4.14
	PRO6	320.4	25.9	12.54	1.34	4.03
7th	WTG4+	260.9	23	11.85	1.41	4.23
	DIP7	313.5	23.7	13.4	1.41	4.23
	PRO7	336.3	28.9	11.77	1.4	4.27

8th	WTG5+	304.15	23.08	13.39	1.48	4.52
	DIP8	357	28.46	21.04	1.48	4.52
	PRO8	426.85	34.07	13.32	1.38	4.22

Even though we see a clear increase in difficulty with each successive grade, a One-way ANOVA test was conducted to see if there are any statistically significant differences between the text factors in the textbooks. We have significant differences in three text categories: words, syllables per word, and characters per word. Nevertheless, it should be stressed that those significant results in the Post hoc Tukey lie between lower grades (e.g., WTG1+) and higher grades from another series (DIP7), so they will not be reported here, but will be included as an appendix (Appendix 1).

Table 17: One-way ANOVA differences of text factors in the textbooks

Variable		**df**	**Mean square**	**F**	**Significance**
Words	Between Groups	14	53635.69	5.76	.00
	Within Groups	126	9313.20		
Sentences	Between Groups	14	198.25	2.07	.02
	Within Groups	126	95.89		
Words per sentence	Between Groups	14	116.07	1.38	.18
	Within Groups	126	84.38		
Syllables per word	Between Groups	14	.05	5.88	.00
	Within Groups	126	.01		
Characters per word	Between Groups	14	.51	7.11	.00
	Within Groups	126	.07		

Table 18 is an overview of the average text factors for each year of learning. We do not have any cases where the 6th grade case played a role in the results, meaning all the results are on a clear ascending path. However, it is interesting to note that the average sentence length for the 4th year of learning is larger than the one for the 5thyear of learning, probably because it is easier for beginners to read shorter sentences.

Table 18: Text factor averages per YOL

YOL	**Words**	**Sentences**	**WpS**	**SypW**	**ChpW**
4	194.83	26.61	7.54	1.31	3.86
5	224.92	21.38	10.51	1.33	4.04
6	284.67	25.03	11.33	1.36	4.14
7	303.57	25.20	12.34	1.43	4.28
8	362.67	28.54	15.92	1.45	4.43

To see if there is any correlation between the year of learning and the (increase) in text factors, we ran a Pearson's Correlation. Most of the results have a 2-tailed significance. YOL correlates positively with all the text factors, meaning that with every following year of learning, the numbers of the text factors increase, widening the vocabulary and syntax of a text. All the results are significant, except sentences. Sentences have negative values with inner sentence factors (WpS) and the factors regarding the words in the sentence. The reason behind that could be that the inner sentence factors are diminished by the increase of the total number of sentences within a text. Although many results have very weak correlations, we can observe that YOL has moderate correlations with the total number of words in the text, as well as characters per word (Ch p W), supporting the theory that words do get more complicated in each grade, as well as the fact that the texts get longer. The strongest correlation in this table is the correlation between syllables and characters (letters) per word, also supporting the theory of more complex words, because when the number of characters per word increases, that means that automatically the number of syllables within the word will increase.

Table 19: YOL and text variables correlations

	Words	**Sents**	**W p S**	**Syl p W**	**Ch p W**
YOL	. 49**	. 14	.28**	.47**	.56**
WORDS		.82**	.01	.23**	.27**
SENTS			-.15	-.04	-.04
WPS				.15	.24**
SYL P W					.86**

After investigating the nature of the relationships between YOL and the text factors, the last One-way ANOVA will focus on the differences between the text factors with YOL as the key variable. In Table 20 we can observe that the results are similar to the ANOVA results in Table 17, where the same three categories have significant results. Since the only significant differences are similar to the results of the Post hoc Tukey results for textbooks and FRE scores, meaning the significance occurs only for the 4th and 7th or 8th grade, we will include them as an appendix (Appendix 2).

Table 20: One-way ANOVA differences of text factors by YOL

Variable		**df**	**Mean square**	**F**	**Significance**
Words	Between Groups	4	119181.51	11.20	.00
	Within Groups	136	10644.39		
Sentences	Between	4	199.49	1.93	.11

	Groups				
	Within Groups	136	103.38		
Words per sentence	Between Groups	4	255.07	3.09	.02
	Within Groups	136	82.62		
Syllables per word	Between Groups	4	.09	10.36	.00
	Within Groups	136	.01		
Characters per word	Between Groups	4	1.29	15.86	.00
	Within Groups	136	.08		

3.5. Discussion

One of the first significant findings of this study is the high variation between both the series and the grades. The first complaint is about the *Way to go plus* series. Since the series has the lowest scores in both the readability tests and text factors, we could say that is generally a very easy series of textbooks. That, however, is not a bad thing, because it can be used perfectly in a classroom with weaker readers. Furthermore, if *Way to go plus* is used from the 4th grade for the 2nd foreign language, then we could argue that the usage of the book is appropriate. Otherwise, the textbook could damage the reading process, as well as the general language acquisition and learning, because it could be too easy and not challenging enough. Just like Yano et al. (1994), we encountered a few cases of oversimplification, mostly in the 4th and 5th grade. It is important to note that those cases occurred with other textbooks as well, not just *Way to go*. Following the results in Table 12, a teacher could use different textbooks for different grades, based on the language level of the students, always providing the students with a suitably challenging text.

One of the main findings of this study is the "6th grade case". It is still unclear if that is just a coincidence based on the scores of the selected texts, of if there are deeper issues that need to be investigated, but, nevertheless, further research in this area is needed. Henceforth, it would be interesting to see if the results would stay the same if we were to change some of the texts. It also remains unclear if the texts just have lower scores in the readability formulas because their syntax and words are simpler, not meaning that their rhetorical organization and flow of thoughts, as well as the listing of ideas and information, is written in a simpler way. One of the

reasons behind these scores could be Thorndike's idea of vocabulary teaching and the role of word frequency. The publishers may deliberately use simpler and previously known words to facilitate and recycle them before moving to the acquisition and learning of new words. Another reason behind this could be that the texts indeed are written with simpler and familiar words to facilitate the vocabulary and grammatical structures, before moving to new material in the last two grades. Given these points, further research with these texts (and their readability scores) and examinees is needed to discover how readability formulas match texts to the needs of Croatian EFL students.

The results in Table 5 and Table 13 show us how great the difference between different readability tests can be. It is understandable why they are often criticized since they give us different scores for different grades. However, we see that all the tests have significant correlations between themselves. If a teacher wants to use readability tests for his/her students, then he/she should be consistent in just using one type of test to avoid such ambiguous situations.

Another important discovery of this study is also the average test score per YOL. Table 13 provides us with a clear insight how hard or easy the average textbooks are for a specific year of learning and grade. It enabled us to see the average text difficulty load per grade. However, the sample in this study is too small, so the results cannot be generalized for all the EFL textbooks in Croatia or further.

After discovering the MEI index, it is still questionable how much we can apply MEI in this or any other study. MEI and FRE need many more studies where they are used together and where we can compare their results with subjective measuring carried out with different learners who differ according to years of learning, and backgrounds in order to prove the impact of the first language on learning and reading in a foreign language. However, MEI is very useful, because of the idea that readability formulas should be adapted to fit the needs of non-native readers. Although readability formulas have been used for over 60 years, there is still the need to develop a formula that takes into account the individual differences of a reader/learner, which is almost impossible, since there is no possible way to measure and adapt a formula for each individuals' individual differences.

Since this is the first time in Croatian EFL that a readability score database has been created with textbook texts, we could say that it is an important step in readability research in Croatia. The different readability scores could be used as a guideline for choosing textbooks in the future. Of course, that means that we need to calculate all the texts not only in these three series but the

other series also used in Croatia, such as *Building Bridges*. If there was an index with scores for the textbooks, there could be further research with a more subjective tone. It could be the usage of the cloze test, where different groups of students would solve cloze tests to double-check the appropriateness of a text. Text leveling or other subjective checklists with different factors considering the reader and not just the text itself should also be used. However, it is hard to believe that even the publishers would be interested in such long and demanding research.

4. Conclusion

This paper has shown a short overview of the history of readability formulas and readability research, followed by a completely new study done in Croatia. The elementary school textbooks have a great variety of reading materials (dialogues included), and now they have matching readability scores for future reference when deciding on the textbook which will be used. We found significant correlations between the readability formulas themselves, discovered that the *Dip in* series is the hardest, *Project* is the series with the longest texts, and *Way to go plus* is the easiest series. Furthermore, the differences between grades and across grades are presented, confirming the hypothesis that the text difficulty increases with succeeding grades. The main discovery of the study is the 6th grade case, which opens the door to further research in the field of readability in Croatia. The last significant discovery is that we found a significant correlation between sentences, words, and years or learning.

The readability formulas should be used further to assign readability scores to other reading texts we plan to use in the classroom. Teachers should also take into account the preferences of their students when selecting the texts, mainly focusing on other factors which are often neglected by readability formulas, such as motivation and pre-knowledge of the reader.

Even though the readability formulas have been used for over half a century, we can say that readability is still a rich area for future research, with a countless number of possible corpora and with a great variety of formulas to be used. However, to get even better results, the researchers should combine the purely mathematical readability formulas with more subjective measurement tools, as well as tests to check the reading comprehension.

5. Bibliography

Alderson, J Charles, Jayanti Banerjee (2001). Language Testing and Assessment-Part I, *Language Teaching, Cambridge University Press* 34: 213-236.

Anić, Višnja(2009). *Way to go 2 plus, udžbenik engleskog jezika za 5.rezred osnovne škole, 5. godina učenja*. Zagreb: Školska knjiga.

Anić, Višnja (2010). *Dip in 7, udžbenik engleskog jezika za 7.razred osnovne škole, 7.godina učenja*. Zagreb: Školska knjiga.

Armbruster, Bonnie, Jean Osborn, Alice Davison (1985). Readability Formulas May Be Dangerous to Your Textbooks. Aviable at: http://www.ascd.org/ASCD/pdf/journals/ed_lead/el_198504_armbruster.pdf (visited in August,2015).

Ban, Suzana (2010). Dip *in 5, udžbenik engleskog jezika za 5. razred osnovne škole, 5. godina učenja*. Zagreb: Školska knjiga.

Ban, Suzana (2009). *Dip in 4, udžbenik engleskog jezika za 4.razred osnovne škole, 4.godina učenja*. Zagreb: Školska knjiga.

Brangan, Sanja (2014). Kvantitativna procjena težine teksta na hrvatskom jeziku. *Rasprave, časopis Instituta za hrvatski jezik i jezikoslovlje*, 40, 1: 35-58.

Brangan, Sanja (2011). *Razvoj formula čitkosti za zdravstvenu komunikaciju na hrvatskom jeziku*. Aviable at: http://medlib.mef.hr/1414/1/Brangan_S_disertacija_rep_1414.pdf (visited in August, 2015).

Breka, Olinka (2008). *Way to go 3 plus,udžbenik engleskog jezika za 6.razred osnovne škole, 6. godina učenja*. Zagreb: Školska knjiga.

Breka, Olinka, Maja Mardešić (2009). *Way to go 4 plus, udžbenik engleskog jezika za 7.razred osnovne škole, 7. godina učenja*. Zagreb: Školska knjiga.

Breka, Olinka (2010). *Dip in 8, udžbenik engleskog jezika za 8.razred osnovne škole, 8.godina učenja*. Zagreb: Školska knjiga.

Brown, James Dean (1998). An EFL readability index. *Japan Association for Language Teaching Journal* 20, 2: 7-36.

Brown, James Dean, Gerriet Janssen, Jonahtan Trace, Liudmila Kozhevnikova (2012). A Preliminary study of cloze procedure as a tool for estimating English readability for Russian students, *Second Language studies* 31, 1: 1-21.

Carell, Patricia (1987). Readability in ESL. *Reading in a Foreign Language* 4, 1: 21-40.

Crossley, Scot, David Allen, Danielle McNamara(2011). Text Readability and intuitive simplification .*Reading in a Foreign Language* 32, 1: 84-101.

Crossley, Scot, Jerry Greenfield, Danielle McNamara (2008). Assessing Text Readability Using Cognitively Based Indices, *TESOL Quarterly* 42, 3: 475-493.

Crossley, Scott, Max Louwerse, Phillip McCarthy, Danielle McNamara (2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal* 91: 15-30.

Crystal, David, (1995).*The Cambridge Encyclopedia of the English Language*. Aviable at: https://www.scribd.com/doc/77462665/The-Cambridge-Encyclopedia-of-the-English-Language (visited in August 2015).

Dale, Edgar, Jeanne Chall (1948). A formula for predicting readability. DuBay, William H., eds. (2006). *The Classic Readability Studies*. Costa Mesa: Impact Information, 63-75.

Dale, Edgar, Jeanne Chall (1949). The Concept of Readability. Dale, Edgar, eds. *Readability-National Conference on Research in English; National Council of teachers of English*. Champaign: 1-7.

Dubay, William H., (2004). *The Principles of Readability*. Costa Mesa: Impact Information.

DuBay, William H., (2006). *Classical Readability Studies*. Costa Mesa: Impact Information.

DuBay, William H., (2007). *Smart Language*. Costa Mesa: Impact Information.

Džeba, Biserka (2006). *Way to go 1 plus, udžbenik engleskog jezika za 4.razred osnovne škole: 4. godina učenja*. Zagreb: Školska knjiga.

Džeba, Biserka (2008). *Way to go 5 plus, udžbenik engleskog jezika za 8.razred osnovne škole, 8.godina učenja*. Zagreb: Školska knjiga.

Fry, Edward (2002). Readability versus leveling. *The Reading Teacher* 56, 3: 286-291.

Fuchs, Lynn, Douglas Fuchs, Deno Stanley (1983). The Nature of Inaccuracy among Readability formulas. University of Minnesota: Institute for Research on Learning Disabilities, 1-29.

Fulcher, Glenn (1997). Text difficulty and accessibility: reading formulae and expert judgment. *System*, 25, 4: 497-513.

Greenfield, Jerry (2004). Readability Formulas for EFL. *Japan Association for Language Teaching* 26: 5-24.

Greenfield, Jerry (2003). The Miyazaki Readability Index. *Comparative culture*9:41-49.

Grey, William and Bernice Leary (1935). What makes a book readable. eds. DuBay, W.H. (2006) *The Classic Readability Studies*, 40-44, Costa Mesa: Impact Information.

Heydari, Pooneh (2012). The Validity of Some Popular Readability Formulas.*Mediterranean Journal of Social Sciences*3: 423-435.

Hutchinson, Tom (2008). *Project 1, udžbenik za engleski jezik, 4. razred, 4. godina učenja*. (3rd edn.). Oxford University Press, Zagreb: Algoritam.

Hutchinson, Tom (2008). *Project 2,udžbenik za engleski jezik, 5. razred, 5. godina učenja*. (3rd edn.). Oxford University Press, Zagreb: Algoritam

Hutchinson, Tom (2008). *Project 3,udžbenik za engleski jezik, 6. razred, 6. godina učenja*. (3rd edn.). Oxford University Press, Zagreb: Algoritam.

Hutchinson, Tom (2008). *Project 4, udžbenik za engleski jezik, 7. razred, 7. godina učenja*. (3rd edn.). Oxford University Press, Zagreb: Algoritam.

Hutchinson, Tom (2009). *Project,udžbenik za engleski jezik, 8. razred, 8. godina učenja*. (3rd edn.). Oxford University Press, Zagreb: Algoritam.

Johnson, Keith (1998).Readability. (Aviable at: http://www.timetabler.com/readable.pdf, visited in August 2015)

Kasule, Daniel (2011). Textbook readability and ESL learners.*Reading & Writing*2:1, 63-76.

Lively, Bertha, S. L. Pressey (1923). A method for measuring the 'vocabulary burden' of textbooks. DuBay, William H., eds. (2006).*The Classic Readability Studies.* Costa Mesa: Impact Information, 7-16.

Lorge, Irving (1949). Readability formulae-an evaluation. Dale, Edgar, eds.*Readability-National Conference on Research in English; National Council of Teachers of English*. Champaign: 8-16.

Mardešić, Maja (2013). D*ip in 6, udžbenik engleskog jezika za 6.razred osnovne škole, 6.godina učenja*. Zagreb:Školska knjiga.

O' Donnel, Mary (2009). Finding middle ground in second language reading: pedagogic modifications that increase comprehensibility and vocabulary axquistition while preserving authentic text features. *The Modern Language Journal* 93: 512-533.

Patty, W.W., W.I. Painter (1931).A technique for measuring the vocabulary burden of textbooks.DuBay, W.H. eds. (2006).*The Classic Readability Studies*. Costa Mesa: Impact Information, 32-40.

Scott,Brian.What Are Readability Formulas?. Aviable at: http://www.readabilityformulas.com/articles/what-are-readability-formulas.php (visited in August 2015)

Vogel, Mabel, Carelton Washburne (1928). An objective method of determining grade placement of children's reading material. DuBay, William H., eds. (2006).*The Classic Readability Studies.* Costa Mesa: Impact Information, 18-24.

Yano, Yusukata, Michael Lang, Steven Ross (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning* 44: 189-219.

Zamanian, Mustafa, Pooneh Heydari (2012). Readability of Texts: state of the art. *Theory and Practice in Language Studies*2, 1: 43-53.

Appendix 1

Table 17B: Post hoc Tukey results for the categories words and characters per word across all the textbooks

TB	Compared Tb	Mean diff. – words	Sig. -words	Mean diff. – Ch p W	Sig. – Ch p W
Wtg1+	WTG2+	-19.79	1.00	-.11	1.00
	WTG3+	-118.16	.54	-.44	.11
	WTG4+	-125.06	.44	-.53*	.01
	WTG5+	-168.32*	.04	-.82*	.00
	DIP4	-122.00	.67	-.25	.95
	DIP5	-159.91	.14	-.46	.10
	DIP6	-143.76	.21	-.54*	.01
	DIP7	-177.66*	.03	-.64*	.00
	DIP8	-221.16*	.00	-.85*	.00
	PRO4	-55.00	1.00	-.21	.98
	PRO5	-87.54	.93	-.45	.13
	PRO6	-184.56*	.02	-.33	.53
	PRO7	-200.46*	.00	-.57*	.00
	PRO8	-291.01*	.00	-.52*	.01
WTG2+	WTG3+	-98.37	.70	-.32	.39
	WTG4+	-105.27	.59	-.41	.08
	WTG5+	-148.52	.05	-.71*	.00
	DIP4	-102.20	.81	-.13	1.00
	DIP5	-140.12	.20	-.35	.37
	DIP6	-123.97	.31	-.42	.06
	DIP7	-157.87	.05	-.52*	.00
	DIP8	-201.37*	.00	-.74*	.00
	PRO4	-35.20	1.00	-.10	1.00
	PRO5	-67.75	.98	-.33	.43
	PRO6	-164.77*	.03	-.21	.92
	PRO7	-180.67*	.01	-.45*	.03
	PRO8	-271.22*	.00	-.41	.05
WTG3+	WTG4+	-6.90	1.00	-.09	1.00
	WTG5+	-50.15	.99	-.38	.06
	DIP4	-3.83	1.00	.19	.98
	DIP5	-41.75	1.00	-.02	1.00
	DIP6	-25.60	1.00	-.10	1.00
	DIP7	-59.50	.98	-.20	.93
	DIP8	-103.00	.42	-.41*	.02
	PRO4	63.16	.99	.22	.95
	PRO5	30.62	1.00	-.01	1.00
	PRO6	-66.40	.96	.11	1.00
	PRO7	-82.30	.84	-.13	.99
	PRO8	-172.84*	.00	-.08	1.00
WTG4+	WTG5+	-43.25	.99	-.29	.38
	DIP4	3.06	1.00	.28	.78
	DIP5	-34.85	1.00	.06	1.00
	DIP6	-18.70	1.00	-.01	1.00

	DIP7	-52.60	.99	-.11	1.00
	DIP8	-96.10	.54	-.32	.22
	PRO4	70.06	.98	.31	.62
	PRO5	37.52	1.00	.08	1.00
	PRO6	-59.50	.98	.20	.93
	PRO7	-75.40	.91	-.04	1.00
	PRO8	-165.94*	.00	.00	1.00
WTG5+	DIP4	46.32	1.00	.57*	.00
	DIP5	8.40	1.00	.36	.17
	DIP6	24.55	1.00	.28	.44
	DIP7	-9.34	1.00	.18	.95
	DIP8	-52.84	.98	-.03	1.00
	PRO4	113.32	.53	.60*	.00
	PRO5	80.77	.86	.37	.13
	PRO6	-16.24	1.00	.49*	.00
	PRO7	-32.14	1.00	.25	.63
	PRO8	-122.69	.09	.30	.23
DIP4	DIP5	-37.91	1.00	-.21	.97
	DIP6	-21.76	1.00	-.29	.73
	DIP7	-55.66	.99	-.39	.25
	DIP8	-99.16	.74	-.60*	.00
	PRO4	67.00	.99	.03	1.00
	PRO5	34.45	1.00	-.20	.98
	PRO6	-62.56	.99	-.08	1.00
	PRO7	-78.46	.96	-.32	.58
	PRO8	-169.01*	.03	-.27	.75
DIP5	DIP6	16.15	1.00	-.07	1.00
	DIP7	-17.75	1.00	-.17	.98
	DIP8	-61.25	.98	-.39	.09
	PRO4	104.91	.78	.24	.93
	PRO5	72.37	.97	.01	1.00
	PRO6	-24.65	1.00	.13	.99
	PRO7	-40.55	1.00	-.10	1.00
	PRO8	-131.09	.15	-.06	1.00
DIP6	DIP7	-33.90	1.00	-.10	1.00
	DIP8	-77.40	.84	-.31	.27
	PRO4	88.76	.90	.32	.56
	PRO5	56.22	.99	.09	1.00
	PRO6	-40.80	1.00	.21	.91
	PRO7	-56.70	.99	-.03	1.00
	PRO8	-147.24*	.03	.01	1.00
DIP7	DIP8	-43.50	.99	-.21	.85
	PRO4	122.66	.47	.42	.14
	PRO5	90.12	.81	.19	.97
	PRO6	-6.90	1.00	.31	.39
	PRO7	-22.80	1.00	.07	1.00
	PRO8	-113.34	.26	.11	.99
DIP8	PRO4	166.16*	.04	.63*	.00
	PRO5	133.62	.13	.40	.06
	PRO6	36.60	1.00	.52*	.00
	PRO7	20.70	1.00	.28	.43
	PRO8	-69.84	.87	.33	.12
PRO 4	PRO5	-32.54	1.00	-.23	.95
	PRO6	-129.56	.38	-.11	1.00

	PRO7	-145.46	.20	-.35	.41
	PRO8	-236.01*	.00	-.30	.58
PRO5	PRO6	-97.02	.71	.12	1.00
	PRO7	-112.92	.47	-.12	1.00
	PRO8	-203.47*	.00	-.07	1.00
PRO6	PRO7	-15.90	1.00	-.24	.79
	PRO8	-106.44	.36	-.19	.92
PRO7	PRO8	-90.54	.64	.04	1.00

Appendix 2

Table 20B: Post hoc Tukey results for YOL and words and characters per word

YOL	**Compared YOL**	**Mean diff. – words**	**Sig. -words**	**Mean diff. – Ch p W**	**Sig. – Ch p W**
4	5	-30.08	.88	-.18	.23
	6	-89.83*	.03	-.28*	.01
	7	-108.73*	.00	-.42*	.00
	8	-167.83*	.00	-.57*	.00
5	6	-59.75	.22	-.09	.74
	7	-78.65*	.04	-.23*	.02
	8	-137.75*	.00	-.39*	.00
6	7	-18.90	.95	-.14	.29
	8	-78.00*	.01	-.29*	.00
7	8	-59.10	.13	-.15	.18