

**Západočeská univerzita v Plzni**

**Fakulta pedagogická  
Katedra anglického jazyka**

**Diplomová práce  
HODNOCENÍ ČTIVOSTI VÝUKOVÝCH TEXTŮ PRO  
ANGLICKÝ JAZYK**

**Bc. Eva Mlejnská**

**Plzeň 2019**

**University of West Bohemia**

**Faculty of Education  
Department of English**

**Thesis**

**ASSESSING READABILITY OF ENGLISH  
LANGUAGE TEACHING TEXTS**

**Bc. Eva Mlejnská**

**Plzeň 2019**

Tato stránka bude ve svázané práci Váš původní formulář *Zadání dipl. práce*  
(k vyzvednutí u sekretářky KAN)

Prohlašuji, že jsem práci vypracovala samostatně s použitím uvedené literatury a zdrojů informací.

*V Plzni dne 1. července 2019*

.....

Jméno Příjmení

## ACKNOWLEDGMENTS

I wish to express my gratitude to the supervisor of my thesis, Mgr. Gabriela Klečková, Ph.D., for her guidance. I also wish to give my thanks to all who supported me along the way.

## ABSTRACT

Mlejnská, Eva. University of West Bohemia. June, 2019. Assessing Readability of English Language Teaching Texts. Supervisor: Mgr. Gabriela Klečková, Ph.D.

The subject of this thesis is the assessment of the readability of English language teaching texts. The accurate assessment of texts for reading in the context of teaching English as a foreign language is a necessity and, acknowledging it as such, there arises the need to ascertain the most useful way of conducting it. There exists a number of established readability formulas for this purpose and several online programmes which can calculate them. These programmes and their use are the main focus of the analysis in this thesis, which aimed to thus draw distinction between them and attempt to conclude whether some of them could be recommended over others. For this purpose fifteen texts were analyzed in four chosen tools and the results were critically examined and compared. In consequence, a combination of the tools rather than a single one had to be recommended. For a quick assessment of readability it was recommended to use the tools which allow for the calculation of readability scores — *Text Inspector*, *Coh-Metrix 3.0* and *Readable*. For an in-depth analysis of the vocabulary, tools which can create the English language profile — *Text Inspector* and *Vocab Kitchen* — were recommended.

## TABLE OF CONTENTS

I. INTRODUCTION.....	1
II. THEORETICAL BACKGROUND.....	2
Reading materials in classrooms .....	2
Readability.....	4
Readability assessment factors .....	5
Working with readability formulas.....	7
Comparing native speakers and EFL students.....	8
Teaching reading .....	10
Automatization .....	14
OCR programmes.....	14
Text Inspector.....	14
Coh-Metrix 3.0.....	15
Readable .....	16
Vocab Kitchen .....	17
III. METHODS.....	18
Hypothesis and Principles .....	18
Textual materials .....	18
Software.....	19
Research procedure.....	19
IV. RESULTS AND COMMENTARIES.....	21
Preparation and Initial Inspection.....	21
Project 1.....	22
Project 1: Text 1 .....	22
Project 1: Text 2 .....	23
Project 1: Text 3 .....	24
Project 2.....	25
Project 2: Text 1 .....	25
Project 2: Text 2 .....	26
Project 2: Text 3 .....	27
Project 3.....	28
Project 3: Text 1 .....	28
Project 3: Text 2 .....	29
Project 3: Text 3 .....	30

Project 4.....	31
Project 4: Text 1 .....	31
Project 4: Text 2 .....	32
Project 4: Text 3 .....	33
Project 5.....	34
Project 5: Text 1 .....	34
Project 5: Text 2 .....	35
Project 5: Text 3 .....	35
EVP.....	36
Commentaries.....	37
V. IMPLICATIONS .....	42
Pedagogical Implications.....	42
Limitation of the Research .....	43
Suggestions for Further Research.....	44
VI. CONCLUSION.....	45
REFERENCES .....	47
APPENDICES .....	50
SHRNUŤÍ.....	68



## LIST OF TABLES

Table 1: Conversion of the Flesh Reading Ease score to school grades .....	7
Table 2: Conversion of the Gunning Fog Index to school grades .....	8
Table 3: Conversion of the Flesch Reading Ease to CEFR levels.....	9
Table 4: The Results of the Initial Inspection of the Different Tools .....	22
Table 5: Analysis of Text 1 from Project 1 .....	23
Table 6: Analysis of Text 2 from Project 1 .....	24
Table 7: Analysis of Text 3 from Project 1 .....	25
Table 8: Analysis of Text 1 from Project 2 .....	26
Table 9: Analysis of Text 2 from Project 2 .....	27
Table 10: Analysis of Text 3 from Project 2 .....	28
Table 11: Analysis of Text 1 from Project 3 .....	29
Table 12: Analysis of Text 2 from Project 3 .....	30
Table 13: Analysis of Text 3 from Project 3 .....	31
Table 14: Analysis of Text 1 from Project 4 .....	32
Table 15: Analysis of Text 2 from Project 4 .....	33
Table 16: Analysis of Text 3 from Project 4 .....	33
Table 17: Analysis of Text 1 from Project 5 .....	34
Table 18: Analysis of Text 2 from Project 5 .....	35
Table 19: Analysis of Text 3 from Project 5 .....	36
Table 20: Recurring mistakes that appear in the tools .....	38

## **LIST OF FIGURES**

Figure 1: The Coh-Metrix L2 Readability in relation to the CEFR levels .....	16
--	----

## I. INTRODUCTION

When teaching English as a foreign language, it is vital that the teachers know and choose an appropriate level of textual material to be provided to the students. Too low a level and the students are left with no room to grow; too high a level and the students' advancement is necessarily hindered by a lack of understanding and, what is more, their growth perspective will possibly decrease due to a lack of motivation from pervading failures. It is believed that the ideal situation finds the student in what is called the *zone of proximal development*, described as the "distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). What this means in practice is that the student is being challenged, but able to overcome the challenge - albeit with some help.

The prevention of issues connected to a faulty estimate of reading difficulty of teaching texts can be, to an extent, achieved by their previous assessment. While not disparaging the value of teachers' personal assessment that draws from their skills and experiences, there exists an option to evaluate textual material based on readability formulas, which can be a valuable asset in the preparation for reading lessons, as it provides a ground for increased objectivity. There are numerous online tools that promise an easy assessment of texts based on the aforementioned formulas as well as other relevant scores. These tools can often be used for free and are easily accessible, which makes them logical helpers in lesson planning.

This is where this thesis aims to contribute with its findings and offer an overview of different approaches and software options that could be used by teachers wanting to ascertain an appropriate level of a chosen text or merely assure themselves of it. The thesis explores the use of reading texts in the classrooms, provides a theoretical background for the readability assessment and investigates the possibilities of its use in the context of teaching English as a foreign language. The research in this thesis considers different online assessment tools and investigates the functions they offer and the reliability of their results, as well as the practicality of their use in order to ascertain whether a potential superiority of one of the tools can be proclaimed.

## II. THEORETICAL BACKGROUND

This chapter contains an overview of the relevant knowledge that is available regarding the assessment of the difficulty of English language teaching texts. Firstly, reading materials, their kinds and use are mentioned. After that, there is the definition of readability. The factors which influence readability are then discussed, along with some history of the text readability research and reasons for their direction. Following is the discrepancy regarding the different readability formulas and their mutual transferability. The question of the second language learners and the options for the assessment in the context of the classroom is discussed afterwards. Methodology of teaching reading is discussed thereafter. Lastly, there is an overview of the relevant programmes that can be used in the readability assessment.

### Reading materials in classrooms

Choosing a text for class reading comes with certain necessary decisions. Teachers usually have two possibilities — either use the reading texts that are contained within the textbook that the class works with on a regular basis or, if the teachers want, they can expand the range by texts taken from the real world. These can be newspaper articles, magazines, literary works, advertisements, but also electronic sources, like blogs, forums, online magazines etc. The teachers' decisions may be influenced by various elements inherent in either type of text.

Commercial textbooks come in many levels, which simplifies the teachers' choice, provided they are familiar with the students' proficiency. This can be measured in several ways. To mention the, arguably, most significant one in the European context, the Common European Framework of Reference for Languages (*CEFR*) divides this proficiency into six categories (A1, A2, B1, B2, C1, C2), which correspond with the individual positions of the proficiency scale. Therefore, B1 level students, for example, are equivalent to intermediate level students and the textual material with which these students work would commonly be an intermediate level course book, workbook, student's book etc. To test the students, they can be subjected to several official tests or exams, like the *TOEFL (Test of English as a Foreign Language)* or the *IELTS (International English Language Testing System)*. This is usually not the case in the school environment, however, where the students' level is assessed in other ways — by the means of the individual institutions' own placement tests. Depending on the outcome of such assessment, the students are then assigned to a class of a particular level in which they

should encounter a particular level textual material. In the case of commercially produced ELT literature, this constitutes texts assessed in advance regarding their level of difficulty, which can mean a reassuring and easy choice for teachers. What also speaks in favour of these materials is the fact that they are often supplemented by various pre-reading and post-reading activities, such as discussion questions, scanning or skimming exercises, vocabulary exercises etc. These ready-made activities offer to the teachers a reduction of preparation time, possibly as well as a different perspective. Be it for whatever reason, they are a part of the teaching process and should not be ignored. Moderation in their use might, however, be beneficent to the learners, as other materials should be included as well.

Authentic material is, according to Harmer (2007, p. 273), one where “no concessions are made to foreign speakers.” It is a “normal, natural language used by native or competent speakers of a language” (Harmer, 2007, p. 273). His preceding commentary of it clearly states his belief in the significance of the use of such material in classroom, as it serves a purpose in later lives of the learners. Based on his reasoning it could be argued that exposure to authentic reading materials prepare the learners better for real-life situation in which they will have to work with similar materials. Whether this exposure could also mean an improvement in overall reading skill is answered by Beresova (2015, p. 199), who observes the difference between learners exposed to authentic texts and texts whose purpose is language instruction and concludes that authentic material leads to improved reading skills. Consequently, a teacher should, knowing this, aim for an inclusion of at least a portion of reading texts from this category, providing that his or her role is to prepare the learners for the real world and not merely an artificial, school-made exam. Budiono (2010) further observes that literary texts in particular also support the students’ motivation through their enjoyment. Literary texts, which are enjoyed, supposedly increase reading comprehension owing to a subconscious development of reading habits. It must be said that this could insinuate he meant reading texts for extensive reading, which is done usually outside classrooms, since it requires a longer period of time than is reasonable to offer during classes, but even then it is another reason to not exclude these kinds of texts.

In order to offer the learners the opportunity to experience authenticity and not demotivate them simultaneously, a problem of which Harmer (2007, p. 273) speaks in relation with a careless approach in choosing authentic reading material, there arises the need to properly analyse the potential readings. Textbooks provide, as already mentioned,

an answer for this directly on their covers — in the form of a code corresponding with a specific learner level. Leaving the matter of accuracy of such a code and granting the commercially produced material reliability in assessing its own difficulty, there still remains the issue of how to successfully measure the difficulty of texts from various other sources, which should constitute a substantial part of the teaching material. Even though many teachers could rely on their own intuition and knowledge of their students, a more professional and analytical approach could be a good way to verify the accuracy of the teachers' estimate.

### **Readability**

There is no progressing without clarifying certain terminology. There are two main expressions connected to the way a reader can perceive text on the hypothetical difficulty scale: readability and legibility. *The Oxford Advanced Learner's Dictionary* (online) lists the two as synonyms, but upon the examination of their individual entries a distinction can be found. Whereas legibility is “the quality of being clear enough to read” (Legibility, n. d.), readability is described as “the fact of being clear and easy to read” — a meaning feasibly synonymous — but also “the fact of being easy, interesting and enjoyable to read” (Readability, n.d.). From this an assertion could be made that both words are connected to the physical quality of the text, such as the font, the clarity of the print or the neatness of the writing, but the meaning of readability can also be extended to the readers' ability to comprehend and become absorbed in the reading. It is this latter meaning of readability that this thesis means to explore.

The field of applied linguistics contains within itself an area of study called *text difficulty* or *text readability* (Fulcher, 1997). Staying true to the aforementioned dictionary definition, these two should not be taken as interchangeable, since *readability* might be indicative of the reader's interest as well as ability, but here they seem to represent identical propositions. For the sake of clarity and since the purpose of the thesis is not to draw distinction between the two terms, readability and difficulty will both be used to describe the ability of a learner to process a text.

As Fulcher observes, this area is a much neglected one. There can be no question of the usefulness of studies pertaining to the assessment of text readability, particularly in pedagogy, so the reason should lie elsewhere. “Rating a text's difficulty is not an exact science”, claims Gilmore (2007, p. 51), touching upon one of the possible problems. Fulcher (1997, p. 498) is more specific on the matter of exactness, putting it down to the

factor of the reader, who presents such variability that any assessment — what Fulcher calls *prediction* — remains within the scope of mere estimates.

### **Readability assessment factors**

As insinuated, the assessment factors can be thought of as being divided into two sorts: the reader related factors and the text related factors. Reader factors would include things such as his or her first language (L1), vocabulary size, motivation, current state of mind etc. Their connecting attribute being that they originate from within the reader. In contrast, textual attributes come from within the piece of reading itself and are in no way influenced by the reader. They are, to mention some, the length of the text, the choice of vocabulary, sentence length, topic etc. It presents itself as without question that the effort needed to process a text is a combination of both the aforementioned factors. Indeed, Fulcher (1997, p. 498) brings into focus the fact that whether or not the reader is interested in the topic of the writing has an even bigger impact on the perceived difficulty than even sentence length.

There have been attempts to narrow down the assessment factors since very early on in the field of this research. Understandably so, since they create the ground for further research and without pinpointing the ones that can create a solid foundation there is no proceeding and every attempt to evaluate the difficulty of a reading text will, as criticised by Fulcher, have the nature of a guesswork. Some of the earlier attempts to describe the difficulty factors were by W. S. Gray (1935), who saw the importance of an exact approach and sought to discover the ones that he would be able to use in predicting the reading difficulty of various texts for adult readers. He made two assumptions (p. 94): “there are elements inherent in reading materials which are significant indicators of difficulty” and “the identification of these elements is an essential step in developing a technique which will help to solve the problem”. His approach was to dismiss the reader factors altogether and focus solely on what he calls “elements of difficulty inherent in reading materials” (p. 94) — structural elements in particular. While not denying influences of other factors (p.94), he finally settles on four out of the forty-four that he originally took into account (1935, pp. 100 – 117). These four were: percentage of monosyllables, percentage of different words, average length of sentences in syllables, percentage of simple sentences (p. 150). Gray looked for factors that would be quantifiable in an objective way (1935, p. 100). He devised reading tests for his subjects and studied which elements have the greatest influence on “poor readers” and which have an influence on “good readers”. The

results suggest that elements of vocabulary influence the former group more with regards to difficulty, whereas structural elements influence the latter (p. 120). As previously mentioned, the quality of being able to be objectively measured was essential for Gray (p. 106), which is understandable, since otherwise precise numbers would be lacking, and the approach would deviate from what constitutes a proper research. In the end he opted for the factors that he perceived as reliable with regards to accurately correlating with the reading results, familiarity and easy recognition and, lastly, convenience of use in the assessment (p. 130–131). The resulting formula consisted of the different elements added together or subtracted — depending on whether they influence the difficulty in a positive or negative way, each multiplied by a coefficient conveying the weighed value (pp. 197–199).

As time went, different factors were used in different ways in attempts to create a reliable formula for the assessment. However, one thing was in common amongst all of them — they used only the factors relating to the text itself. To illustrate by mentioning several of the, arguably, more significant: the Kincaid-Flesch formula uses the sentence length, number of syllables and number of words (Kincaid, Fishburne, Rogers, & Chissom, 1975, pp. 39–40); the Fog index is calculated using words per sentence, number of words and a number of words of three or more syllables, the SMOG formula depends on the same except the number of words per sentence (Wang, Miller, Schmitt, & Wen, 2013, p. 506).

Dismissing the reader factors is seen as a shortcoming of most readability formulas by Danielson, who would consider them “above and beyond the formulas” (1987, p. 185). This concern of hers is logical. The problem is that she does not put forward any solution for it. The reason for this is evident, as one cannot simply calculate using the infinite number of potential readers in all their potential states of mind. Certainly not if one strives to create a universal formula that would have the biggest possible range of application — which is, no doubt, the case for most scholars in this field. The only reader that remains, the only one who can be taken into consideration in this situation, is an “idealized ‘average’ reader”, who is mentioned by Fulcher (1997, p. 498). This reader presents a model for a given age, as proposed by Fulcher (1997, p. 498), but age can effectively be easily substituted for a level of language mastery. This level may be expressed by the already mentioned CEFR proficiency categories or any other descriptors, such as school grade or even individual difficulty scales.



### Working with readability formulas

There is an issue that comes with the previously insinuated variedness by which it is possible to express the difficulty scale — they do not always translate mutually. Zamanian and Heydari (2012) describe what they found to be some of the most frequently used formulas. Their overview illustrates this very well. The Flesch Reading Ease formula uses a scale of 0 to 100, where the highest possible reading difficulty is expressed by the value 0 and the lowest by the value 100. The Dale-Chall formula expresses the same by numbers below 4,9 to 10 and above, where the lower the value, the lower the reading difficulty. The Gunning's Fog index also attributes the lower end of the scale to lower reading difficulty, but the numbers by which this scale is expressed range from 6 to 17. Fry Readability Graph describes difficulty on the scale of 1 to 12, with 1 being the least difficult and 12 the most difficult reading text (Zamanian and Heydari, 2012, pp. 44–46).

This issue is partially remedied by the fact that most formulas — and all of the above-mentioned ones — include a conversion to the school grade that approximately matches the given level of reading difficulty. This means that a textual material of a given difficulty level should ideally be well understood by a standard student of a particular grade and beyond. The range, however, differs amongst the different formulas, and where some reach as far as the first grade — namely the Fry Readability Graph — some start only at sixth grade (Zamanian and Heydari, 2012, pp. 44–46). The Flesch-Kincaid Grade level value should correspond with the school grade of the reader (Kincaid et al., 1975). The conversion of the Flesch Reading Ease is displayed in *Table 1* and the conversion of the Gunning Fog Index in *Table 2*.

*Table 1:* Conversion of the Flesch Reading Ease score to school grades

Reading Ease Score	Estimated Reading Grade
90 to 100	5th grade
80 to 90	6th grade
70 to 80	7th grade
60 to 70	8th and 9th grade
50 to 60	10th to 12th grade (high school)
30 to 50	13th to 16th grade (college)
0 to 30	college graduate

Note: Data collected from Flesch and Gould (1949)

Table 2: Conversion of the Gunning Fog Index to school grades

Fog Index	Reading Level by Grade
17	College graduate
16	College senior
15	College junior
14	College sophomore
13	College freshman
12	High-school senior
11	High-school junior
10	High-school sophomore
9	High-school freshman
8	Eighth grade
7	Seventh grade
6	Sixth grade

Note: Data collected from Gunning (1968, p. 40)

This situation is not ideal, but it means that it is possible to at least roughly compare the results of different formulas applied to the same text. The results are, however, unlikely to match every time and without exceptions, which is due to different factors being taken into consideration and different ways in which these are implemented in the various formulas.

### Comparing native speakers and EFL students

As Zamanian and Heydari (2012, p.49) mention in their findings, most of what was researched in the field of readability was aimed at native speakers of English. The Kincaid-Flesch formula, for example, was first developed in order to better assess the ability of military personnel to comprehend the technical reading material presented to them (Kincaid et al., 1975, p. 1). Wang et al. (2013) studies their use in the context of health care. Flesch and Gould (1949, pp. xi–xii) originally intended for the readability formula to be used by the authors of literary works. Gray (1935, p. 59) focused on adult readers only. Wang et al. (2013, p. 504) further mentions the use of readability formulas by teachers. He does not specify this, but it might be reasonable to assume that he means teachers in native speaker environments, since it has been established that that is the area in which these formulas are used the most.

This can mean that the results, if used on calculating readability for a learner of English as a foreign language (EFL) can be somewhat dubious. This concern is put forward, among others, by Zamanian and Heydari (2012, p.49), who call into question the

validity of such result in and EFL context. There remains the question of the extent to which these formulas can be used to evaluate a reading text for such students. The answer that intuitively offers itself is that since the more difficult texts will still be more difficult than the easy ones, the relative difficulty could remain unchanged and the only problem — and not an unsolvable one by any means — would be to ascertain the exact position that the EFL students hold relative to the native speakers. Undoubtedly, this position would be of much older students further down the less proficient end of the scale. There have, naturally, been attempts at converting the traditional formulas to CEFR levels, such as can be seen in Table 3.

*Table 3: Conversion of the Flesch Reading Ease to CEFR levels*

<b>Flesch-Kincaid readability scores</b>	<b>CEFR LEVELS</b>
0–50	C2
50–60	C1
60–70	B2
70–80	B1
80–90	A2
90–100	A1

Note: Data collected from Linguapress (n.d.)

It would be prudent to describe the two terms here, in order to clarify. EFL speakers are those users of English who do not use this language in their country, not even for official administrative purposes (Jenkins, 2015, pp. 2–5). Jenkins (2015, p.5) claims that in the context of the expanding use of English, this term is falling out of favour, since it does not take into account the fact that English has increasingly been used as a Lingua Franca (ELF) or as an International Language (EIL). According to her, the term EFL better expresses only those users who intend to communicate with native English speakers. For the purposes of this thesis, the terms are interchangeable, as they serve to merely make a distinction between those users, who acquire the language on purpose at a later stage in their lives and those, who use it as their mother-tongue. That is why EFL will be used in this thesis exclusively. Mc Kay and Brown (2016, pp. xiiv–xiv) discuss the issue of native speakers and the definition of this label, which is rather difficult to reach, as there might be many points of view from which to reach it — such as whether the speaker uses the language with a certain degree of competence, whether he or she was raised in an English speaking environment or whether dialect belongs to one of the dominant varieties. This

distinction, again, is inconsequential when it comes to this thesis, and the native speaker will be used to describe a person who uses English as a primary means of communication was raised in an English-speaking environment and thus possesses a corresponding level of competence.

Being that both the aforementioned groups have undergone a different process of language acquisition and language learning, their shift on the proficiency scale, too, is likely very different. That, for example, fourth grade native speakers' English proficiency is higher than an EFL learners' English proficiency is doubtless, but the advancement they achieve in a year's time is something that could be subject to inquiry. The situation that presents itself as the most logical is that native speakers, who are, without a doubt, subject to more extensive exposure and practice, advance at a much faster pace than EFL students, whose exposure and practice is limited to just several school sessions. To put this into perspective, in the context of the Czech Republic, this time consists of minimum three hours a week (Metodický portál, n.d.a). What this leads to is that whatever advancement the native speakers make, the EFL students should make it much slower. Therefore, a positive shift in the difficulty level of a reading text that is generally understood by native speakers in a year's time should be attained in a considerably longer period of time by EFL students.

The proficiency of EFL students also likely does not have the potential to reach the level of native speakers, which would result in a smaller range on the difficulty scale when it comes to understandable, therefore appropriate, reading materials. There is only a small portion of learners who achieve a proficiency that rivals that of the native speakers (Ellis, 2009, p. 14). Owing to this, the reading texts that can be worked with in an EFL learning environment will not surpass a certain level of difficulty. In the context of lower secondary education in the Czech Republic, the minimum level of foreign language proficiency upon its completion is determined to be A2 (Metodický portál, n.d.b). Even though the average proficiency is likely to be higher than the minimum, it is very clear that the difficulty levels that lower secondary English teachers will aim for are low.

### **Teaching reading**

Even though the right choice of text is arguably amongst the most crucial aspects of teaching reading, it still comes with certain specifics that go far beyond a mere choice of reading text. When teachers want to execute a reading lesson correctly, it is vital that they

understand these things and consider them in their lesson planning. Some of the key issues will be considered in this chapter.

Reading is adaptive when it comes to time, so the readers can pace themselves according to their needs. This could be a good thing, but it could also present some challenges to the readers. Scrivener (2005, pp. 184–185) mentions the fact that readers, who do not, presumably, worry about time constraints, will take an unnecessarily long time to read a text in an effort to understand every word of it. Richards (2015, p. 449) presents this problem as something that second language learners with limited proficiency face. It is not a very efficient way to process a piece of writing and, as Scrivener (2005, pp. 184–185) says, it is one of the main problems that readers face. With it, he mentions, comes a lack of vocabulary, which slows the readers down and, as a result, their motivation to continue reading decreases. Another problem he puts forward is that readers may understand the words they read, but the general meaning of the text remains elusive for them.

These two areas can be explained by two theories, covered by Scrivener (2005, p. 178), both of which address the processing of the input when learning a receptive language skill. The first theory is called ‘bottom-up’ and it is a strategy which has the readers decipher the individual words and from them piece together the meaning of the whole message. The second theory, ‘top down’, is the opposite, which means that the readers first get the gist of what a piece of writing means even if it means not understanding everything that is written down. Scrivener (2005, p. 178) puts forward the notion that the former is not very reasonable, and it is clear to see why. Second language learners, who, in many cases, struggle to understand the language, exert too much energy on translating words and their individual meanings, understanding morphological and syntactic constructions and, in the process, neglect to focus on the general meaning. When approached the other way round, the students get familiar with the message of the text and later, when they re-read it closely, the message is already in their memory and they are more free to focus on the more detailed aspect of the text, such as difficult vocabulary or grammar. This is why Scrivener (2005, p. 187) suggests that the sequencing of a reading activity should go from gist-oriented tasks, where students work with the overall meaning, such as make predictions and then check if they prove to be true, or work with the title of the text, or put items into correct order etc. to a more detailed work with focus on isolated parts of the text. Richards (2015, p. 451) suggests that these two processes combine in what is called ‘interactive processing’ and that they function together, as reading is a process of constant

interaction between the readers and the text, but it is something that can only occur at higher levels of proficiency. This is useful to realize, since it shows what a complex activity understanding a text is.

Before the start of the lesson teachers need to realize what kind of purpose the reading will serve and according to it choose the appropriate text as well as strategies. There are two kinds of reading in the learning context that can be put into contrast to each other — intensive and extensive. Scrivener (2005, p. 188) comments on these while putting them into perspective regarding classroom scenarios and real-life situations. Intensive reading, he says, is inherent to ELT classes, as it involves reading shorter expanses of text more than once, often with different focus each time. Extensive, on the other hand, he claims to be more natural, as it means reading in a more leisurely manner and rarely coming back to what has already been read. He further claims that: “there is a great deal of evidence that extensive reading has a powerful impact on language learning” (Scrivener, 2005, p. 188). This may seem to imply that an intensive reading is the inferior of the two and that it should not have a place in the classroom, but it is not so, since it serves its purpose as well.

On the topic of determining goals, Richards (2015, p. 457) pinpoints how many different kinds of reading classes there are and how the kind of reading and the choice of text depend heavily on the needs of the students. It might, therefore, be a good idea for the teachers to first realize what it is that they want the students to gain at the end of the lesson. Richards (2015, p. 457) divides the reading purposes into general improvement in reading and specific, which include preparation for different exams.

Intensive reading is often used to introduce a particular grammar rule or to furnish the students with a specific, topic related vocabulary. This is useful for lower level students, as it does not burden them with expansive reading texts. Language items in these texts are usually densely distributed and serve the purpose of illustrating their use in discourse. This does not mean that there cannot be specific items that the students focus on in extensive reading as well, but it is perhaps more reasonable to include it in work with a higher proficiency readers in order to broaden their vocabulary, grammar awareness or awareness of specific language structures. It could be argued that an intensive reading that happens in classes does not necessarily always happen for the sake of reading itself, but rather in order to introduce and teach relevant language items. Even then this kind of

activity should be approached with the same attitude as any kind of reading, as it creates an opportunity for practice and could be useful for the students later.

Since the lack of vocabulary knowledge hinders the students in their advancement, it is always a good idea to prevent this by helping them understand in advance and thus dispensing with an excessive amount of delaying moments. Richards (2015, p. 459) claims that pre-teaching vocabulary prior to the reading task itself will reduce the difficulty level of the text. It might be prudent to point out that the readability will remain the same, as that is something that does not change without changing the text, but the subjective ease with which the students read will undoubtedly increase.

In the phase that precedes reading it is helpful to also include other than just vocabulary knowledge. Richards (2015, p. 459) mentions providing some background knowledge on the topic and activating schemas, which gives the students a purpose to read. He claims that the pre-reading sessions are the most important for the students, because they are a source of motivation and prepare the students for the reading itself by giving them strategies and prior knowledge.

As the students read the text, it is sometimes useful that they know what to focus on, especially in lower levels of proficiency. Scrivener (2005, p. 171) describes a scenario regarding listening activities that illustrates a problem, which can be easily applied to reading as well. In this scenario the students are given an input, which they should focus on and comprehend, and which is immediately followed by a set of questions. This is problematic, because, as Scrivener (2005, p. 171) points out, the fact that the students can or cannot answer these questions may not be indicative of whether they processed the input correctly, as they were not aware of what parts of it they should have considered important. He goes on to add that this is not what happens in real life, from which it is easy to surmise that this strategy does not support real-life skills and is therefore not a meaningful and useful strategy. Scrivener (2005, p. 186) suggests adapting a reading strategy according to the type of text and how it would naturally be read were it not a part of a language class. Depending on this, various kinds of reading tasks, such as skimming or scanning can be implemented. Scrivener (2005, p. 185) describes skimming as quick reading that has the reader process the overall gist of a passage and scanning as looking for a specific information in the text. Whatever the task is, it is vital that the students know what they are supposed to do.

After the reading there comes a stage that is arguably as important as the preceding ones. In it, students should review the reading strategies as well as the reading itself (Richards, 2015, p. 462). Activities that Scrivener (2005, p. 187) suggests for this stage are debate, role-play or writing tasks.

### **Automatization**

Given the complexity of some formulas and the length of the potential text a question of the worth of a potential readability analysis naturally offers itself. An answer for this question undoubtedly lies in software options that are available to help calculate readability. There have been numerous attempts to implement already existing readability formulas or create new ones in a functional tool that would quickly analyse a piece of writing and inform the user about its readability. Some of the software options are described in this chapter, albeit not in excessive detail, as this serves as an introduction for each of them. This shall be remedied in the later chapters if need be.

### **OCR programmes**

Automatic assessment of electronic text formats is an easy task, as the properties of such texts allow for an easy transfer to the assessment tool, but there are instances where this is not possible. When it comes to printed material or electronic materials of other than text formats, this presents a problem, as these data cannot be easily copied and analysed. For these purposes there is the optical character recognition (OCR) system. OCR systems first pre-process the image, which means adjusting the contrast and removing noise and background, then segment the image into pages, lines, characters etc., extract features that identify the segments, identify the characters and finally, in the post-processing stage, detect misspellings and non-grammatical utterances and adjust the output accordingly (Vijayarani and Sakila, 2015, pp. 20–21). The performance of OCR tools has been found to be good for the purposes of converting characters (Vijayarani and Sakila, 2015, pp. 28–29). The accuracy of OCR software is of varying degrees depending on the type used. Commercially produced versions generally produce better results than freeware options (Záhorová, 2017, pp. 33–34). Even the best OCR tools are, however, susceptible to errors, which means that neither of them is to be trusted invariably and it is a good idea to always inspect their output.

### **Text Inspector**

Text Inspector is an online tool for text analysis that is supported by the English Profile. The version that is embedded on the English Profile website, available at:



<https://languageresearch.cambridge.org/wordlists/text-inspector> offers one of its functions, the English vocabulary profile (EVP), for free use with texts of up to 500 words. EVP informs the users about the occurrence of specific tokens (words) and the CEFR proficiency level which they belong to. The total count of the token types is also included. Each token is accounted for according to its lowest difficulty value by default, but the use of it can be changed to a higher level of difficulty, if that is the case in the analysed text (Text Inspector, n.d.a).

The official domain of the tool is available at: <https://textinspector.com> and the number of functions it offers is substantially higher. It shows general statistics such as number of sentences, number of tokens, the ratio of tokens per types, the average length of sentences, number of syllables etc. Three readability scores are also included, namely the Flesch Reading Ease, the Flesch-Kincaid Grade and the Gunning Fog index. It measures lexical diversity, the number of elements according to parts of speech, counts spelling errors and more (Text Inspector, n.d.b).

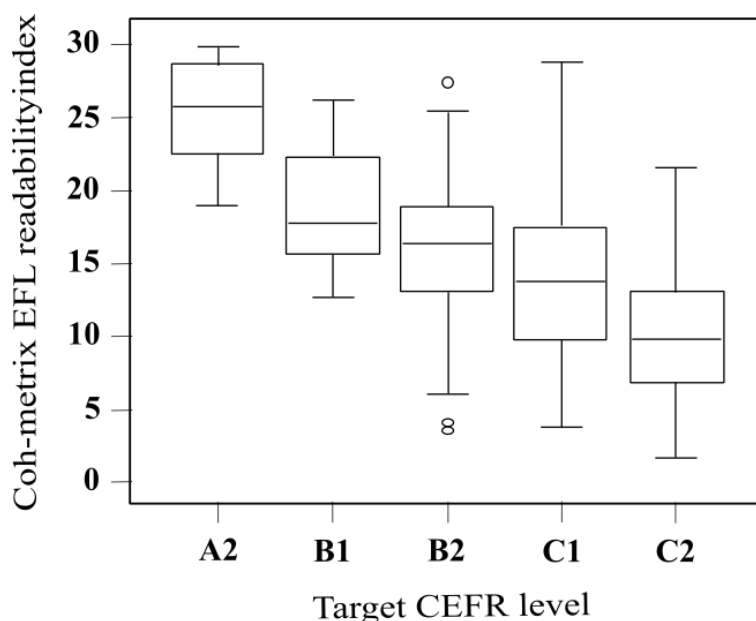
The free version allows for the analysis of texts of up to 250 words, but with free subscription plan it increases to 400 words worth of text. Paid subscription plans offer an increase in word count and add more possible users, as well as unlock certain functions, like the EVP (Text Inspector, n.d.b), which, curiously, is a premium function only, despite being available for free use through the English Profile.

### **Coh-Metrix 3.0**

The Coh-Metrix tool was developed for the purposes of analysing the cohesion and coherence (McNamara, Graesser, McCarthy, & Cai, 2014, pp. 1–2). It is currently available at: <http://www.cohmetrix.com/> and it analyses a wide variety of textual features, precisely 106, all of which are grouped according to the type they belong to. These groups include the basic information such as word count, sentence length, average word length etc. It also includes lexical diversity and syntactic complexity. The aim was to create a tool that would go beyond the surface characteristics of the text (McNamara et al., 2014, p. 83). There are three readability scores available, which are the Flesch Reading Ease, Flesch-Kincaid Grade level and the Coh-Metrix L2 Readability, which is the tool's own formula for readability prediction that focuses on the readability for second language learners and consists of three factors — the content word overlap, sentence syntactic similarity and word frequency (McNamara et al., 2014, p. 80). The values of this readability index converted to CEFR levels can be seen in *Figure 1*. The formula for its calculation is

supposed to be superior to the classical formulas when it comes to simplified texts meant for the learners of English as a second language (McNamara et al., 2014, p. 81).

Figure 1: The Coh-Metrix L2 Readability in relation to the CEFR levels



Note: Data collected from Green, Trim, and Hawkey (2012, p. 132)

The tool is free to use, but the limit is set to a recommended maximum of 15 000 characters (Coh-Metrix 3.0, n.d.). This maximum is only recommended, however, and the tool will work with larger amounts of text. The processing time for such instances, of course, increases.

### Readable

Readable is described as a tool that helps establish how easy to read a piece of text is and offers tips on how to improve its readability (Readable.com, n.d.), This makes it primarily oriented at producers of text, which is inconsequential in the context of this thesis, as it is still a valid tool for readability assessment. It is available at:

<https://app.readable.com/text/> and it offers a wide variety of analytical features that help establish readability. It contains its own readability rating, which scales from A to E, with A being the most readable. The recommendation on the website is that a text for the general public should not be lower than B (Readable.com, n.d.). It further contains the Flesch-Kincaid Grade level, Flesch Reading Ease and the Gunning Fog index, along with recommendations of what to aim for. It further contains clues about grammar and spelling issues, a count of longer sentences and words, all of which can be highlighted

(Readable.com, n.d.). The tone of the text on the scale of formal to conversational is also included (Readable.com, n.d.).

The free version can be used with an unlimited amount of text. A subscription can be purchased for additional features. These comprise of further readability formula scores, detection of particular word groups that influence the writing style, the composition of text etc. (Readable.com, n.d.).

### **Vocab Kitchen**

The vocabulary profiler that is available on the Vocab Kitchen website at: <http://vocabkitchen.com/> assesses a piece of text from the point of vocabulary usage. There are two lists against which the vocabulary can be checked, the CEFR and the Academic Word List (AWL) (VocabKitchen, n.d.a). The CEFR profiler sorts the individual words according to the proficiency level to which they belong, along with the percentage of the occurrence of such words in the text (VocabKitchen, n.d.b). The use of these words cannot be further adjusted regarding their CEFR proficiency level. The AWL profiler displays the words that belong to the AWL in percentage contained within the text as well as the list of the individual words (VocabKitchen, n.d.c).

The use of this tool is free, and registration is not necessary (VocabKitchen, n.d.a). A limit is not set for the amount of text that can be pasted into the profiler. The original website mentions the use for the purposes of assessment of text difficulty levels for education (VocabKitchen, n.d.a).

### III. METHODS

This chapter includes the description of the research methods as well as the reasons for them. It also includes the aim of the research. The research materials and tools used for the research are also to be found in this chapter.

#### **Hypothesis and Principles**

The hypothesis that was assumed before the research was that by the means of using graded text with different online assessment tools a difference between the tools that would indicate their suitability for the assessment of readability of ELT texts would surface. For this testing of the different online tools the following methods were chosen. Firstly, eleven factors by which readability can be assessed were compiled and the possibility of their assessment via the four tools was asserted. Then fifteen graded texts from the Project Student's books were transferred by the means of an OCR programme and tested in the four online tools (Text Inspector, Coh-Metrix 3.0, Readable, Vocab Kitchen). Each text was assessed based on eleven factors and the results were compared in order to indicate the differences between the four tools. The research for this thesis was conducted with the presumption that the Project students' books were correct in their assessment of their level.

#### **Textual materials**

Fifteen text were used at total. They were taken from the fourth edition Project Student's books by Tom Hutchinson, levels 1 to 5. Their CEFR levels were:

- Project 1: A1 (Hutchinson, 2014a)
- Project 2: upper A1 to A2 (Hutchinson, 2014b)
- Project 3: upper A2 (Hutchinson, 2014c)
- Project 4: upper A2, lower B1 (Hutchinson, 2014d)
- Project 5: upper A2, lower B1 (Hutchinson, 2014e)

Three texts were chosen from each level at random. If there were headings available, the headings were included in the testing, as these could also be a source of incidence of features that might influence readability. The only principle that was followed in the choosing process was the absence of any disruptive features, such as blank spaces, which would necessitate that the text be read in a different fashion from what is usual.

The reasons for choosing the Project Student's books were mainly two. The first one was the fact that they were graded, which means that they could serve as a control group, provided that their self-assessment was reasonably accurate. Second reason was their spread across the lower secondary schools in the Czech Republic, since they are seemingly the popular choice. This fact was also asserted by Štefanová (2011, p.39), who found out that they were the most frequently used. It would appear that the situation has not changed significantly since then.

### **Software**

For the purposes of this thesis only free online software was tested. If the software was available through paid subscription but offered the option to be used with a limited amount of text and other features for free, the limited free version was used and tested. A free online OCR programme, the *Online OCR*, available at: <https://www.onlineocr.net/> (Online OCR, n.d.) was used to render texts from the images of the texts taken from the books. This programme was used only as a means of gaining the textual material in an acceptable format and as such, was not tested relative to other similar software performance-wise. The four programmes that were tested were *Text Inspector*, *Coh-Metrix 3.0*, *Readable* and *Vocab Kitchen*. These were assessed from the point of view of the range of functions they provided, their accuracy and their practicality regarding the use by teachers for ELT purposes.

### **Research procedure**

The research started with finding three random reading texts from each of the five books. The texts were then transformed into a picture format and subsequently to a text format by the means of an OCR programme. They were checked against the original to make sure that their form and content matched and all mistakes that originated during the conversion process were corrected in order for the research to be able to proceed.

The second step in the research was the compiling of different readability assessment factors that could be traced in the texts. The factors were chosen according to the findings from the theoretical background. They were of three types — formulas, quantitative measures and qualitative measures. The tested formulas were the Flesch Reading Ease, the Flesch-Kincaid Grade, the Gunning Fog Index and the Coh-Metrix L2 Readability. The quantitative measures included the number of words and sentences, as length should contribute to the difficulty of the text, but also in order to measure the accuracy with which the tools could assess these basic textual features. Moreover, the

average length of words and sentences was chosen. The qualitative measures were chosen to be the type/token ratio, which measures the incidence of word repetition and therefore the lexical diversity of the texts, and the EVP, which assigns the specific tokens (words) to appropriate CEFR levels.

Finally, the texts were each tested in the four tools and the resultant data were recorded. The data were compared and any discrepancies were traced — first via a critical survey of the texts, meaning that possible areas that could be problematic, such as punctuation or special characters were identified if apparent, and second by the testing of the potentially problematic areas in the tools in a more isolated fashion and, if necessary, repeatedly. All problems and inaccuracies were taken into account during the assessment of the tools, together with any characteristics that might impact the practicality and agreeableness for the user.

## IV. RESULTS AND COMMENTARIES

This chapter contains the results of the research that was conducted. The data is presented in an order that corresponds with the order of the texts as they were tested, which is from the preparation and initial inspection of the tools to the Project 1 texts to the Project 5 texts. The initial parts describe the word and sentence count, as well as the averages that resulted from them, as well as the four chosen formulas (Flesch Reading Ease, Flesch-Kincaid Grade, Gunning Fog Index, Coh-Metrix L2 Readability) in the individual texts. Latter part is dedicated to the EVP in the *Text Inspector* and *Vocab Kitchen* and the very last part contains commentaries relevant to the research. The results are displayed in tables that clearly show the differences between the software tools. The important distinctions are described and, if necessary, explained and further commented on. For the texts that the research worked with, see Appendices A–O.

### **Preparation and Initial Inspection**

It was necessary to first create texts from the images taken from the books. This was achieved using a free OCR programme called *Online OCR*, as was already mentioned. The resulting texts were surprisingly free of mistakes, however, sections of text that were blurry or otherwise distorted contained several mistakes. Likewise, the letter ‘L’ that was lowercase had been in several instances transferred as uppercase ‘I’. Apostrophes sometimes caused interference too. Overall, it was necessary to check each text in order to ascertain that they were assessed in the correct form.

The assessment itself was preceded by an initial inspection of the four tools, in which the maximum amount of analysable text as well as the range of different functions their free version offered was determined. The findings can be seen in *Table 4*. It was asserted that the maximum amounts of text was unlimited for both *Readable* and *Vocab Kitchen*, whereas *Text Inspector* and *Coh-Metrix 3.0* had a maximum amount set. This limit was hard-set for *Text Inspector* and could not be exceeded, but the limit for *Coh-Metrix 3.0* was only recommended and larger amounts of text could be analysed. This recommendation possibly had its origins in an effort to maintain the processing time within reasonable limits. *Table 4* also clearly indicates the amount of functions that the tools offer. The largest amount of functions is offered by the *Text Inspector* and the smallest amount by the *Vocab Kitchen*, which specializes in the EVP. *Coh-Metrix 3.0* is the only tool that calculates the Coh-Metrix L2 Readability.

Table 4: The Results of the Initial Inspection of the Different Tools

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Maximum Text length (for Free)	400 words	15000 characters (recommended)	unlimited	unlimited
Flesch Reading Ease	✓	✓	✓	✗
Flesch-Kincaid Grade	✓	✓	✓	✗
Gunning Fog Index	✓	✗	✓	✗
EVP	✓	✗	✗	✓
Average Sentence Length [words]	✓	✓	✗	✗
Type/token ratio	✓	✓	✗	✗
Coh-Metrix L2 Readability	✗	✓	✗	✗
Average Syllables per Word	✓	✓	✓	✗
Number of Sentences	✓	✓	✓	✗
Number of Words	✓	✓	✓	✓

### Project 1

The texts in this section all came from the Project 1 Student’s book by Tom Hutchinson. This book’s level should be A1 according to the information on the cover (Hutchinson, 2014a). For the individual texts discussed in this section, see Appendices A–C.

#### Project 1: Text 1

The text was a description of a flat, its rooms and the furniture and equipment within. It was a short and simple text at first glance, only with a few contractions. It was a very basic descriptive text. The text can be seen in Appendix A.

The results of the analysis can be seen in *Table 5*. The difference between the sentence count in the result is explained by *Text Inspector* not counting headings as separate sentences. This tool also, together with the *Coh-Metrix 3.0* also counted contractions such as ‘there’s’ as two separate words, which explains the difference in word count. Flesch Reading Ease score was very similar with all the tools, placing the text at close to 100, which translates to roughly 5th grade and which should correspond with the A1 CEFR proficiency. Gunning Fog Index was off the scale and also very similar. Flesch-Kincaid Grade corresponded with the 1st grade and was very close in all the tools as well. The Coh-Metrix Readability was off the scale at approximately 33, which would place the text at less than A2.



Table 5: Analysis of Text 1 from Project 1

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	99,37	98,795	98,2	✘
Flesch-Kincaid Grade	1,50	1,489	1,4	✘
Gunning Fog Index	4,74	✘	4,4	✘
Average Sentence Length [words]	8,48	8,091	✘	✘
Type/token ratio	0,36	0,363	✘	✘
Coh-Metrix L2 Readability	✘	33,267	✘	✘
Average Syllables per Word	1,17	1,180	1,2	✘
Number of Sentences	21	22	22	✘
Number of Words	178	178	164	164

### Project 1: Text 2

The second text was a compilation of five simple description of people who live in Britain but have their roots in other countries. There was no heading. It appeared slightly more difficult than Text 1 due to longer sentences and several occurrences of direct speech. There were also dashes and non-English names present in the text. This text can be seen in Appendix B.

For the results of the analysis see *Table 6*. During the analysis, it transpired that unlike *Text Inspector* and *Readable*, *Coh-Metrix 3.0* registered dashes as punctuation that separates individual sentences and not just individual clauses, therefore the sentence count differed in this regard. *Coh-Metrix 3.0* also registered two apostrophes as words, which is a mistake that occurred in *Vocab Kitchen* with dashes. Flesch Reading ease in *Text Inspector* and *Readable* was very similar, while it was slightly higher in *Coh-Metrix 3.0*, but all the tools placed the difficulty at 7th grade, which translates to B1 level. *Text Inspector* and *Readable* also calculated the Flesch-Kincaid Grade to be equivalent to 5th grade, whereas *Coh-Metrix 3.0* evaluated the text as easier again, placing it at the 4th grade level. A big difference occurred in the Gunning Fog Index. Whereas *Text Inspector* placed the text at 8th grade, *Readable* placed it off the scale, which could count as a difference of more than two grades. The Coh-Metrix L2 Readability score was even slightly higher than in Text 1, meaning that it could be placed into lower than A2 level too.

Table 6: Analysis of Text 2 from Project 1

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	74,74	78,177	74,6	✘
Flesch-Kincaid Grade	5,14	4,252	5,1	✘
Gunning Fog Index	8,02	✘	5,5	✘
Average Sentence Length [words]	9,29	8,733	✘	✘
Type/token ratio	0,44	0,436	✘	✘
Coh-Metrix L2 Readability	✘	34,408	✘	✘
Average Syllables per Word	1,45	1,416	1,5	✘
Number of Sentences	28	30	28	✘
Number of Words	260	262	250	252

### Project 1: Text 3

Text 3 was a narrative — a fable about a fox and a crow. It was considerably longer than the previous texts, but it contained a lot of repetition, which should, in theory, have made it simpler. There was some reported speech and contraction in the text. It also contained a heading. For the actual text, see Appendix C.

The result for this text can be seen in *Table 7*. *Text Inspector* would not allow for the analysis of this text, as it was longer than the free limit, which left it with only the EVP for this tool. The sentence count differed due to the *Coh-Metrix 3.0* separating sentences at dashes, but also seemingly randomly connecting sentences that included reported speech, such as: “*Why aren't you singing today, Mrs Crow? Everyone says that you've got a beautiful voice.' Now everyone knows that the crow can't sing.*” which was counted as two sentences. The rather big difference in the word count was caused by *Coh-Metrix 3.0* counting contractions as two words. Flesch Reading Ease was calculated by the remaining tools at 100 or more, meaning that the text should be A1 level or lower. Flesch-Kincaid Grade did not reach even 1 in both tools and Gunning Fog Index was off the scale. Curiously, the Coh-Metrix L2 Readability was over 24, placing the text at somewhere between A2 and B1, which is much more advanced than the traditional formulas calculated it.

Table 7: Analysis of Text 3 from Project 1

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	✘	100,000	103,4	✘
Flesch-Kincaid Grade	✘	0,757	0,6	✘
Gunning Fog Index	✘	✘	3,9	✘
Average Sentence Length [words]	✘	7,907	✘	✘
Type/token ratio	✘	0,271	✘	✘
Coh-Metrix L2 Readability	✘	24,182	✘	✘
Average Syllables per Word	✘	1,124	1,1	✘
Number of Sentences	✘	54	57	✘
Number of Words	✘	427	397	401

## Project 2

This section contains texts from Project 2 Student’s book by Tom Hutchinson. The book should be around the A1 to A2 CEFR level (Hutchinson, 2014b). The three texts that are discussed in this chapter are seen in Appendices D–F.

### Project 2: Text 1

This text was a description of the narrator’s routines. There was no heading or reported speech, but it contained contractions. The vocabulary and syntax were quite simple, but there were certain special items, such as ‘8A’ or ‘PE’. For this text, see Appendix D.

Table 8 displays the results of the analysis of this text. The difference in sentence count is explained by *Readable* mistakenly counting: “*I’m in class 8A. My favourite subjects are Maths, History and PE.*” as one single sentence. The differences in the word count can be explained by *Text Inspector* and *Coh-Metrix 3.0* dividing contractions. Flesch Reading Ease was very similar in all the tools, placing the level of the text at A1, or the 5th grade. The Flesch-Kincaid Grade placed it at 2nd grade in all the tools and the Gunning Fog Index were off the scale, again displaying similar results. Coh-Metrix L2 Readability of the text was close to 27, which should make it an easier A2 level in CEFR.

Table 8: Analysis of Text 1 from Project 2

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	93,74	93,792	92,8	✘
Flesch-Kincaid Grade	2,37	2,376	2,5	✘
Gunning Fog Index	5,04	✘	5,1	✘
Average Sentence Length [words]	8,81	8,852	✘	✘
Type/token ratio	0,51	0,510	✘	✘
Coh-Metrix L2 Readability	✘	26,965	✘	✘
Average Syllables per Word	1,23	1,230	1,2	✘
Number of Sentences	27	27	26	✘
Number of Words	238	239	226	226

## Project 2: Text 2

Text 2 was a description of festivals in Britain. The text comprised of longer sentences than the previous one and it made it appear as though the texts would get progressively more difficult. There was a heading in the text and brackets appeared in it for the first time. There were apostrophes in it, too. The text can be seen in Appendix E.

The results of the analysis are shown in *Table 9*. Apart from the already established differences between the tools it was now found out that *Vocab Kitchen* registered numbers as words. Curiously, so did *Text Inspector*, despite claiming otherwise. Brackets did not cause any interference in the results. Flesch Reading ease as calculated by all tools indicated the level was somewhere between 8th and 9th grade, which is the equivalent of B2 level. Gunning Fog Index results showed the same, with the lower value calculated by *Readable*. The Flesch-Kincaid Grade was, also in agreement amongst the tools, between 6th and 7th grade. The Coh-Metrix Readability evaluated the text as more difficult, roughly at the border of A2 and B1.

Table 9: Analysis of Text 2 from Project 2

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	68,98	70,783	68,0	✘
Flesch-Kincaid Grade	6,96	6,538	6,9	✘
Gunning Fog Index	9,17	✘	8,5	✘
Average Sentence Length [words]	13,39	12,684	✘	✘
Type/token ratio	0,56	0,562	✘	✘
Coh-Metrix L2 Readability	✘	22,494	✘	✘
Average Syllables per Word	1,47	1,456	1,5	✘
Number of Sentences	18	19	19	✘
Number of Words	241	241	236	239

### Project 2: Text 3

Third text was a narrative and could be also described as a fable. There was some repetition in the syntactic patterns as well as in the vocabulary. Reported speech and contractions were quite frequent and one sentence contained a dash. The language appeared somewhat more complex when compared to the Project 1 texts at first glance. For this text, see Appendix F.

Table 10 shows the result of the analysis of this text. The difference in word and sentence count can be explained by the previously described issues, however, some new issues transpired in this analysis. *Vocab Kitchen* counted dashes as words and *Coh-Metrix 3.0* counted some apostrophes as words — but not all. *Readable* divided the sentence: “*To London?’ says Foxy Loxy.*” into two. It did not divide the sentence with the dash despite doing so previously. *Text Inspector* counted the sentences: “*Follow me.’ So Chicken Licken, Henny Penny, Ducky Lucky and Goosey Loosey follow Foxy Loxy.*” and so did *Coh-Metrix 3.0*, which also connected the sentences: “*Do you want to join us?’*” and: “*Yes, we must come with you,’ say Ducky Lucky and Goosey Loosey.*” and three other ones, which displayed the similarity of containing two apostrophes next to each other. It could therefore be assumed that this was the reason for the mistake. Flesch Reading Ease was calculated at the equivalent of 6th grade by all the tools. This should place the text at the A2 CEFR level. Flesch-Kincaid Grade was pointing towards the 3rd grade in *Text Inspector* and *Readable*, but in *Coh-Metrix 3.0* this was the 4th grade. Gunning Fog index was, again, off the scale. Coh-Metrix L2 Readability score corresponds with the B1 level.

Table 10: Analysis of Text 3 from Project 2

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	84,18	83,659	86,2	✘
Flesch-Kincaid Grade	3,76	4,075	3,3	✘
Gunning Fog Index	3,73	✘	3,5	✘
Average Sentence Length [words]	9,03	10,000	✘	✘
Type/token ratio	0,30	0,306	✘	✘
Coh-Metrix L2 Readability	✘	20,664	✘	✘
Average Syllables per Word	1,34	1,336	1,3	✘
Number of Sentences	36	33	39	✘
Number of Words	325	330	316	317

### Project 3

The following section contains the analysis of three texts from Project 3 by Tom Hutchinson. The CEFR level is marked at upper A2 (Hutchinson, 2014c). For the discussed texts, see Appendices G–I.

#### Project 3: Text 1

The first text was a story about a boy from New Zealand who moved to Great Britain. It contained a heading and some contractions. The sentences were rather short and the whole text seemed easy to comprehend. This text can be seen in Appendix G.

The research results are seen in *Table 11*. There were not many differences in the word and sentence count amongst the tools and all of them could be explained by the previously discussed issues, such as headings and contractions. Flesch Reading Ease was equivalent to 5th grade everywhere, therefore it would belong to the A1 level. Gunning Fog Index was off the scale, evaluating the text as easier than scale measures. Flesch-Kincaid Grade level was very low too, as it corresponded with the 1st grade. Coh-Metrix L2 Readability was lower than A2, therefore it could be assumed that the result is A1.

Table 11: Analysis of Text 1 from Project 3

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	98,84	100,000	98,5	✘
Flesch-Kincaid Grade	1,79	1,525	1,6	✘
Gunning Fog Index	4,04	✘	3,7	✘
Average Sentence Length [words]	9,32	9,000	✘	✘
Type/token ratio	0,43	0,417	✘	✘
Coh-Metrix L2 Readability	✘	33,336	✘	✘
Average Syllables per Word	1,16	1,153	1,2	✘
Number of Sentences	28	29	29	✘
Number of Words	261	261	247	247

### Project 3: Text 2

Second text from Project 3 was a description of the human history. It was very dense regarding the information it contained and there were some potentially difficult words. It also had a heading and contained many large numbers. Dashes appear in the text as well. The text can be seen in Appendix H.

Table 12 displays the results of the analysis of this text. The discrepancies in the word and sentence count can be, again, explained by the previously mentioned issues. Newly discovered issue transpired in *Text Inspector*, which not only counted numbers as words, but if there was comma in the numbers — such as was the case with numbers above one thousand — it divided the numbers at the comma and counted them as more than one word. Flesch Reading Ease was between 60 and 70 in all cases, which is the equivalent of 8th and 9th grade and the B2 level in CEFR. Flesch-Kincaid Grade placed the text at 6th grade in all the tools. There was difference in Gunning Fog Index, which was off the scale in *Readable*, but placed at the equivalent of 8th grade in *Text Inspector*. Coh-Metrix L2 Readability was slightly above 27, which corresponds with A2 level.

Table 12: Analysis of Text 2 from Project 3

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	66,53	66,933	68,1	✘
Flesch-Kincaid Grade	6,85	6,505	6,3	✘
Gunning Fog Index	8,48	✘	5,4	✘
Average Sentence Length [words]	11,54	10,393	✘	✘
Type/token ratio	0,48	0,478	✘	✘
Coh-Metrix L2 Readability	✘	27,164	✘	✘
Average Syllables per Word	1,52	1,529	1,5	✘
Number of Sentences	26	28	27	✘
Number of Words	300	291	279	291

### Project 3: Text 3

The last text from Project 3 is a description of different transport options. It was divided into four sections, each with its own heading. There was a heading for the whole text as well. Contractions, numbers and text in brackets appeared in the text, as well as several special tokens, such as ‘mph’ or ‘M25’. There were compound words with a hyphen in the text for the first time. For the text, see Appendix I.

The results of the analysis of the text are seen in *Table 13*. Apart from the usual issues, which were very palpable in the sentence count due to the multiple headings in this case, there appeared a new problem regarding the compounds. Whereas *Vocab Kitchen* and *Coh-Metrix 3.0* counted a compound such as ‘double-decker’ and ‘man-made’ as one word, *Readable* and *Text Inspector* counted them as two. Flesch Reading Ease was similar enough for all the tools and the score indicated that it was equivalent to what would count as a text for the 7th grade, which translated to B1 level. However, the score was very close to a higher level. Flesch-Kincaid Grade was calculated to belong to 5th grade by the *Text Inspector* and 4th grade by the other two tools. A difference of nearly two grades appeared in the Gunning Fog Index, which indicated 6th grade in *Readable* and 7th, but nearly 8th grade in *Text Inspector*. Coh-Metrix L2 Readability suggests that the text belongs to A2 CEFR level.



Table 13: Analysis of Text 3 from Project 3

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	77,41	79,260	79,2	✘
Flesch-Kincaid Grade	5,64	4,937	4,9	✘
Gunning Fog Index	7,95	✘	6,0	✘
Average Sentence Length [words]	12,78	11,000	✘	✘
Type/token ratio	0,48	0,479	✘	✘
Coh-Metrix L2 Readability	✘	24,436	✘	✘
Average Syllables per Word	1,38	1,376	1,4	✘
Number of Sentences	32	37	37	✘
Number of Words	409	407	397	397

### Project 4

This section concerns the analyses of three text from Project 4 by Tom Hutchinson. The CEFR level of this students' book is estimated to be upper A2 to lower B1 (Hutchinson, 2014d). The texts discussed can be seen in Appendices J–L.

#### Project 4: Text 1

This text was a description of a job, which was introduced through two real persons. It seemed to fit the trend of the texts getting progressively more difficult, as it contained somewhat longer sentences and some potentially challenging grammar and vocabulary. There was also a heading, several numbers and contractions, as well as an abbreviation. For the text, see Appendix J.

Table 14 displays the result of the analysis, which shows several differences in the word and sentence count, all of which can be explained by previously discussed issues. Flesch Reading Ease and Flesch-Kincaid Grade results were in agreement amongst the tools, but the former placed the text at the 7th grade — the equivalent of B1 level, whereas the latter indicated that it belonged to the 5th grade. Gunning Fog Index in *Readable* placed the text at 7th grade, but *Text Inspector* calculated it to just short of the High-school freshman grade, which is a difference of nearly two grades. Coh-Metrix L2 Readability placed the text at the B1 level.

Table 14: Analysis of Text 1 from Project 4

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	75,45	76,384	74,9	✘
Flesch-Kincaid Grade	5,70	5,462	5,6	✘
Gunning Fog Index	8,99	✘	7,5	✘
Average Sentence Length [words]	11,93	11,500	✘	✘
Type/token ratio	0,50	0,495	✘	✘
Coh-Metrix L2 Readability	✘	20,309	✘	✘
Average Syllables per Word	1,41	1,404	1,4	✘
Number of Sentences	27	28	28	✘
Number of Words	322	322	309	310

### Project 4: Text 2

Second text from Project 4 was a narrative about King Arthur. It contained contractions and reported speech and the sentences were, at times, rather long. There was also a heading and one sentence contained a dash. The text can be seen in Appendix K.

The results of the analysis are to be seen in *Table 15*. The usual issues transpired regarding dashes and headings, but there appeared several new ones, some of which could be described as irregular. They were the joining of the sentences: “*The sword just fell into the water.*’ *King Arthur was very angry.*” in *Coh-Metrix 3.0* and the division of the sentence: “*Have you done it?*’ *asked King Arthur, when the knight returned.*” in *Readable*. *Coh-Metrix 3.0* also counted seemingly random apostrophes as words. Flesch Reading Ease and Flesch-Kincaid Grade were, again, similar enough throughout the different tools, but different between each other. The former evaluated the text at 6th grade, which should be roughly A2 level, and the latter placed it at 3rd grade. Gunning Fog Index differed again, with the higher value being calculated by *Text Inspector*. Both, however, were off the scale. *Coh-Metrix L2 Readability* placed the text somewhere between A2 and B1.

Table 15: Analysis of Text 2 from Project 4

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	88,57	88,878	87,1	✘
Flesch-Kincaid Grade	3,45	3,353	3,5	✘
Gunning Fog Index	5,84	✘	4,7	✘
Average Sentence Length [words]	10,24	10,026	✘	✘
Type/token ratio	0,38	0,377	✘	✘
Coh-Metrix L2 Readability	✘	23,726	✘	✘
Average Syllables per Word	1,28	1,274	1,3	✘
Number of Sentences	38	39	40	✘
Number of Words	389	391	382	383

### Project 4: Text 3

The last text from Project 4 described the country of Australia, with its history, inhabitants, fauna etc. It was a longer text with lots of information, including numbers — such as years or percentages. It contained contractions, dashes and a heading. This text can be seen in Appendix L.

The results are displayed in *Table 16*. This was the second text that *Text Inspector* assessed as over the limit. The big difference in word count could be explained by *Coh-Metrix 3.0* counting characters such as apostrophes or dashes as words — not all of them, however. *Coh-Metrix 3.0* and *Readable* both placed it at 6th grade according to the Flesch-Kincaid Grade and at 8th to 9th grade — the equivalent to B2 level — according to the Flesch Reading Ease. Gunning Fog Index placed it at early 9th grade. Coh-Metrix L2 Readability assessed it as A2 CEFR level.

Table 16: Analysis of Text 3 from Project 4

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	✘	69,003	65,1	✘
Flesch-Kincaid Grade	✘	6,331	6,9	✘
Gunning Fog Index	✘	✘	7,9	✘
Average Sentence Length [words]	✘	10,854	✘	✘
Type/token ratio	✘	0,520	✘	✘
Coh-Metrix L2 Readability	✘	25,272	✘	✘
Average Syllables per Word	✘	1,499	1,5	✘
Number of Sentences	✘	41	39	✘
Number of Words	✘	445	431	437

## Project 5

The last three texts come from Project 5 by tom Hutchinson. The level of this students' book should be the same as the previous one, that is upper A2 to lower B1 (Hutchinson, 2014e). For the individual text discussed in this section, see Appendices M–O.

### Project 5: Text 1

The first text from Project 5 described the education system in the USA from first grade all the way to university. The description included certain values and traditions. The language was not too difficult, but difference from the lower level texts was obvious. This text contained a heading and some apostrophes and brackets. For the text, see Appendix M.

*Table 17* displays the results of the analysis. The usual issues explain most of the differences that appeared in the analysis. Apart from them, it now transpired that even though *Text Inspector* usually divides words at apostrophes, it made an exception for the word 'o'clock', which it counted as a single word. *Coh-Metrix 3.0* took several irregular characters for words again. Flesch Reading Ease and Flesch-Kincaid Grade were calculated to be similar enough in the three tools, but the former placed the text at 7th grade — the equivalent of B1 level — and the latter at only 5th grade. Gunning Fog Index was higher again in *Text Inspector* — placing the text at 8th grade — and lower in *Readable*, which placed it at 5th grade only. *Coh-Metrix L2 Readability* evaluated the CEFR level at B1.

*Table 17:* Analysis of Text 1 from Project 5

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	75,91	78,669	75,4	✘
Flesch-Kincaid Grade	5,54	5,102	5,5	✘
Gunning Fog Index	8,32	✘	7,3	✘
Average Sentence Length [words]	11,55	11,333	✘	✘
Type/token ratio	0,45	0,457	✘	✘
Coh-Metrix L2 Readability	✘	22,208	✘	✘
Average Syllables per Word	1,41	1,379	1,4	✘
Number of Sentences	29	30	30	✘
Number of Words	335	340	330	329

## Project 5: Text 2

This text concerned a looming disaster in the shape of a supervolcano. It was filled with information and could be challenging regarding imagination. The text contained some potentially challenging vocabulary as well as different numbers, a heading, some apostrophes and, for the first time, degrees of Celsius. This text can be seen in Appendix N.

The results of the analysis of this text can be seen in *Table 18*. There were no new issues discovered during this analysis and the differences could all be explained by the usual issues — like apostrophes and numbers, some of which were larger than a thousand. Again, Flesch Reading Ease and Flesch-Kincaid Grade were very similar across the three tools, but the first formula assigned the text to 8th to 9th grade — therefore the B1 level — and the second to 7th grade. Gunning Fog index was higher in *Text Inspector* at the equivalent of High-school junior and it was very nearly High-school junior in *Readable*. Coh-Metrix L2 Readability placed it at B2 level, displaying the usual tendencies in evaluating the texts as somewhat more difficult.

*Table 18*: Analysis of Text 2 from Project 5

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	64,34	64,982	63,5	✘
Flesch-Kincaid Grade	7,63	7,379	7,4	✘
Gunning Fog Index	11,21	✘	8,9	✘
Average Sentence Length [words]	13,46	12,518	✘	✘
Type/token ratio	0,53	0,524	✘	✘
Coh-Metrix L2 Readability	✘	15,869	✘	✘
Average Syllables per Word	1,52	1,523	1,5	✘
Number of Sentences	26	27	27	✘
Number of Words	350	346	326	334

## Project 5: Text 3

The very last text was a brief history of the English language. It was, just like the rest of the texts from Project 5, palpably more advanced when compared to the texts from Project 1 or 2. The text contained a heading, several numbers and contractions, some potentially challenging words, abbreviation and a compound. The text was originally followed by a set of examples of words, but these were not included in the analysis, as they

could not be considered part of the text, but rather its complement. For this text, see Appendix O.

*Table 19* displays the result of the analysis of this text. There transpired no new issues during the analysis. All the discrepancies are due to apostrophes, numbers, compound words and the heading. Flesch Reading Ease was very similar amongst the tools, evaluating the text at 8th to 9th grade difficulty, which translates to B2 level. Flesch-Kincaid Grade placed it at 8th grade in *Text Inspector* and 7th grade in the remaining tools. Gunning Fog Index, again higher in *Text Inspector*, assessed the difficulty to be at High-school freshman, whereas *Readable* placed it at 7th grade. Coh-Metrix L2 Readability result indicated that the level of the text is B1.

*Table 19: Analysis of Text 3 from Project 5*

	Text inspector	Coh-Metrix 3.0	Readable	Vocab kitchen
Flesch Reading Ease	63,12	63,963	62,9	✘
Flesch-Kincaid Grade	8,01	7,729	7,8	✘
Gunning Fog Index	9,41	✘	7,8	✘
Average Sentence Length [words]	14,32	13,652	✘	✘
Type/token ratio	0,53	0,527	✘	✘
Coh-Metrix L2 Readability	✘	21,449	✘	✘
Average Syllables per Word	1,53	1,525	1,5	✘
Number of Sentences	22	23	23	✘
Number of Words	315	314	309	310

## EVP

The following section displays the results of the analysis of EVP in both tools that offer this function (*Text Inspector*, *Vocab Kitchen*). The results are first described in general manner, from which general tendencies are drawn, and later several examples of specific tokens are discussed as well as the differences resulting from them. For the general results the analyses of all the text are taken into consideration, but the specifics are discussed in a way that does not cover all of the texts.

The results of the analyses of EVP for all the texts are seen in Appendix P. From what the results show, several things can be concluded. *Text Inspector* placed more tokens in the ‘Unlisted’ category — it did so in 9 out of the 15 cases — which could mean that it has a smaller dictionary at its disposal. I did, however, disperse the rest of the tokens throughout the scale much more than *Vocab Kitchen*, which placed an overwhelming

majority of the tokens into the category 'A1', where it always displayed a higher number than *Text Inspector*. This means that *Vocab Kitchen* generally assessed the texts as easier than *Text Inspector*. A trend that was apparent in both tools is their evaluation of the texts as gradually more difficult as the level of the books increased. This was broken only by Project 4, the texts of which seemed more difficult regarding their vocabulary than those of Project 5. Project 4 and 5 should, however, have the same level.

Certain issues transpired when the analysis of the EVP was looked at more closely. For example, in the case of Text 3 from Project 1, 16 out of the 18 token types that were 'Unlisted' category were tokens that contained apostrophes — either ones that were connected to the words or ones that belonged to the words themselves, meaning contractions and possessive case — which would negate the theory about *Text Inspector* having a smaller dictionary and would instead mean that there was a problem with the tool's detection of apostrophes. Another difference that could be seen in the same text was the evaluation of the phrase 'looks up' as B1 word in *Text Inspector*, whereas the same was evaluated as A1 in *Vocab Kitchen* due to the fact that this tool divided the phrase into 'looks' and 'up'. This could be seen as a fault in *Vocab Kitchen*, as it did not correctly detect the use of the individual words when they appeared next to each other and their meaning and use therefore slightly changed. The same could be observed in Text 1 from Project 4, where the phrasal verb 'end up' was correctly assessed as such and subsequently placed in the B1 category in *Text Inspector*, whereas it was assessed as individual words by *Vocab Kitchen*, which then placed both words into the A1 category. The same text also uncovered another issue. The word 'pirates' was assessed by *Text Inspector* and placed into B1 category, but *Vocab Kitchen* put the same into the *unlisted* category. The same was true for words such as 'kangaroo, schoolchildren, cattle, copper, cricket' etc. There were, however, instances where the situation was the other way round. Words such as 'America, English' or 'done' appeared to not be present in the dictionary of *Text Inspector*, but they were assessed and accordingly placed by *Vocab Kitchen*.

### **Commentaries**

The aim of the research was to assess the different online tools and attempt to draw distinctions between them regarding their usefulness in reading text assessment. This could be viewed from several perspectives. The first is the amount of functions the tools offer, then their accuracy, their speed, and lastly their user-friendliness. This section is dedicated to the individual viewpoints and the way in which they influence the overall results.

When it comes to the amount of functions it is clear that *Text Inspector* holds the prime position. It is closely followed by *Coh-Metrix 3.0*, which, however, offers an overwhelming amount of different functions for free, none of which were relevant to this research. The least amount of functions is offered by *Vocab Kitchen*. It should be stated that the amount of functions goes together with the relevance of the functions and, regarding the practicality of text assessment in the context of the classroom, the EVP — a function that *Vocab Kitchen* offers — is especially useful, as it allows for the classification of individual words. *Text Inspector* also offers the EVP analysis. The readability formulas (Flesch Reading Ease, Flesch-Kincaid Grade, Gunning Fog Index) are all offered by *Text Inspector* and *Readable*, while *Coh-Metrix 3.0* does not calculate one of them — Gunning Fog Index. Instead, this tool calculates readability according to its own formula. These formulas are taken into consideration later on.

Accuracy is another, no less important aspect of the tools. It would be fair to say that none of the tools were without problems. Apart from some irregularities that were encountered in the testing of the tools, several recurring mistakes appeared in the tools, the most significant of which can be reviewed in *Table 20*. From this summarization an unfavourable conclusion can be drawn regarding *Text Inspector*, which might offer the most functions of the four, but it also shows the most mistakes. These mistakes, unfortunately, influence all the results that the tools calculate.

*Table 20*: Recurring mistakes that appear in the tools

	<b>Text inspector</b>	<b>Coh-Metrix 3.0</b>	<b>Readable</b>	<b>Vocab kitchen</b>
<b>Numbers as Words</b>	✓	✓	✗	✓
<b>Dashes as Words</b>	✗	✗	✗	✓
<b>Apostrophes divide words</b>	✓	✓	✗	✗
<b>Divides compounds with a hyphen</b>	✓	✗	✓	✗
<b>Comma divides numbers</b>	✓	✗	✗	✗
<b>Dashes divide sentences</b>	✗	✓	✗	?
<b>Titles not Separate Sentences</b>	✓	✗	✗	?
<b>Quotation marks connect sentences</b>	✗	✓	✗	?

The EVP is influenced by the mistakes, since it depends on accurate detection of words and phrases. This means that the inability of *Vocab Kitchen* to take the surrounding words into consideration inhibits the accuracy of the tool, but so do the issues that *Text Inspector* encounters with apostrophes. In the case of *Text Inspector*, this issue can be



worked around to a certain extent by deleting all apostrophes that appear in reported speech, but it would be both time consuming and only partially efficient, since apostrophes also appear in possessive case and contractions. The readability formulas are also influenced by the detection mistakes, because their score relies on the amount of words, sentence length, type/token ratio etc. Incorrect detection can therefore lead to an incorrect score. Assuming that the books were precise in their assessment of their own level, this was matched in less than half the cases — 7 out of 15 texts — by Flesch Reading Ease and even less than that by Flesch-Kincaid Grade, which usually assessed the texts as easier by several grades, and also Gunning Fog Index, which was usually off the scale due to the score being too low. A slightly better results were calculated using the Coh-Metrix L2 readability formula, which correctly assessed 8 out of the 15 texts. For the comparison of the performance of Flesch Reading Ease and Coh-Metrix L2 Readability in relation to the official levels of the texts, see Appendix Q. This somewhat disappointing accuracy could, of course, be caused by several things — the already discussed detection issues, an ineptness of the formulas themselves or a faulty conversion to CEFR level. To be fair, it could also be caused by either an incorrect self-assessment of the books — the opposite of which is the presumption that this research is based upon — or, more likely, the fact that it is the whole book's overall level that should correspond with the proclaimed level, not every single text or exercise, which could, logically, be subject to certain level fluctuations.

Speed is another characteristic that should be discussed. The fastest tools are, understandably, the ones which calculate the least amount of results — *Readable* and *Vocab Kitchen*. The slowest of all of them is, by far, *Coh-Metrix 3.0*, which is burdened by the sheer amount of calculations it needs to make. These differences are not overwhelming — they are in tens of seconds at most — but they might be something to consider in the case of evaluating either many texts or one very long text, as these differences will naturally grow.

The last important viewpoint that should be mentioned is the user-friendliness of the tools. This aspect is, needless to say, rather subjective, but it still offers some space for discussion. There are three subcategories discussed here — the appearance, the clarity of arrangement and the intricacy of use.

Appearance is perhaps the most subjective of all the categories. The most presentable of all the tools appears to be *Readable*, as it makes use of colour coding and the website has a unique, pleasing look that speaks in favour of the work of graphics that

handled the creation of the website. The rest looks somewhat more bland — especially *Coh-Matrix 3.0*. It is up to everybody to decide whether this aspect should be taken into considerations, as it has the least influence on the overall usefulness of the tools.

The clarity of arrangement impacts the use slightly more. It pertains to the way individual items are distributed throughout the site. This is negatively influenced by the amount of items that are there to be found, as is most perceptible in *Coh-Matrix 3.0*, which displays such a vast amount of data in one place that it necessitates a rather painstaking search every time a specific information is needed. The analysed text in this tool is placed within a small window and often cannot be seen in its whole extent. *Text Inspector*, which places different functions in clearly distinguished tabs that can be easily switched. The same is true for *Readable*. This tool also displays the analysed text next to the results, but it is placed in a large window and sections of it are marked by highlights that increase the orientation in the text. *Text Inspector* also displays the analysed text, which is even divided into individual sentences. In the case of EVP analysis that *Text Inspector* offers, the detected tokens are not only to be displayed under the individual sections pertaining to the CEFR levels, but they are also chronologically listed below, together with colour coding and the levels displayed next them. It is useful that the use of the tokens can be manually adjusted here, but the amount of information does not allow for fluent reading of the analysed text. This is better handled by *Vocab Kitchen*, as it displays the text only with colour coding. Such text can then be comfortably reviewed, and any difficult word is easily spotted. The clarity is here at the expense of the possibility of use adjustment, however.

The last subcategory that needs to be mentioned is the intricacy of use — the user friendliness. The only tool the free version of which needs to be opened in two browser windows for the use of all the functions worked with in this research is *Text Inspector*, which offers the EVP only in the version embedded on the English Profile website. The rest of the tools can be displayed in a single window. As for the work with the text, this is done through either typing or pasting of said text into a given window and then clicking a confirmation button that starts the analysis. The last part is not true for *Readable*, because this tool starts the analysis immediately on its own. The text can be revised and adjusted, and the analysis happens in real time. In all tools, except for *Text Inspector*, the text can be adjusted. The page needs to be refreshed in *Text Inspector* in order to do this. There is a further inconvenience that should be mentioned with regards to the confirmation of the analysis. *Coh-Matrix 3.0* has a verification code that needs to be transcribed before the

proceeding with the confirmation, which can be rather burdensome if several analyses need to be made.

All of the above-mentioned facts should be taken into consideration before deciding on the use of a single tool. Most of the issues discussed above influence the use of the assessment tools and it is most important that the potential user first addresses the question of what it is that he or she is looking for in terms of functions, accuracy ease of use etc. This leads to a perhaps unsatisfactory conclusion. The aim of this thesis was to evaluate the different tools that could be used for the assessment of readability of ELT texts and to attempt to pick from them the most suitable based on the testing that was conducted. However, with the overwhelming amount of differences in both advantages and disadvantages of the different tools, this is very hard to declare with certainty. The relevant issues in all the tools have been brought to attention in this research and it is for each user to review his or her needs and choose accordingly. What is vital, is that all the issues are known and taken into consideration before use, so that the tool can be worked with in a way that makes the most use of its functions. It is perhaps a good idea to not limit oneself to the use of a single tool, but rather use a combination of several and draw from the results of each with the simultaneous employment of a personal judgement. The tools that offer readability formulas can be used for a quick evaluation, whereas the ones that offer the EVP can help build a base for a deeper analysis of the readings. None of the tools have been found to be perfect, but all of them were useful in their own way. They are, however, only tools and the results from each of them need to be taken only as guiding information, not as a finished assessment.

## **V. IMPLICATIONS**

This chapter covers the circumstances and impact of the research. It contains an elaboration on the possibility of practical application of the subject of the research, as well as its value, including the limitations. The possible expansion of this research and several options for its broadening are also discussed.

### **Pedagogical Implications**

As was already mentioned, the reading materials in EVP classrooms are of two kinds — official, textbook materials and authentic ones. The research in this thesis was done on official reading texts, since the focus could thus be aimed at the testing of the tools, but the contribution it brings would likely influence the way that the latter kind of reading materials is considered, since it is usually an authentic text that needs to be assessed in terms of readability with more care. This research suggests several teaching implications.

It is very useful to have something at hand when assessing the overall level of a new reading text for a particular group of students. The tools discussed in this thesis could contribute to this by offering a chance to quickly assess these texts to approximately determine their level. When it comes to larger amounts of text aimed at extensive reading exercises, the formulas are particularly useful, as they establish the overall readability without focusing on details. As was already mentioned, it is not necessary to understand every single expression that appears in the texts, as long as the gist of it is understood, so the use of software tools to determine a numeric value according to a readability formula should be sufficient for the purposes of extensive reading.

When it comes to intensive reading, one where more focus is given to the grammar and lexis that appear in the text, the software tools could be useful as well. Even though they cannot help with the choice of texts that are used as an introduction to a particular grammar rule, they can help greatly with vocabulary. It was already established that the lack of vocabulary is often very limiting for readers of low proficiency, which means that the identification of potentially problematic lexical items before the lesson increases the chances of success for the students. The EVP function that the software tools offer is especially useful for this, as it can help detect the expressions that correspond with a higher level of proficiency than that of the students. Profiling of the texts beforehand could help determine what the focus of the pre-reading phase should be.

Regarding the use of specific tools for the assessment it was already established that no single one could be, based on this research, recommended as superior. Since they all display their own specific combinations of issues that influence their results, it would be best to use more than one of them. Each of them also offers a specific set of functions. *Text Inspector*, *Coh-Metrix 3.0* and *Readable* could all be used in the general assessment of texts, since they all feature the option to assess the texts based on readability formulas. *Coh-Metrix 3.0* even offers its own formula. The assessment should, logically, be more reliable if more tools and formulas are applied. For the EVP only two tools out of the ones discussed here could be used — *Text Inspector* and *Vocab Kitchen*. Even though their results more or less matched in this research, certain differences in their assessment were nonetheless found and it would be advisable to also use both of them for the analysis of the EVP.

### **Limitation of the Research**

There are certain aspects of this research that could have an impact on the results and the broadness of their potential application. These aspects should be taken into consideration when reviewing the results. The major ones are examined further in this section.

This research was conducted on the total of fifteen texts from five students' books by the same author. The reason for this was the need to have a bulk of samples that was consistent in their style and also their official assessment, however, given a larger space for research, this could have been broadened to other books and other kinds of textual material, which would result in more general results. The issue of space was also solved by the total number of samples, which, however, means further limitations caused by the sample volume.

A limitation that was already touched upon in the commentary section concerns the reliability of the official assessment of the tested samples. The accuracy of the results is partially dependent on the difficulty assessment of the control group. This problem could not be resolved in any other way than to trust the official assessment of the books, since the aim of the research was to test the individual tools, not the textual material. This assessment, however correct it might be, concerns the whole publications and not their individual parts. This could logically mean that the individual text, though likely not exceedingly more or less difficult than their proclaimed level, might be of slightly varying difficulties.

### **Suggestions for Further Research**

This research was aimed at the inspection of several software tools that could help with the assessment of text readability in the ELT context. The research, however, is far from exhausted. There are several possibilities regarding the expansion of this research.

The first, logical step, would be the identification of further software tools and their testing. Such tools could also offer further functions, which might be included alongside the ones present in this thesis. Some of these tools might only be available with a paid subscription, therefore it would make sense that the currently explored tools be bought in their full versions as well. It would also be interesting to see the differences between those and their freeware counterparts and subsequently ascertain whether the paid software is worth the money both in the context of the accuracy of the results, but also in the context of the demands and possibilities of English teachers. A survey could be made that would assess their needs and options and the openness to use such software. The openness of their school management to potentially supply funds for such tools could also be assessed.

It would also be valuable to expand the research in terms of the scope of texts tested. Text of varying length could be included, such as books that belong to the category of graded readers. Different topics and styles of texts might also be applied, since it is possible that the tools performance could increase or decrease depending on a specific type of text.

Practical testing is another area that could be explored. It is one thing to mechanically assess readability, but the reality could be different for schoolchildren and their teachers. It would be interesting to see whether the actual perception of the texts corresponds with the assessment based on the software tools and where the personal assessment of real English teachers stands.

## VI. CONCLUSION

This thesis aimed to explore the possibility of the assessment of text readability. The issue of different readability formulas was elaborated on and the options of their application in ELT practice was explored. For these purposes the research recorded in this thesis examined four different online tools that offered the option of free readability assessment.

Of the relevant functions that could be used in the context of language teaching *Text Inspector* was found to offer the largest amount. However, it also displayed the most mistakes in the detection of lexical items. This tool was then closely followed by *Coh-Metrix 3.0* in both aspects. The most common detection mistakes that appeared in the counting of numbers as words, the division of words that contained apostrophes and the division of hyphenated compound words.

Since the word and sentence count appear as variables in some of the readability formulas that the tools calculate, these mistakes then necessarily influence the final readability scores. However, it was found that despite the uncovered problems, the tools agreed on the Flesch Reading Ease score in all but one case in an extent sufficient enough for the texts to be placed in identical CEFR levels.

The CEFR levels calculated based on Flesch Reading Ease were the most successful out of the traditional readability formulas in matching the official levels of the books which the texts came from — placing 7 out of 15 text correctly. This was surpassed by the count of 1 by the Coh-Metrix L2 Readability formula offered by only one of the tools — *Coh-Metrix 3.0*. However, due to the possibility of level fluctuations within the books this result should not be taken as decisive.

For the assessment of the overall level of texts three of the four tools could be used — *Text Inspector*, *Coh-Metrix 3.0* and *Readable*. These three tools can calculate readability using readability formulas. Two tools, namely *Text Inspector* and *Vocab Kitchen*, can be used to create the EVP, which is effective in preparation for intensive reading lessons.

During the appraisal of the overall practicality it was concluded that *Readable* was the easiest and fastest to use, while *Coh-Metrix 3.0* was slowed down by the validation code and its results were the least well arranged. The free version of *Text Inspector* also limits the amount of text that can be analyzed, making it inappropriate for longer texts. The

EVP results are better displayed in *Vocab Kitchen*, but unlike *Text Inspector*, this tool does not correctly detect phrasal verbs.

It remains to conclude that each of the different tools offers its own specific combination of both functions, but also problems. The best advice that could be given based on this research is to use not one, but rather a combination of the tools, as that puts the resultant information into better perspective. What could be declared with certainty is that the analysis of reading texts in the context of ELT is, especially when it comes to authentic materials, a way to a better preparation for lessons.



## REFERENCES

- Beresova, J. (2015). Authentic Materials – Enhancing Language Acquisition and Cultural Awareness. *Procedia - Social and Behavioral Sciences*, 192, 195-204. doi:10.1016/j.sbspro.2015.06.028
- Budiono, D. (2010). Literary Texts as the Material In Teaching Reading Comprehension. *Magister Scientiae*, 27. Retrieved June 13, 2019, from <https://pdfs.semanticscholar.org/a882/c506a1f84298f4f90332d3edd67a4eed1fd4.pdf>.
- Coh-Metrix 3.0. (n.d.). Retrieved from <http://tool.cohmetrix.com/>
- Danielson, K. E. (1987). Readability Formulas: A Necessary Evil? *Reading Horizons*, 27(3), 178-188.
- Ellis, R. (2009). 1. Implicit and Explicit Learning, Knowledge and Instruction. *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*, 3-26. doi:10.21832/9781847691767-003
- Flesch, R., & Gould, A. J. (1949). *The art of readable writing*. New York: Harper & Row.
- Flesch, R., & Gould, A. J. (1949). *The art of readable writing*. New York: Harper & Row.
- Fulcher, G. (1997). Text difficulty and accessibility: Reading formulae and expert judgement. *System*, 25(4), 497-513. doi:10.1016/s0346-251x(97)00048-1
- Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40(2), 97-118. doi:10.1017/S0261444807004144
- Gray, W. S., & Leary, B. E. (1935). *What makes a book readable, with special reference to adults of limited reading ability*. Chicago, Illinois: University of Chicago Press.
- Green, A., Trim, J. L., & Hawkey, R. (2012). *Language functions revisited: Theoretical and empirical bases for language construct definition across the ability range*. Cambridge: Cambridge University Press.
- Gunning, R. (1968). *The technique of clear writing*. New York, NY: McGraw-Hill.
- Harmer, J. (2007). *The Practice of English language teaching* (4th ed.). Oxford, UK: Pearson Longman.
- Hutchinson, T. (2014a). Project 1 (4th Ed.) Oxford: Oxford University Press
- Hutchinson, T. (2014b). Project 2 (4th Ed.) Oxford: Oxford University Press
- Hutchinson, T. (2014c). Project 3 (4th Ed.) Oxford: Oxford University Press
- Hutchinson, T. (2014d). Project 4 (4th Ed.) Oxford: Oxford University Press
- Hutchinson, T. (2014e). Project 5 (4th Ed.) Oxford: Oxford University Press

- Jenkins, J. (2015). *Global Englishes a resource book for students*. Milton Park, Abingdon, Oxon: Routledge.
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. doi:10.21236/ada006655
- Legibility. (n.d.). In *Oxford Learners Dictionaries*. Retrieved from: <https://www.oxfordlearnersdictionaries.com/definition/english/legibility?q=legibility>
- Linguapress. (n.d.). The Flesch-Kincaid readability scales and their relevance for the EFL / ESL class. Retrieved from <https://linguapress.com/teachers/flesch-kincaid.htm>
- McKay, S., & Brown, J. D. (2016). *Teaching and assessing EIL in local contexts around the world*. New York: Routledge.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. New York, NY: Cambridge University Press.
- Metodický portál . (n.d.b). Retrieved from <https://clanky.rvp.cz/clanek/k/z/3024/SPOLECNÝ-EVROPSKÝ-REFERENCNI-RAMEC-PRO-CIZI-JAZYKY-V-CR.html/>
- Metodický portál. (n.d.a). Poznámky ke vzdělávacím oblastem. Retrieved June 16, 2019, from <https://digifolio.rvp.cz/view/view.php?id=10963>
- Online OCR. (n.d.). Retrieved from <https://www.onlineocr.net/>
- Readability. (n.d.) In *Oxford Learners Dictionaries*. Retrieved from: <https://www.oxfordlearnersdictionaries.com/definition/english/readability?q=readability>
- Readable.com. (n.d.). Retrieved from <https://app.readable.com/text/>
- Richards, J. C. (2015). *Key issues in language teaching*. Cambridge: Cambridge University Press.
- Scrivener, J. (2005). *Learning teaching* (2nd ed.). Macmillan.
- Štefanová, J. (2011). *Školní vzdělávací program a učebnice anglického jazyka pro základní školy*(Master's thesis). Filozofická fakulta Univerzity Palackého, Olomouc.
- Text Inspector. (n.d.a). Retrieved from <https://languageresearch.cambridge.org/wordlists/text-inspector>
- Text Inspector. (n.d.b). Retrieved from <https://textinspector.com/>

- Vijayarani, S., & Sakila, A. (2015). Performance Comparison of OCR Tools. *International Journal of UbiComp*, 6(3), 19-30. doi:10.5121/iju.2015.6303
- VocabKitchen. (n.d.a). Retrieved from <http://vocabkitchen.com/>
- VocabKitchen. (n.d.b). Retrieved from <http://vocabkitchen.com/profiler/cefr>
- VocabKitchen. (n.d.c). Retrieved from <http://vocabkitchen.com/profiler/awl>
- Wang, L., Miller, M. J., Schmitt, M. R., & Wen, F. K. (2013). Assessing readability formula differences with written health information materials: Application, results, and recommendations. *Research in Social and Administrative Pharmacy*, 9(5), 503-516. doi:10.1016/j.sapharm.2012.05.009
- Záhorová, J. (2017). *Principy převodu textu z tištěné do digitální podoby a způsob aplikace v osobních počítačích*(Bachelor's thesis). Západočeská univerzita, Plzeň.
- Zamanian, M., & Heydari, P. (2012). Readability of Texts: State of the Art. *Theory and Practice in Language Studies*, 2(1). doi:10.4304/tpls.2.1.43-53

## APPENDICES

### Appendix A

#### Project 1: Text 1

##### FLAT TO LET

There are four rooms in the flat and a hall. The hall is room D. There's a small table in the hall. There's a telephone on the table.

Room E is the living room. There's a sofa here and there are two armchairs, too. The television is here. There's a desk here, too, with a chair. There's a computer on the desk.

Room B is the bathroom. There isn't a bath here, but there's a big shower. There's a toilet and a washbasin, too. There's a big mirror on the wall.

Room A is the kitchen. There are lots of cupboards here. There's a fridge, a sink and a cooker. There's also a television and a radio here. We eat here, so there's a small table and there are four chairs. Room C is the bedroom. There are two beds here. There's a big wardrobe, and a chest of drawers. There's a lamp on the chest of drawers and a clock.

## Appendix B

### Project 1: Text 2

People from all over the world live in Britain. These young people were all born in Britain, but their parents or grandparents are from other countries.

Gabi lives in Edinburgh in Scotland. Her grandfather is from Hungary. 'We usually go to Budapest every summer, but I don't speak Hungarian. It's a very difficult language. Luckily, all my cousins there learn English at school.'

Mei's family is from Singapore. They live in Birmingham. Her parents work in a hospital there. 'I speak English and Chinese,' says Mei. 'We always speak Chinese at home, but at school I only speak English. Some of my friends are Chinese, too, but we always speak English to each other'

Desmond's grandparents are from Nigeria in Africa. Desmond lives in Manchester 'Everyone in Nigeria speaks English. It's the national language,' he says. 'There are a lot of African languages there, too, but I don't speak any of them. I only speak English. I'm learning French at school.'

Kathir lives in Oxford. His parents are from Sri Lanka. 'We speak English at home,' he says. 'My parents also speak Tamil — an Indian language — but I don't. I can speak Spanish. We learn it at school and it's my favourite subject.'

This is Emre. His family is from Turkey. 'There are a lot of Turkish people in our part of London, so I speak English and Turkish. I speak English most of the time, but we visit my grandparents in Turkey every year and they don't speak English.'

## Appendix C

### Project 1: Text 3

#### The Fox and the Crow

The fox is hungry. He's looking for food, but he can't find anything to eat. Then he sees something. It's a piece of cheese — nice, yellow cheese. The fox loves cheese.

He runs to get it, but then he sees a bird — a big black bird. It's a crow. The crow likes cheese, too. The big, black crow lands next to the cheese, picks it up in her beak and flies up into a tree.

The fox isn't happy. He's still hungry, but now he can't eat the piece of cheese. He can see it. It's in the crow's beak and the crow is sitting on a branch in the tree.

'I really want that piece of cheese,' thinks the fox. 'But how can I get it? I can't climb trees. And the crow can fly away.'

The fox is very clever and he has an idea. He walks to the tree and he looks up. The crow is sitting on a branch. She's holding the piece of cheese in her beak. She sees the fox, but she doesn't fly away. She knows that the fox can't get her. She's safe in the tree.

'Good morning, Mrs Crow,' says the fox and he smiles. 'How are you today?' The crow doesn't say anything. She can't speak because she's holding the piece of cheese in her beak.

'Isn't it a lovely day?' says the fox. 'The sun is shining. The birds are singing.'

The crow doesn't say anything.

'But you aren't singing,' says the fox. 'Why aren't you singing today, Mrs Crow? Everyone says that you've got a beautiful voice.'

Now everyone knows that the crow can't sing. She can only say 'Caw. Caw.' But the fox says: 'Please sing for me. Mrs Crow. I want to hear your beautiful voice.'

The crow wants to sing. She wants to have a beautiful voice.

'Please sing your beautiful song for me,' says the fox again.

The crow wants to sing. She opens her beak and says: 'Caw. Caw.'

But the fox isn't listening to her. When she opens her beak, the piece of cheese falls. It lands in front of the fox. He looks at the cheese. The crow looks at the cheese.

'Thank you,' says the fox. Then he picks up the piece of cheese — the nice, yellow cheese — and he eats it.

## Appendix D

### Project 2: Text 1

My name's Henry Mills. I'm twelve years old and I live in London. We live in a flat. It's on the third floor.

I go to Central High School. I'm in class 8A. My favourite subjects are Maths, History and PE. I don't like English or Physics.

I walk to school with my friend, Nadim. He lives on the seventh floor of our block of flats. We leave home at half past eight. School starts at twenty to nine with registration and assembly, and our first lesson starts at nine o'clock. School finishes at half past three. After school we go home and I do my homework before dinner. In the evening I watch TV or I play a computer game. I practise the guitar, too.

On Thursdays I don't go home at half past three. I have a guitar lesson at school. I'm not very good yet but I really like it. I want to be in a band when I grow up.

On Saturday mornings Nadim and I play football at the sports centre.

I've got a sister. Her name's Alison. She doesn't go to high school. She's seventeen years old, so she goes to sixth form college. The college isn't near our flat, so she takes the bus. She leaves home at ten past eight and her lessons start at quarter to nine.

## Appendix E

### Project 2: Text 2

#### FESTIVALS

Christmas is the biggest festival of the year. Families usually come together to celebrate it. People put up decorations and decorate a Christmas tree. In many countries people open their presents and have their Christmas meal on Christmas Eve (24 December), but in Britain people normally do these things on Christmas Day (25 December). The traditional meal is turkey with vegetables, followed by Christmas pudding. The next day (26 December) is a holiday, too. We call it Boxing Day. On Boxing Day, people often visit friends, or go to sports matches. They go shopping, too, because a lot of shops start their winter sales on Boxing Day.

On New Year's Eve, a lot of people have a party to 'see in' the new year. They often switch on the radio or the TV to hear Big Ben in London strike midnight. Then they usually join hands and sing Auld Lang Syne. In London, Edinburgh and several other cities there are big street parties with music and fireworks.

We also celebrate Easter in Britain, but it isn't a very big festival, like Christmas or New Year. The Easter holiday lasts four days from Good Friday to Easter Monday. People eat sweet bread buns called hot cross buns on Good Friday and on Easter Sunday people usually give chocolate Easter eggs. Some families have an Easter egg hunt in the garden. They hide small Easter eggs for children to find.



## Appendix F

### Project 2: Text 3

Chicken Licken is in the farmyard. He's eating under a tree when suddenly a nut falls on his head. 'Oh, no,' says Chicken Licken. 'The sky is falling down. I must go to London and tell the King.'

Chicken Licken runs down the road. When he meets Henny Penny, he stops. 'Where are you going?' she says. 'Oh, Henny Penny,' says Chicken Licken. 'The sky is falling down and I'm going to London to tell the King.' 'Well, I must come with you,' says Henny Penny. So Chicken Licken and Henny Penny run down the road. In the village, Ducky Lucky and Goosey Loosey are swimming on the pond. When Chicken Licken and Henny Penny see them, they stop. 'Hello,' says Ducky Lucky. 'Where are you running to?' 'Oh, Ducky Lucky,' says Chicken Licken. 'The sky is falling down and we're going to London to tell the King. Do you want to join us?' 'Yes, we must come with you,' say Ducky Lucky and Goosey Loosey. Chicken Licken, Henny Penny, Ducky Lucky and Goosey Loosey run down the road. At the side of the road is Foxy Loxy. He's sitting under a tree. When they see him, they stop. 'Hello,' says Foxy Loxy, 'where are you going?' 'Oh, Foxy Loxy,' says Chicken Licken. 'The sky is falling down and we're going to London to tell the King.' 'To London?' says Foxy Loxy. 'Well, you're going the wrong way. That isn't the road to London. You must go this way. Follow me.'

So Chicken Licken, Henny Penny, Ducky Lucky and Goosey Loosey follow Foxy Loxy. But, oh dear, he doesn't take them to London. He takes them to his den. His wife and children are waiting for him — and they're very hungry!

Chicken Licken, Henny Penny, Ducky Lucky and Goosey Loosey never get to London and they never tell the King that the sky is falling down.

## Appendix G

### Project 3: Text 3

#### MY NEW HOME

My name's Carl. I live in London now, but I wasn't born in England. I was born in Auckland in New Zealand. We moved to England two years ago. My dad got a new job here. I didn't want to leave New Zealand. I had lots of friends there. The weather was great and our house was near the sea. In summer, I went to the beach after school two or three times a week. I went sailing a lot, too.

I didn't like it in England when we first arrived. I didn't have all my friends here. They were on the other side of the world. We weren't near the sea, and it was winter, so the weather wasn't very nice. It was colder than Auckland, and when we arrived it rained a lot! I just wanted to go back to New Zealand.

Things got better when I started school. There are kids from lots of different countries at my school here, so I wasn't unusual. I soon made a lot of friends.

I still miss New Zealand, but I like it here now. London's a great city. I can't go to the beach after school, but I go to the sports centre with my friends. I don't go sailing now. I play football. I didn't play it in New Zealand. We played rugby at my school. I prefer football and I'm quite good at it. I'm in the school team this year.

## Appendix H

### Project 3: Text 3

#### THE HUMAN STORY

Modern human beings (homo sapiens) first appeared in Africa 150,000 years ago. For thousands of years they stayed in Africa, but about 60,000 years ago, they started to migrate into Asia.

From there, some went north and arrived in Europe about 35,000 years ago. Europe was much colder than today, as this was the time of the Ice Age. However, there was plenty of food there. Our ancestors hunted huge animals, like mammoths and woolly rhinoceroses. There were already other people in Europe, too. These were the Neanderthals. They left Africa 200,000 years ago. They were bigger and stronger than our ancestors, but by 20,000 years ago, the Neanderthals were extinct. We don't know why.

Other groups of modern humans went east, further into Asia. From there, some went south into Australia. The first people reached Australia about 45,000 years ago. Today Australia is an island, but during the Ice Age a lot of the Earth's water was ice. So sea levels were 20 metres lower than today. The seas between Asia and Australia were much shallower and in many places there were land bridges. So it was easy to cross into the new continent.

There was also a land bridge in the north between Siberia and Alaska. About 15,000 years ago, people walked across this land bridge into North America. They moved down through Central America and reached South America about 11,000 years ago. Human beings were now in every continent except Antarctica.

Soon after that — about 10,000 years ago — the last Ice Age ended. As the ice melted, sea levels rose. Australia became an island again and the land bridge to North America disappeared. No new people arrived in these places until modern times.

## Appendix I

### Project 3: Text 3

#### Transport

##### ROADS

Vehicles in Britain and Ireland travel on the left, so the steering wheel is on the right. If you live in a country where vehicles travel on the right, you must be extra careful when you cross the road. Always look right first! The busiest motorway in Britain is the M25. It goes round London. It's almost 200 kilometres long and is one of the longest ring roads in the world. In some countries you have to pay a toll to use the motorways, but in Britain and Ireland most of them are free. Distances in Britain are in miles (three miles is approximately five kilometres). And speeds are in miles per hour (mph).

##### RAILWAYS

For long journeys, the train is usually the fastest way to travel. You can also travel by coach. It's cheaper, but it usually takes longer. The main railways connect London to other cities. London hasn't got a central station. Different companies built the railways, and they all had their own station in London. Victoria and Waterloo, for example, are the stations for trains to the south. If you want to go to the west, however, you have to go from Paddington. From St Pancras Station you can take a Eurostar train through the Channel Tunnel to cities in France and Belgium.

##### TRANSPORT IN LONDON

The fastest way to travel in London is on the Underground (or the Tube, as it's also called). But you can't see any of the famous places down there, so if you want to see London's sights, it's better to take one of the famous red double-decker buses. You'll get a good view if you sit upstairs. However, like a lot of big cities, the traffic in London can be very bad, so the bus often takes a long time. You can also take one of London's black taxis, or even a cycle rickshaw. They can be quicker, but they're both expensive. Often the best thing to do in London is to walk.

##### AIRPORTS

There are several big airports in Britain. London has got five. Heathrow is the biggest, and it's the busiest airport in Europe. A plane takes off or lands there every 75 seconds. The most popular destination is New York. Some people think that London needs a new airport. There are plans to build one on a man-made island in the River Thames.

## Appendix J

### Project 4: Text 1

#### STUNT DOUBLES

Have you ever done anything really dangerous? Have you fallen off a bridge? Have you walked through a fire? Rick English and Angela Meryl have. You've probably never heard of Rick or Angela, but you've seen them when you've watched a film at the cinema or on a DVD. They've been in a lot of famous films, including Kill Bill, Pirates of the Caribbean, Casino Royale and the Harry Potter films. You haven't heard of them, because they're stunt doubles. Most film stars don't want to do dangerous things in films, so stunt doubles do them. Rick and Angela have been stunt doubles for some of the biggest movie stars.

They've done a lot of very dangerous stunts. Rick has driven cars through walls and into rivers. Angela has fallen off bridges, high buildings and motorbikes. She's also fallen out of windows and cars. Stunts look good in films, but they can be very dangerous. In 2002, Angela fell through a glass table in the film Kill Bill. The glass cut her hand very badly and she ended up in hospital.

Probably the most dangerous stunt ever was in the film Cliffhanger. The stuntman, Simon Crane, climbed from one aeroplane to another, while they were flying at nearly five thousand metres. That was also the most expensive stunt ever. Simon got a million dollars for it.

'Danger is part of our life,' says stuntman Frank Street. 'A lot of stunt doubles have died and most of them have been in hospital several times. In fact, I've just come out of hospital myself. I broke my arm when I jumped out of a helicopter last week.' Because stunts are so dangerous and expensive, film directors have started to use CGI (Computer Generated Imagery) instead. It's safer and cheaper. So, in the future, there probably won't be any stunt doubles.

## Appendix K

### Project 4: Text 2

#### Sir Bedivere and Excalibur

Long, long ago there lived a famous king. His name was Arthur and he was the king of the Britons. King Arthur and his Knights of the Round Table fought many battles. Arthur's last battle was against his own son, Mordred, at Camlan. The armies started fighting early in the morning and the battle lasted all day. When they finished fighting, Mordred was dead and the king was badly wounded. Only one of Arthur's knights survived — Sir Bedivere.

After the battle of Camlan, Sir Bedivere took the wounded king to the island of Avalon. The island was in the middle of a lake. Arthur knew the lake well. He remembered the day when he received his magic sword, Excalibur, from the Lady of the Lake. 'That was a long time ago,' he thought. Then he said to Sir Bedivere: 'I need to return Excalibur to the lake before I die,' But he was too weak to move, so the knight offered to take the sword. 'Throw the sword out into the middle of the lake,' said the king. Sir Bedivere promised to do it.

The knight stood on the shore of the lake and looked at Excalibur. It was such a beautiful sword. He couldn't imagine throwing it away. He decided to keep it. On the way back to the king, he stopped to hide the sword in the bushes. 'Have you done it?' asked King Arthur, when the knight returned.

'Yes, Your Majesty,' said Sir Bedivere.

The king smiled. 'Tell me,' he said. 'What did you see?'

'Nothing,' said the knight.

The king stopped smiling. 'You didn't throw the sword into the lake,' he said. 'Go back and do it.'

Sir Bedivere didn't want to throw the sword into the water, but he agreed to do it and went back to the lake. He picked up the sword and started to throw it, but he couldn't do it. He hid the sword again and went back to the king. Again the king asked him: 'What did you see?'

'Nothing happened,' he said. 'The sword just fell into the water.' King Arthur was very angry. He refused to believe the knight's story. 'I see that I must do it myself,' he said, and he tried to stand up.

## Appendix L

### Project 4: Text 3

#### AUSTRALIA

Australia is the sixth-largest country in the world. However, fewer than 20 million people live there. The first people there were the Aborigines. They arrived in Australia about 10,000 years ago. During the last Ice Age. At that time, sea levels were lower and people could easily cross from Asia. When sea levels rose again, Australia was cut off until it was discovered by a Dutch explorer, Abel Tasman, in the late 17th century. Today, only 1.5% of Australia's population are Aborigines.

The name 'Australia' comes from the Latin terra australis, which means 'southern land'. In the 18th century, the British explorer, Captain James Cook, claimed the land for Britain. At first, Australia was used as a prison. Criminals from Britain were sent there. Later, other people from Britain and Ireland went to live there. The British influence is still very strong. The official language is English, cars drive on the left, and the national game is cricket. In recent years, however, many immigrants have come from other European countries and from Asia, too.

Australia was cut off from the other continents a long time ago, so a lot of Australia's animals, like the kangaroo, the koala and the platypus aren't found anywhere else. Australia also has some of the most dangerous animals in the world. There are poisonous snakes and spiders. Several people are bitten every year, and schoolchildren have to learn first aid for snake and spider bites. In the north, there are huge saltwater crocodiles and in the sea, around the coast, there are sharks and poisonous jellyfish.

The largest city is Sydney with its famous Harbour Bridge and Opera House. The 2000 Olympics were held in Sydney. Sydney, however, isn't the capital of Australia. The capital is Canberra.

Australia is a huge country — approximately 4,000 km across — but most Australians live in the south-east corner between Sydney and Melbourne. This is because 90% of the land is very dry. Australians call this dry area the outback. There are some huge farms there with millions of sheep and cattle. There are nine sheep for every person in Australia. The children who live on these farms can't go to school, because the nearest town is too far away. They study at home with the School of the Air. They talk to their teacher by radio and, nowadays, by telephone, the Internet and webcams, too. There is also a 'flying doctor' service. The doctor comes to you by plane.

Australia is a rich country. Meat, fruit, vegetables, wool and wine are produced. Gold, silver, iron, copper, coal and diamonds are found there, too.



## Appendix M

### Project 5: Text 1

#### EDUCATION IN THE USA

Children In the USA start Grade School when they are six years old. When they are eleven, they go to Junior High School, and at the age of fourteen to Senior High School.

The school day usually runs from seven or eight o'clock in the morning to half past two or three o'clock. Pupils have homework to do after that. In most schools, the day starts with a flag ceremony. Pupils and teachers stand with their hand on their heart and make a pledge of loyalty to the flag and the country while the American flag is raised.

American schools try to create a strong sense of community. Schools have their own teams for American football, baseball, ice hockey and basketball. They usually have bands and drama groups, too. When sports teams play against other schools, it's a big event. Pupils and their parents come to watch. The cheerleaders dance and the school band plays, too.

Pupils can leave school when they are sixteen, but most stay at school till they are eighteen. Then they go to university. (Americans normally call it 'college'.) They usually spend four years at university. They don't always go to a local university. A lot of students study in cities hundreds of miles from their home.

State education in the USA is free up to the age of eighteen. There are some private schools that parents pay for, but most children go to state schools. College isn't free. Parents have to pay, and students often have a part-time job, too. They work in shops, restaurants and petrol stations.

When students leave school (including Grade School and Junior High School) or university, they have a Graduation Day. This is a very big event. The students who are graduating wear gowns and caps. Their parents come to watch their children receiving their graduation certificates. At the end of High School there's usually a big dance, too. It's called the 'High School Prom'.

## Appendix N

### Project 5: Text 2

#### SUPERVOLCANO

Under the quiet green forests of Yellowstone National Park in the USA is a sleeping monster. Nobody knows its exact size, but it's about 80 kilometres long and 45 kilometres wide, and it looks like a peaceful valley covered in forests. In fact, it's a huge crater full of magma, or molten rock. If it erupted, the explosion would be bigger than 1,000 atomic bombs. Scientists call it a 'supervolcano'.

Volcanoes don't often cause damage over a wide area, but the eruption of a supervolcano would be a global disaster. It would destroy everything within 1,000 kilometres. Lava would cover an area as big as Europe. The explosion would be so strong that it would cause earthquakes, tsunamis and hurricanes. The hot lava would start forest fires all over North America.

People in the rest of the world wouldn't escape the destruction. The supervolcano would pollute the atmosphere with billions of tons of ash and dust. The wind would blow this pollution around the world and we would enter 'a volcanic winter'. When the ordinary volcano Tambora in Indonesia erupted in 1815, it produced 'the year without a summer' of 1816. There was so much dust in the atmosphere that even in Europe people didn't see the sun for several weeks that year. A supervolcano would be much worse. We wouldn't see the sun for a couple of years. Temperatures would fall by up to 15°C in North America and 3-5°C worldwide. Food production would stop, because plants wouldn't grow and animals would die. Famine and wars would follow.

If the monster under Yellowstone woke up, it would be the end of life as we know it. And it isn't the only one. There are supervolcanoes in California, near Naples in Italy, in Indonesia and in New Zealand. But could it really happen? Scientists at Yellowstone don't think it will happen soon, but it has erupted before. In fact, it usually erupts every 600,000 years, and the last eruption was 640,000 years ago!

## Appendix O

### Project 5: Text 3

#### The English language

The story of the English language began in the 5th century when tribes from north-western Europe invaded Britain. We call these people the Anglo-Saxons. They spoke a language that was similar to modern German. A lot of common English words, like good, man and break, come from the Anglo-Saxon language. Over the next thousand years, however, this language changed, because other groups of people came to live in Britain.

The first change started in about 800 AD, when England was invaded by the Vikings from Norway and Denmark. They brought words like sky, get and husband into English. The 'ths' sounds in words like these, thing and father also come from the Viking languages.

In 1066, the Normans from northern France conquered England. The Normans spoke French. For the next 300 years, there were two languages in England. The ordinary people spoke English, but the aristocracy spoke French. The two groups didn't understand each other very well, so the grammar that they used became much simpler. A lot of French words, like parliament, language and beef, came into English, too. Slowly the modern language of English was born.

From the 17th century, English spread to many other parts of the world. Most of the early European settlers in North America came from Britain, so English became the language of the United States and Canada. Later, the British Empire took English to Australia, New Zealand, South Africa, India and several other countries in Africa, Asia and the Caribbean.

Today, over 400 million people speak English as their first language, and billions of people use it as a second language. English has become the international language of business, entertainment, music, sport, science, transport and the Internet.

Languages change all the time, so English isn't the same all over the world. Here are some differences between British and American English.

## Appendix P

### The comparison of EVP results from Text Inspector and Vocab Kitchen

Text	Tool	A1	A2	B1	B2	C1	C2	Unlisted
Project 1: Text 1	Text Inspector	77,99	8,81	1,89	0,00	0,00	0,00	11,32
	Vocab Kitchen	96	6	0	2	0	0	5
Project 1: Text 2	Text Inspector	69,26	4,67	1,17	0,00	0,00	0,00	24,90
	Vocab Kitchen	83	5	1	0	0	0	14
Project 1: Text 3	Text Inspector	72,52	5,46	2,48	5,20	1,49	0,00	18,87
	Vocab Kitchen	86	5	3	0	1	0	10
Project 2: Text 1	Text Inspector	83,04	7,14	1,34	0,45	0,00	0,45	7,59
	Vocab Kitchen	90	7	2	0	0	0	6
Project 2: Text 2	Text Inspector	72,96	5,15	8,15	1,29	0,00	0,00	12,45
	Vocab Kitchen	74	5	5	0	0	0	15
Project 2: Text 3	Text Inspector	50,16	13,79	4,39	0,94	0,94	0,00	29,78
	Vocab Kitchen	67	15	3	0	0	0	18
Project 3: Text 1	Text Inspector	80,66	8,23	0,82	2,06	0,00	0,00	8,23
	Vocab Kitchen	91	7	2	0	0	0	6
Project 3: Text 2	Text Inspector	57,76	21,30	7,22	1,44	0,36	0,00	11,91
	Vocab Kitchen	62	16	7	0	0	0	18
Project 3: Text 3	Text Inspector	68,87	15,48	4,82	0,25	0,25	0,00	10,41
	Vocab Kitchen	73	14	5	0	0	0	11
Project 4: Text 1	Text Inspector	58,75	19,47	0,99	3,63	0,00	0,33	16,83
	Vocab Kitchen	69	19	0	0	0	0	15
Project 4: Text 2	Text Inspector	55,93	16,24	6,96	5,15	0,00	0,00	15,72
	Vocab Kitchen	65	17	4	2	0	0	13
Project 4: Text 3	Text Inspector	60,28	15,37	6,86	3,07	0,47	0,00	13,95
	Vocab Kitchen	62	14	7	2	0	0	16
Project 5: Text 1	Text Inspector	73,77	10,80	5,86	3,40	0,00	0,00	6,17
	Vocab Kitchen	77	12	5	2	0	0	7
Project 5: Text 2	Text Inspector	58,33	15,12	7,41	5,25	0,00	1,54	12,35
	Vocab Kitchen	64	14	6	3	0	0	18
Project 5: Text 3	Text Inspector	62,58	11,92	2,98	3,61	0,66	0,33	17,88
	Vocab Kitchen	70	11	3	1	1	0	16

## Appendix Q

**The comparison of the performance of Flesch Reading Ease and Coh-Metrix L2 Readability in relation to the official levels**

	Official level	Flesch Reading Ease	Coh-Metrix L2 Readability
<b>Project 1: Text 1</b>	A1	A1	A2
<b>Project 1: Text 2</b>	A1	B1	A2
<b>Project 1: Text 3</b>	A1	A1	A2 to B1
<b>Project 2: Text 1</b>	upper A1 to A2	A1	A2
<b>Project 2: Text 2</b>	upper A1 to A2	B2	A2 to B1
<b>Project 2: Text 3</b>	upper A1 to A2	A2	B1
<b>Project 3: Text 1</b>	upper A2	A1	A1
<b>Project 3: Text 2</b>	upper A2	B2	A2
<b>Project 3: Text 3</b>	upper A2	B1	A2
<b>Project 4: Text 1</b>	upper A2, lower B1	B1	B1
<b>Project 4: Text 2</b>	upper A2, lower B1	A2	A2 to B1
<b>Project 4: Text 3</b>	upper A2, lower B1	B2	A2
<b>Project 5: Text 1</b>	upper A2, lower B1	B1	B1
<b>Project 5: Text 2</b>	upper A2, lower B1	B1	B2
<b>Project 5: Text 3</b>	upper A2, lower B1	B2	B1

## SHRNUTÍ

Předmětem této diplomové práce je hodnocení čtivosti výukových textů pro anglický jazyk. Přesné vyhodnocení textů pro čtení je v rámci výuky anglického jazyka naprosto nezbytné a tím nutně vzniká potřeba zjistit jakým způsobem jej nejlépe provádět. Existuje množství různých zavedených vzorců pro výpočet čtivosti a také online programů, které je dokáží vypočítat. Na zmíněné programy a jejich užití se tato práce soustředí s cílem zjistit odlišnosti mezi nimi a pokusit se tak vyvodit závěr ohledně případného doporučení konkrétních programů. Za tímto účelem bylo ve vybraných programech analyzováno patnáct textů a výsledky byly následně porovnány a kriticky zhodnoceny. Ve výsledku bylo doporučeno užití kombinace více programů spíše než jednoho konkrétního. Pro rychlé zhodnocení čtivosti bylo doporučeno použít programy, které umožňují vypočtení skóre čtivosti, tedy Text Inspector, Coh-Metrix 3.0 a Readable. Pro detailní analýzu slovní zásoby byly pak doporučeny programy Text Inspector a Vocab Kitchen, které umožňují vytvoření profilu anglického jazyka (English language profile).