Journal of Faslo el-khitab مجلة فصل الخطاب
ISSN:1071-2335/ E-ISSN:2602-5922/ Legal Deposit N°: 2012-1759
Vol 11, N° 01, March 2022, pp 535 - 554

تاريخ الاستلام(2021/11/14) تاريخ القبول (2022/02/03) تاريخ النشر (2022/03/30)

BNP Formulas for Algerian Middle School EFL Learners: Predicting Readability through Estimated Reading Time

Younes BEHIRA[1], **Abdelghani NAIT-BRAHIM**[2], **Ulrike PADO**[3]

[1] Mohamed Ben Ahmed University of Oran 2 (Algeria), behirayounes@gmail.com
[2] Higher Normal School of Oran (Algeria), nait.abdelghani@ens-oran.dz
[3] University of Applied Sciences of Stuttgart (Germany), ulrike.pado@hft-stuttgart.de

Summary:
Readability formulas for English are normed to predict grades in the United States (Flesch 1948; Chall and Dale 1995). This limits their usability outside of their creation context and schooling system, especially for EFL teaching. This study proposes to predict student reading times as an empirically observable operationalization of readability. Specifically, we induce formulas to be used by Algerian classroom teachers in predicting the readability of the reading texts for middle school EFL learners through estimated reading time. Learners in different middle schools in Algeria participated in the study by reading selected texts to get the approximate reading time for each category of readers. BNP formulas are the first that combine a characteristic of the text, i.e., character count, with a characteristic of the intended readers, i.e., reading speed, to predict the readability of the text to EFL readers through estimating their reading time.
Keywords: EFL, Reading; Readability Formulas; Estimated Reading Time; reading Speed; Character Count.

1. INTRODUCTION

The fact that the Algerian BEM[1] English examination is exclusively of the written mode is the main reason behind the importance given by most classroom teachers to the reading skill. In this exam, learners have to read a text to do the subsequent exercises. However, it has been noticed that middle school teachers find difficulties in selecting and/or adjusting the reading texts to their learners' level. Such task is an impression-based professional activity for the majority of teachers, as they have no specific scientific tools, such as readability formulas, to make use of. The existing

∗ Corresponding author: **Younes BEHIRA**, e-mail: behirayounes@gmail.com

formulas, whether for readability prediction or reading time estimation, cannot be adopted by Algerian classroom teachers due to the many differences between the contexts they are devised for and the Algerian EFL context. Hence, devising specific formulas for Algerian middle school EFL learners becomes an urgent need to facilitate the selection and/or adaptation of Algerian middle school EFL textbook reading texts for both textbook writers and classroom teachers.

Additionally, teachers usually have to manage their session time by allotting specific time to each activity learners do in class including reading. The latter is the most challenging activity for both the learners, whose understandability of the text determines the extent to which they can do the accompanying tasks, and the teachers, whose main preoccupation is providing learners with a readable text, i.e., a text that can be read and understood in a specific time. Gérard and Roegiers (2009: 244) affirm that “the degree of readability does not depend only on the text per se (and its support), but also on the reader’ characteristic[2].” One of the readers’ characteristics is the reading speed which differs among Algerian middle school learners as an English beginner takes more time reading a text than an intermediate learner. Therefore, a formula that estimates the reading time of a text can help the teacher choose appropriate texts and plan their reading sessions. For instance, let’s estimate that the maximum reading time set for a specific group of learners is 6 minutes which is equivalent to 10% out of a 60-minute session. If the estimated reading time of the text is longer than the time set for reading, then the teacher will find the text unsuitable for his learners, and thus either adapt it to fit the set reading time or search for another one.

BNP formulas provide teachers with approximate estimated reading time for a text to help them manage their time in the classroom and predict the suitability of the text for the target readers. Moreover, the formulas can be used by Algerian textbook writers to select appropriate reading texts that match the target learners’ reading ability. Using these formulas, textbook writers can also provide classroom teachers with the approximate estimated reading time for each selected reading text to help them anticipate and manage their reading sessions. Additionally, the formulas will help textbook designers maintain length gradation of the textbook reading texts and consistency along middle school levels (MSLs). Furthermore, BNP formulas are developed for both intensive and extensive reading sessions. A text that takes 30 minutes to be read by learners is too long to be taught in a 60-minute session; however, teachers can use the text for an extensive reading session.

In should be noted that BNP formulas are by no means deemed here the unique tools to predict the readability of the texts as they are based on measurable data. It is admitted that other non-measurable text characteristics, such as composition, obscurity and topic, are also to be taken into consideration when selecting and/or adapting a text. For instance, with regard to topic, a text on football is predicted to be less difficult to read, thus requiring less reading time, than a text discussing globalization due the differences in vocabulary as it can be hypothesized that beginners are more familiar with football-related vocabulary than the one of globalization. Thus, BNP formulas focus on measurable data related to the length of the text and the target reader's reading speed.

1.1 Purpose of the Study

To achieve the purpose of the study, we derive linear regression models to predict student reading times in different grade levels given properties of the text such as the number of words, syllables or characters. These properties are well-established in the literature as useful for reading difficulty prediction, yet easily observable from the text without complex pre-processing. Thus, the following research questions are addressed:

1) To what extent can mathematical formulas that combine both a text characteristic and a target reader's characteristic predict the suitability of a reading text for Algerian middle school EFL learners?

2) To what extent can estimated reading time help predict the suitability of a reading text for the intended readers?

3) How can mathematical formulas help textbook designers and classroom teachers select and/or adapt Algerian middle school EFL textbook reading texts?

This paper reviews the pioneering and most important readability tools that have marked the readability literature with reference to the incompatibility of the existing readability formulas with the Algerian EFL context. It also discusses the methodology and techniques adopted in devising BNP formulas which are, according to McLaughlin (1969: 640), "mathematical equation[s] derived by regression analysis" to find "the equation which best expresses the relationship between two variables." The devised formulas are tested on different sample texts to identify the variables that best correlate with each other, and hence the best equations are adopted. The study ends up by discussing the results and the application of the BNP Formulas by Algerian classroom teachers and textbook writers.

1. Literature Review

We now discuss the concept of readability and previous work on readability prediction, both manually derived, and, more recently, using methods of machine learning and computational linguistics.

1.1 Readability

The concept of readability has been used to refer to the ease with which readers comprehend a reading text (Lorge 1944: 404; Richards et al. 1992: 306). Gérard and Roegiers(2009: 244) define readability as "the measure by which the reader may easily receive the author's message[3]." Dale and Chall(1949) provide a more comprehensive definition to readability by referring to the readers' reading speed and their interest, in addition to their understandability of the text, stating that:

> ... readability is the sum total (including the interactions) of all those elements within a given piece of printed material that affects the success that a group of readers have [sic]with it. The success is the extent to which they understand it, read it at an optimum speed, and find it interesting. (23)

In the same respect, Harmer (2007: 99) maintains that the usefulness of reading "for language acquisition" depends on readers' understanding of the text they read. The difficulty level of a text influences the reading process either positively, when the text is not difficult to process, or negatively, when the text is beyond the level of its reader. Hence, it is a must, as stated by Westwood (2008:35), "to ensure that the difficulty level of the texts they [readers] are required to read is compatible with their current reading ability."

1.2 Readability Formulas

The pioneering work of Sherman (1893) on the objective analysis of linguistic forms of literary works sparked other researchers' enthusiasm to develop scientific tools that transcend human's subjectivity in predicting the readability of a written text. Thorndike's word-frequency lists (1921), and his extended list with Lorge (1944) attempted to classify words according to their difficulty level on the basis of their frequency of use in written prose: the more frequent a word, the easier it is. Lively and Pressy worked on developing a statistical approach using the weighted median index number in Thorndike's list to predict the readability level of textbooks (Lively and Pressy 1923). These works inspired other researchers to develop other predictive methods of text difficulty giving birth to many readability formulas used in different fields such as education, military, publishing, and healthcare.

Gray and Leary (1935: 98-99) list 82 "expressional elements" related to words, sentences, and paragraphs that may indicate the difficulty level of a written text. Since it would be impossible to integrate all these elements, only some significant ones, thought to be better indicators of difficulty level, were selected for devising the different readability metrics, among which 20 formulas have been extensively tested and proved for their feasibility in predicting the readability of texts for different reading contexts. Most of these theorems are premised upon two variables: (1) sentence length and (2) word difficulty. The first variable is represented in the formulas by the average sentence length, which is computed by dividing the total word count by the number of sentences in the text. The second variable takes different forms such as the average word length in characters or syllables, percentage or number of difficult words, percentage or number of monosyllabic words, or the percentage or count of polysyllabic words. Other attempts at capturing vocabulary difficulty are computing the (average) number of syllables for the text or for 100 words or the number of unfamiliar words (which introduces the need to specify a vocabulary of familiar words for the different learner levels).

Vogel and Washburne created the first readability formula that uses the linguistic characteristics of texts as readability variables. They analyzed the correlations of 19 textual elements of 152 books with "the median reading score [a regression score that can be matched to grade level] of the children who read" these books (1928:375). They tried different combinations of elements and compared their correlations to find out that 4 elements correlate best (r = 0.845): number of different words in 1000 words, number of prepositions in 1000 words, number of uncommon words in 1000 words, and number of simple sentences in 75 sample sentences. Using the 4 elements, they developed a regression equation, named Winnetka Formula, to get the reading score (RS) of the evaluated text (Vogel and Washburne 1928). This formula therefore attempts to capture both the vocabulary and syntactic complexity of the reading text. However, the Winnetka Formula was described as complicated, time consuming and unpractical for short texts, thus motivating other researchers in the field to create other simpler and more practical formulas.

Flesch devised his first readability formula in 1943 using three variables: average sentence length, number of affixes, and number of references to people. As stated by Flesch (1948: 221), the "wide application" of the formula by many "academic institutions" encouraged him "to re-examine the formula and to analyze its shortcomings". In addition to sentence length and word complexity, he aimed to capture how interesting the texts might be to a reader. He introduced "two [new]multiple-correlation regression formulas" using the variables of number of syllables and average sentence length in 100-word samples of the assessed reading material for the first formula that predicts the reading ease, and the variables of personal words and personal sentences in 100-word samples for the second formula that assesses the human interest. For the RE formula, according to Flesch, "the longer the words and sentences, the harder to read," while for the HI formula "the more personal words and sentences, the more interesting is the text." The formulas rate texts on a 100-point scale: the higher the score, the easier it is to understand the text and the more interesting it is for the reader (229-230). Each score range refers to an estimated US schooling grade (Flesch 1949: 149-151). This formula was adapted into the Flesch-Kincaid Grade Level Readability Formula by Kincaid and his research group. The formula is still used by the US army and the economic sector in the US (Kincaid et al. 1975).

Dale-Chall Readability Formula was created based on a 769-word list described as familiar to 80% of 4th American graders. Based on the percentage of unfamiliar words in a text, the formula predicts the US school grade in which the target readers will be able to answer at least 50% of the comprehension questions on the evaluated text (Dale and Chall 1948: 41, later updated in Chall and Dale 1995).

A large number of formulas, many in active use today and well-known to practitioners, are based on estimating text complexity by syllable count: This includes the FOG Index (Gunning 1952), the Fry Readability formula (Fry 1968), the SMOG[4] Readability Formula (McLaughlin 1969: 639, 643) or the FORCAST formula (Caylor et al. 1973). More recently, Solomon developed the Strain Index from the FOG index: It needs as input just the syllable count of the first three sentences of the evaluated text (Solomon 2007).

Another interesting expansion of readability prediction is the choice of new target variables. While the vast majority of readability formulas aim at predicting a specific reading age or grade, Nguyen and Uitdenbogerd (2019) predict judgments of text difficulty on a Likert scale, working with Vietnamese learners of English. Finally, Nishikawa et al. (2013) and Weller et al. (2020) predict document reading times. Their motivation is to estimate how long it will take a native speaker (of Japanese and English, respectively) to read news items. While Weller et al. demonstrate the feasibility of predicting average reading times as a dependent variable, Nishikawa et al. are only interested in predicting which of two documents will be read relatively faster.

Despite this large number of proposals for predicting text readability, none of these metrics match the Algerian middle school EFL context for many reasons. First, most of the formulas were primarily devised for non-educational contexts (Caylor et al. 1973; Kincaid et al. 1975). Second, the metrics that were developed for the US schooling grades (Flesch 1948; Gunning 1952; McLaughlin 1969; Chall and Dale 1995) do not match the requirements of the Algerian educational context where English enjoys a foreign language status and is taught starting from the 6th schooling grade. Third, formulas devised for native speakers cannot easily be adapted to language learner environments. For instance, some formulas (Lively and Pressy 1923; Vogel and Washburne 1928; Chall and Dale 1995) depend on a list of words that are considered to be familiar or easy words which “may be viewed as the most elemental words in the English language … these words and their meanings are known without formal schooling” (Chall and Dale 1995: 13), but the same may not hold for Algerian EFL learners. Finally, our goal of providing teachers and text book designers with a technically simple way of assessing not only the readability but also the expected reading time of a text in the non-native context has not yet been addressed.

2. Methodology

This section provides a general overview for the target research population and the study sample. It also deals with the techniques adopted in collecting and analyzing the study data, and includes the training and testing of linear regression models (LRMs) to select the most consistent variables to adopt in BNP Formulas.

3.1 Study Participants and Data Collection

Algerian children start their 5-year primary schooling at the age of 6. They are taught different subjects using the institutional formal language 'standard Arabic', being itself a language to be learned as Algerian children speak, at this age, either the Algerian Dialectal Arabic and/or a Berber variety as their mother tongue. In their primary third year, they start learning French as the first foreign language. At middle school, Algerian learners (aged approximately 11 years) start studying English as the second foreign language. Learners are taught English for four years. 1st and 2nd year learners study English in 2 one-hour sessions a week, while 3rd and 4th year learners have 3 one-hour sessions a week. Additionally, a one-hour tutorial session takes place every week for a number of learners of each class for all levels.

The study data were collected from different middle schools in three Algerian cities, namely Oran, Tiaret, and Relizane. Some practitioner teachers who have taught English for more than 10 years accepted to take part in the study during 2018/19 and 2019/20 school years. Teachers had to select a sample of participants among their learners according to their school achievement in English. The participants were carefully selected to represent the target research population. They were categorized into four categories according to their average score in English out of a maximum of 20 (Category 1: 16-20; Category 2: 13-15.99; Category 3: 10-12.99; Category 4: 0-9.99). First year learners were categorized by their teachers on the basis of their participation, in-class activities and tests during the first trimester of the school year. Teachers were provided with reading time sheets for each level. Texts to be read by learners were selected from the official middle school English textbooks. Each participant read the selected texts during English classes throughout the school year. 34 middle school learners from the four levels (Y1toY4) were selected by their classroom teachers to participate in the study taking into consideration the participants' level, gender, age, and category.

Table1: Study participants' profile by learner level (Y1-4: year of instruction), gender and ability (category 1 is highest ability).

Level		Y1	Y2	Y3	Y4
Gender	**Male**	5	4	4	4
	Female	5	4	4	4
Category	**1**	2	2	2	2
	2	2	2	2	2
	3	3	2	2	2
	4	3	2	2	2
Total		**10**	**8**	**8**	**8**

Table 1 includes information on the investigated sample. Two participants were selected from each category except for the 3rd and 4th categories ofY1 which include 3 participants.

3.2. Reading Texts

The reading texts of the study are selected from the institutional middle school textbooks for two reasons. First, to involve the classroom teachers who participated in the study in the selection process. Teachers were asked to select, at the first stage, a number of reading texts from the four middle school textbooks taking into consideration two main characteristics: (a) length, to use texts of different lengths, and (b) topics, so that the selected texts are of different topics and spread over the textbook units. The final set of selected texts used in the study includes the texts that were selected by most teachers. Second, to get the target readers' reading time in an ordinary classroom setting, where learners read the textbook texts rather than texts on papers. Table (2) includes the word count, syllable count, character count and percentage of polysyllabic words for each sample text.

Table 2: Counts of sample textbook reading texts showing word (W), syllable (SY) and character (C) counts and percentage of polysyllabic words (PSY).

Y1					Y2				
Text	**W**	**SY**	**C**	**PSY%**	**Text**	**W**	**SY**	**C**	**PSY%**
1	31	37	111	3	**1**	165	226	642	4
2	17	21	54	0	**2**	236	318	1000	4
3	55	70	208	7	**3**	249	356	1047	7
4	49	66	193	8	**4**	207	308	924	12
5	66	86	271	2	**5**	96	184	515	24
6	40	63	186	10					
7	168	239	733	10					
8	63	82	249	5					
Mean	**61.1**	**83**	**250.6**	**5.6**	**Mean**	**190.6**	**278.4**	**825.6**	**10.2**
Median	**52**	**68**	**200.5**	**6**	**Median**	**207**	**308**	**924**	**7**
Y3					**Y4**				
1	113	171	487	12	**1**	165	265	811	17
2	130	194	583	9	**2**	285	430	1248	10
3	144	211	601	10	**3**	476	686	2118	11
4	181	265	785	12	**4**	302	507	1506	16
5	215	320	972	11	**5**	526	773	2326	13
6	243	389	1229	12					
Mean	171	258.3	776.2	11	**Mean**	350.8	532.2	1601.8	13.4
Median	162.5	238	693	11.5	**Median**	302	507	1506	13

Table 2 shows that 8 textbook reading texts were read by Y1 participants. Seven texts are less than 100 words long, while all the texts contain less than 10%

of polysyllabic words (0-10%, with a median of 6%). Generally, percentage of polysyllabic words rises as students' progress (Y2 has a median of 7%, Y3 of 11.5% and Y4 of 13%). A notable outlier is one text selected for Y2 participants, which is short at less than 100 words, but contains 24% of polysyllabic words. Text lengths increase drastically from a median of 52 words in Y1 to 207 words in Y2, but Y3 texts are again shorter than Y2 and Y4 texts with a median of 162.5 words. Y4 texts for the most advanced learners are longest at a median of 302 words.

We also show the number of syllables and characters per text. For both syllables and characters, the absolute frequencies follow the same pattern as the number of words per text (very low – large increase - reduction - large increase), but the ratio of average number of syllables by average number of words of course follows the same monotonically rising pattern as the percentage of polysyllabic words. Interestingly, the same is true for the ratio of average number of characters by average number of words. This is an indication that the number of characters in a text might stand in as a predictor for both of these variables.

3.3 Models

The linguistic characteristics (words, syllables, characters and polysyllabic word counts) of the sample texts are used in training linear regression models and analyzing their prediction results. These variables are selected among others for many reasons. First, average sentence length and average word length, which are computed using the word count and character count, are the main variables of the existing readability formulas adopted for their positive correlations with the difficulty level of the text. Second, in addition to their efficiency in predicting the readability level of a text (School Renaissance Institute 2000), they are measurable compared to other non-measurable text characteristics. Third, some readability experts classified English words according to their frequency of use: the less frequent a word, the more difficult it is; thus, most of the less frequent listed words are polysyllabic words (Thorndike 1921; Thorndike and Lorge 1944; Spache 1953). This explains why the count or percentage of syllables or polysyllabic words per textare considered in some readability formulas (Gunning 1952; Chall and Dale 1995; Solomon 2007).

The machine learning software WEKA[5] was used to develop an overall LRM for each learner year. The training data included the word counts, syllable counts, and character counts of the sample texts. The data were grouped according to the learner years. Available target variables were the reading time in seconds (RTS) and, derived from it, the average reading time per word (ARTW: calculated by dividing the RTS by the word count), average reading time per syllable (ARTSY: calculated by dividing the RTS by the syllable count), and average reading time per character (ARTC: calculated by dividing the RTS by the word count), for each data set. The models were developed using the three variables of word count, syllable count and character count and their corresponding average reading time (ART), to compare the differences in estimated reading time. The ARTs of all participants are averaged again across the participants for each category and level. The category-overall ART and level-overall ART are used in models for prediction. The models were trained on a 10-fold cross-validation mode and were evaluated using the correlation coefficient (Pearson's *r*)[6] as well as the root mean squared error (RMSE)[7] between observed and predicted reading times.

3. Experiments

Our first experiment addresses the core of our goal, prediction of observed reading time. We train linear models based on the number of words, syllables and characters specifically for each year of instruction. We then verify the plausibility of the selected model by generating predictions for new texts that have the same number of words, but differ in the percentage of polysyllabic words. Based on the assumption that longer words are harder to read, we expect to see longer reading times for texts with more polysyllabic words and verify that the model predicts this pattern. Finally, we show that we need prediction models that are specific to each learner ability cohort in order to accurately estimate the time needed by a mixed-ability group to read a text in the classroom. The sample texts used in experiment 2 were chosen randomly from different online sources taking into consideration their lengths and topics that learners might be acquainted to.

3.1 Experiment 1

Table 3: Linear regression models for the four learner years based on words, characters and syllables. Pearson's r and RMSE. N: number of training instances. ERT: Estimated Reading Time.

Level	N	V1	V2-Mean		Overall LRMs	*r*	RMSE
Y1	**80**	W	ARTW	1.952	ERT= (2.4245 x W) + (54.2911 x ARTW) −122.6929	0.96	**33**
		SY	ARTSY	1.458	ERT= (1.7345 x SY) + (72.8195 x ARTSY) −118.6968	0.96	33.06
		C	ARTC	0.492	ERT= (0.5764 x C) + (210.8408 x ARTC) −116.8006	0,96	33.66
Y2	**40**	W	ARTW	1.681	ERT= (1.7783 x W) + (161.7829 x ARTW) −303.1264	0.94	33.22
		SY	ARTSY	1.111	ERT= (1.0985 x SY) + (273.0461 x ARTSY) −301.3609	0.98	**17.76**
		C	ARTC	0.378	ERT= (0.3804 x C) + (792.4276 x ARTC) −305.6891	0,97	20.12
Y3	**48**	W	ARTW	1.369	ERT= (1.3444 x W) + (156.4254 x ARTW) − 212.9387	0.97	**17.64**
		SY	ARTSY	0.911	ERT=(0.8608 x SY) + (233.8766x ARTSY) −204.3972	0.97	18.75
		C	ARTC	0.307	ERT= (0.2787 x C) + (683.6585 x ARTC) −195.1134	0.96	21.19
Y4	**40**	W	ARTW	1.275	ERT= (1.234 x W) + (334.5605 x ARTW) − 428.6095	0.96	40.20
		SY	ARTSY	0.827	ERT= (0.8168 x SY) + (514. 8797x ARTSY) −429.5547	0.97	**36.22**
		C	ARTC	0.275	ERT= (0.2686 x C) + (1531. 7886x ARTC) −421.2003	0.97	36.75

Our first experiment compares the appropriateness of the three simple predictor variables number of words, number of syllables and number of characters for the task of predicting observed reading times. We train separate linear regression models for each of the four years of instruction, since reading speed in learners depends on ability in the foreign language. Table 3 shows the number of training instances per year. For each year of instruction, we train three separate models: One based on the number characters in the text, one on the number of words and one on the number of syllables.

The results show very good positive correlations among the variables of each overall LRM (*r* = 0.94-0.98). In two cases, the word-based models show the lowest RMSE, in two cases, the syllable-based methods have the lowest error. When the syllable-based methods do not yield the lowest RMSE, they are numerically very close to the best performance.

Lorge (1944) classifies the polysyllabic word count/percentage in a text as one of the measures of the variable of "vocabulary load" in a text "used as a predictor in every study of readability" (405), which explains the reason why it is used in some readability formulas with other variables (Gunning 1952; Kincaid et al. 1975; Klare 1975) or as a sole variable (McLaughlin 1969; Solomon 2007). Therefore, it is not surprising that the syllable-based measure should perform well. However, the number of syllables in a text is harder to determine in practice than the number of words or characters and is therefore not the ideal measure for a robust, easy-to-use tool for teachers and school book designers.

4.2 Experiment 2

Experiment 1 showed that we can train LRMs on the aggregated reading time data for each year of instruction and reliably predict the reading times for specific texts. We still need to differentiate between the three input variables and identify the one that is most useful for robust automated predictions in the field.

Hence, we analyze the overall LRM predictions for texts of different lengths and different polysyllabic word percentages to (1) study the effect of polysyllabic word count in a text on the plausibility of the overall LRMs in providing estimated reading times, and (2) identify the best variables to adopt among the three variables ARTW, ARTSY, and ARTC.

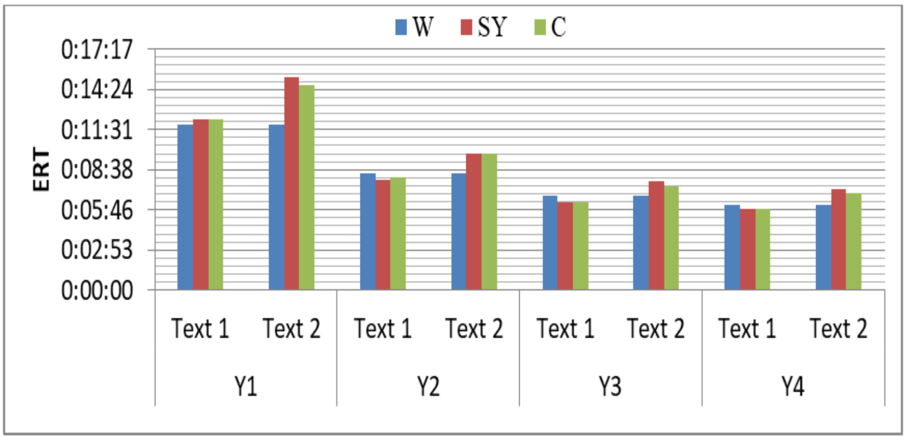
As illustrated in Table 4, two sample texts (STs) of the same word count and different polysyllabic word-percentages, and subsequently, different syllable counts and character counts, are selected to test the overall LRMs to examine the effect of the variable of polysyllabic word count in their plausibility. Note that we do not have observed reading times for these texts; we are interested in the match between hypothesized patterns and model predictions to test the models' plausibility. We expect that the word-based models will make the same reading time predictions for Text 1 and Text 2, and that the predictions of the syllable-based model will differ between the two texts. We also expect that the predictions of the character-based models will differ between the texts, since polysyllabic words are longer and contain more characters than monosyllabic words. We do not know, however, how large the gap in predicted reading times will be between the syllable- and character-based models.

The testing results are presented in Figure 1.

Table 4: Sample texts with different amounts of polysyllabic words.

	Text1	Text2
% polysyllabic words	7	22
Word count	300	300
Syllable count	432	535
Character count	1295	1556

Figure 1: Plausibility check of the year-specific LRMs for sample texts with different amounts of polysyllabic words. ERT: expected reading time; W: word, SY: syllable, C: character.



The results in Figure1 show at first glance that the four LRMs capture increasing reading speeds for advanced learners. Further, as expected, the word-based models predict the same reading times for both texts in all cases. In contrast, an increase for the estimated reading time per syllable as well as the estimated reading time per character is noticed for Text 2 as compared to Text 1 for all models. The estimated reading times for Text 1 show a gap of no more than ½ minute between the word-based, syllable-based and character-based models – all models predict essentially the same reading time. Compare this to a gap of up to 4 minutes for Text 2, between the word-based models on one side and the syllable- and character-based models on the other side. The figure also shows a high similarity of the estimated reading times from the syllable- and character-based models, especially for Text 1.Note that we cannot determine how well these predictions fit reality due to the lack of empirical reading time data for Texts 1 and 2; we can however conclude that the error will be very similar for both models.

The experimental results confirm the expected result that a longer reading time is estimated for a text with a higher polysyllabic word count. In contrast, the syllable-based and character-based overall LRMs provide different estimated reading times for texts of different polysyllabic word counts and the gaps between their estimated reading times are very low. Our Experiment 1 indicated that syllable-based models are robust and accurate. Experiment 2 confirms that for unseen texts, the gap in predicted reading times for syllable- and character-based models is very small. In addition, determining the number of characters in a text is much simpler than determining the number of syllables. This is relevant to our goal of providing robustly automated reading time predictions for new texts to serve teachers in the field. Therefore, the focus on the next training phases will be on the character-based models.

4.3 Experiment 3

We now turn to training individual LRMs not only for each year, but also for each learner category in each year. The number of instances available for training shrinks, in this case, but we gain more fine-grained models of the reading times at the high and low ends, which are consistently underestimated by the overall model, as shown above.

Table 5: Character-based individual LRMs for years of instruction and categories of ability. Ca: Category. N: Number of instances.

Level	Ca	N	ARTC	Individual LRMs	r	RMSE
Y1	1	16	0.249	ERT = (0.3463 x C) + (185.4177 x ARTC)– 62.0577	0,98	11.98
	2	16	0.386	ERT = (0.4967 x C) + (223.3422 x ARTC)– 101.3559	0,95	32.99
	3	24	0.528	ERT = (0.6489 x C) + (195.7637 x ARTC)– 120.8358	0.99	10.83
	4	24	0.690	ERT = (0.7165 x C) + (172.8332 x ARTC)– 125.9143	0.99	**6.75**
Y2	1	10	0.30	ERT = (0.3135 x C) + (902.8845 x ARTC)– 285.855	0.99	5.24
	2	10	0.329	ERT = (0.3396 x C) + (956.5228 x ARTC) – 326.4739	0.99	**2.83**
	3	10	0.378	ERT = (0.3797 x C) + (857.5144 x ARTC)– 328.6899	0.99	7.6
	4	10	0.505	ERT = (0.5007 x C) + (703.686 x ARTC)– 358.5641	0.98	12.25
Y3	1	12	0.204	ERT = (0.1941 x C) + (671.9616 x ARTC)– 131.0749	0.99	**4.21**
	2	12	0.263	ERT = (0.2432 x C) + (712.2978 x ARTC)– 177.29	0.98	8.47
	3	12	0.34	ERT = (0.2965 x C) + (628.5916 x ARTC)– 189.2485	0.97	11.6
	4	12	0.42	ERT = (0.3524 x C) + (563.8412 x ARTC)– 196.8003	0.98	11
Y4	1	10	0.206	ERT = (0.2003 x C) + (51421.0975 x ARTC)– 290.0477	0.99	5.58
	2	10	0.24	ERT = (0.2289 x C) + (1389.9964 x ARTC)– 325.0492	0.99	8.07
	3	10	0.298	ERT = (0.2842 x C) + (1471.1005 x ARTC)– 431.9652	0.99	10.94
	4	10	0.357	ERT = (0.3445 x C) + (1315.1448 x ARTC)– 459.0464	0.99	**4.99**

Table 5 demonstrates that character-based individual LRMs were developed for each category of years of instruction. The results show a very good positive correlation among the LRMs variables (*r* = 0.95-0.99) and RMSEs are also very low. Note that with the small number of training instances available, model overfitting is a valid concern. Therefore, we analyze the over LRMs for each year of instruction and the category-specific LRMs for 3 texts of different lengths to check their plausibility in providing approximate estimated reading times.

Figure 2: Overall and category-specific estimated reading times for 100/200/300-word sample texts for the category-specific models and the overall LRM for the year.

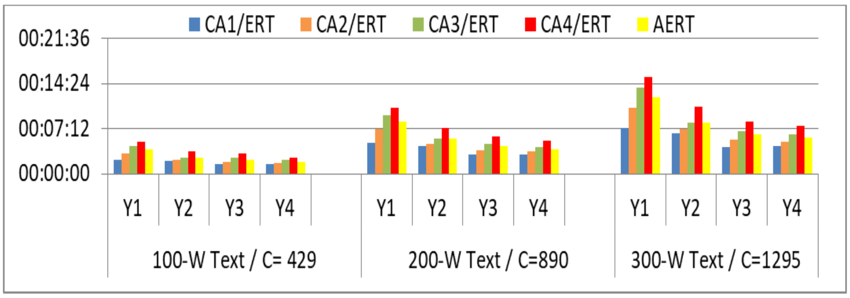


Figure 2 illustrates that the lowest estimated reading times among categories are noticed for category 1 while the highest are for category 4 of all years. This shows that on new texts, the suite of category-specific models correctly captures the differences in student reading times caused by differences in their ability to read and understand English. This is despite the small number of data points available for training.

Additionally, we see that the overall LRM underestimates the reading times for the weakest learners in each year (category 4) by 2-3 minutes while overestimating the reading times of the ablest learners in category 1 across all years of instruction. This mirrors the picture shown in Figures 2-5 above by plotting observed reading times and the overall model's predictions. To further illustrate this point, we show the estimated reading times for the ability categories as predicted by the individual models.

We therefore conclude that the character based LRMs for each year of instruction and category are the best to adopt as BNP Formulas for estimating the reading time of texts for MS learners. First, the overall LRMs were confirmed to underestimate the estimated reading times for low-ability and overestimate the estimated reading times for high-ability readers. In addition, the estimated reading time provided by word-based LRMs is not affected by a change in the polysyllabic word count in a text despite the effect this variable has on the text difficulty. A significant increase in estimated reading time provided by the syllable- and character-based for texts with higher polysyllabic word-count is observed compared to texts of the same word count with low polysyllabic word-count. Furthermore, very good positive correlations are noticed for character and syllable-based LRMs with low RMSEs. Finally, the estimated reading times predicted by the character and syllable-based LRMs are quite similar. Thus, it is easier and more accurate to compute the character count of the evaluated text on electronic devices, which will be used in the application of the formulas, than the syllable count.

4. Conclusion

Algerian middle school EFL teachers are required to set, in their lesson plans, the estimated time for each classroom activity. However, it has been noticed that most teachers find difficulties in estimating the reading time that their learners need to read a text due to the differences in learners' reading abilities. The reading speed, therefore, was thought to be a good reader's characteristic that can be adopted in devising a set of formulas that estimate the reading time of a text and which will help classroom teachers plan their lessons and select the appropriate reading texts for their learners. The second variable, i.e., the linguistic characteristic of the assessed text, was identified by training LRMs in WEKA software using the variables which have been described by readability experts as good readability indicators, mainly the word count, syllable count and character count.

WEKA training and prediction results demonstrated that the overall LRMs predictions are too low compared to low-ability readers and too high compared to high-ability readers, which induced the development of individual LRMs for each MSL category. In addition, very high positive correlations and low RMSEs are noticed for the character-based LRMs. Furthermore, a consistency is observed through the estimated reading times of the character and syllable-based LRMs despite the polysyllabic word-count differences in the sample texts compared to the estimated reading times predicted by the word-based LRMs. Moreover, the character-based LRMs estimate quite similar reading times to the ones estimated by the syllable-based LRMs. Besides, variance in estimated reading times across the MSLs and their categories confirm that the higher the MSL, the longer the estimated reading time is; and the lower the target reader's level of English, the longer the estimated reading time is. Hence, the character count, as a linguistic characteristic of the evaluated text, and the average reading time per character, as a target reader's characteristic, are proved to be the best variables to adopt in BNP formulas to get the approximate estimated reading times of texts for the Algerian middle school EFL learners.

A website is planned to be published to facilitate the use of BNP formulas by Algerian classroom teachers and textbook writers when selecting and/or adapting appropriate texts that match the Algerian middle school EFL requirements, mainly time management. The formulas can also help textbook writers select reading texts to maintain gradation and consistency through MSLs and their textbooks. Additionally, classroom teachers can use BNP formulas to select texts for both intensive and extensive reading sessions compared to other formulas that were proved to be efficient just for long texts that are not appropriate for intensive reading classes. However, it should be noted that users of BNP formulas should take into consideration other non-measurable characteristics, such as sentence structure, abstractness and coherence, when selecting and/or adapting a reading text.

References

1BEM is an acronym for Brevet d'Enseignement Moyen (Middle School Certificate). By the end of their middle school 4th year, Algerian learners sit for the national BEM examination which determines whether they will be able to proceed to secondary school.

2 Our translation.

3 Our translation.

4 SMOG stands for Simple Measure of Gobbledygook.

5 WEKA was developed by the Machine Learning group at the University of Waikato and is available free of charge at
https://www.cs.waikato.ac.nz/ml/weka/.

6 A measure of how well sets of data are related.

7 The square root of the average squared error of the regression. It measures the overall accuracy of the trained model to compare it to other trained models.

1. **Books**

- Caylor, J. S., et al. (1973), Methodologies for Determining Reading Requirements of Military Occupational Specialties, Human Resources Research Organization, Virginia, USA.
- Chall, J. S., and Dale, E. (1995), Readability revisited: The new Dale-Chall readability formula, Brookline Books, Massachusetts, USA.
- Flesch, R. (1949), The Art of Readable Writing, Harper and Row, New York, USA.
- Gérard, F. M. and Roegiers, X. (2009), Des manuels scolaires pour apprendre: concevoir, évaluer, utiliser, De Boeck Université, Bruxelles, Belgium.
- Gray, W. S. and Leary, B. E. (1935), What Makes a Book Readable, Chicago University Press, Chicago, USA.
- Gunning, R. (1952), The Technique of Clear Writing, McGraw-Hill, New York, USA.
- Harmer, J. (2007), How to teach English, Pearson Longman, Essex, UK.
- Richards, J. C., Platt, J. and Platt, H. (1992), Longman dictionary of language teaching and applied linguistics, Longman, London, UK.
- School Renaissance Institute (2000), The ATOS Readability Formula for Books and how it Compares to other Formulas, School Renaissance Institute, Wisconsin, USA.
- Sherman, L. A. (1893), Analytics of Literature: A Manual for the Objective Study of English Prose and Poetry, Ginn & Company, Boston, USA.

- Thorndike, E. L. (1921), The Teacher's Word Book, Columbia University, New York, USA.
- Thorndike, E. L. and Lorge, I. (1944). The Teacher's Word Book of 30,000 Words, Columbia University, New York, USA.
- Westwood, P. (2008), What teachers need to know about Reading and writing difficulties, Australian Council for Educational Research, Victoria, Australia.

2. **Articles**

- Dale, E., and Chall, J. S. (1948), A Formula for Predicting Readability: Instructions, Educational Research Bulletin, 27(2), 37-54.
- Dale, E., and Chall, J. S. (1949), The Concept of Readability, Elementary English, 26(1), 19-26.
- Flesch, R. (1948), A New Readability Yardstick, Journal of Applied Psychology, 32(3), 221-233.
- Fry, E. (1968), A Readability Formula that Saves Time, Journal of Reading, 11, 513-516.
- Kevyn, C. T. (2014), Computational assessment of text readability: A survey of current and future research, ITL-International Journal of Applied Linguistics, 165(2), 97–135.
- Kincaid, J. P., et al. (1975), Derivation of New Readability Formulas (Automated Readability Index, Fog count, and Flesch Reading Ease Formula) for Navy Enlisted Personnel, Research Branch Report. 8-75. Memphis: Chief of Naval Technical Training: Naval Air Station.
- Lorge, I. (1944), Predicting readability, Teachers College Record, 45, 404-419.
- Lively, B. A. and Pressey, S. L. (1923), A method for measuring the 'vocabulary burden' of textbooks, Educational administration and supervision, 9, 389-398.
- McLaughlin, G. (1969), SMOG Grading: A New Readability Formula, Journal of Reading, 12(8), 639-646.
- Nguyen, T., and Uitdenbogerd, L. A. (2019), Measuring English Readability for Vietnamese Speakers, Proceedings of The 17th Annual Workshop of the Australasian Language Technology Association.
- Nishikawa, H., et.al. (2013), A Pilot Study on Readability Prediction with Reading Time, Proceedings of the 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations, pages 78–84, Sofia, Bulgaria, August 4-9
- Solomon, N. W. (2007), Strain Index: A new readability formula, Journalism Online Newsletter, Retrieved on 13/06/2019
- from www.ogg.osu.edu/media/documents/health_lit/STRAIN% 20INDEX.pdf

- Spache, G. (1953), A New Readability Formula for Primary-grade Reading Materials, Elementary School Journal, 53, 410-413.
- Vogel, M., and Washburne, C. (1928), An Objective Method of Determining Grade Placement of Children's Reading Material, The Elementary School Journal, 28(5), 373-381.
- Weller, O., et al. (2020), You Don't Have Time to Read This: An Exploration of Document Reading Time, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 1789-1794.