SPIKE: Online Reading Competencies Measure for Malay Language

Nur Amalina Mohamad Hazawawi, Mohd Hafiz Zakaria, Syariffanor Hisham
Creative Media Lab, Faculty of Information and Communication Technology,
Universiti Teknikal Malaysia Melaka, Melaka, Malaysia.
hafiz@utem.edu.my

Abstract— **The study presented in this paper aims to propose an online tool to measure the readability level of Malay reading materials and to suggest the suitability of the text material based on the reader's age. The motivation of this study was conceived due to the fact that there is no available tool that can measure the readability level of Malay texts. The development of this tool named Sistem Penilaian Kebolehbacaan Bahasa Melayu (SPIKE) is a part of a continuous work done to evaluate readability competencies among young adults with dyslexia. Several articles from Malay newspaper and education magazines were sampled, then analysed to match the suggested level of reader competencies from the highest to the lowest readability level. Results show that SPIKE can benefit general public particularly people with reading difficulties including dyslexics in measuring their reading competencies and to check whether a reading material is suited for their age. For future works, it is expected that this tool can help to increase the accuracy of measuring Malay texts and may be beneficial for authors or content providers to assess the suitability of their work for targeted readers.**

Index Terms— **Readability test tool; Reading difficulty; Readability level; Readability.**

I. INTRODUCTION

The Malay language readability test tool named Sistem Penilaian Kebolehbacaan Bahasa Melayu (SPIKE) is developed to assist readers in knowing the score of a Malay text and acquire the reading level and grade level based on several readability formulas [1]. Two commonly used readability tool for checking the readability of a document is in Microsoft Word is using Flesch Reading Ease and Flesch Kincaid Grade Level which can be accessible from Spelling and Grammar check tool [2].

The existing readability test tools are made primarily for English text with options either typing in, pasting or providing a webpage link. The scores are calculated based on several readability formula like Flesch Kincaid Grade Level and SMOG Index. Text parameters such as character count, syllable count, word count, sentence count, characters per word, syllables per word and words per sentence are used for calculating score. Each readability scoring formula would vary slightly as the calculation is constructed based on average words per sentence and average syllables per words [1] [2] [3].

The main focus of this paper is to discuss the development of an online tool that can measure the readability level of reading materials to help readers with reading difficulties. In addition, the paper also aims to discuss the readability level of printed materials and provide some insight of the suitability of reading materials available for public according to their age.

II. READING COMPETENCIES IN MALAYSIA

Competency is the ability to apply or utilize a set of related knowledge and skills required to successfully perform 'critical work functions' or tasks in a characterized work setting. In reading competency, we can define it as the practical skills and knowledge that empower individual to read very well [5]. Reading has been defined as the ability to interpret an arrangement of entrenched images (symbols), for example, Latin letter arranged in a meaningful sequence in order to derive meaning. Reading is one of the four basic languages ability that an individual will develop at an early phase of life [6].

The Star Online Newspaper reported that Malaysians prefer to read light reading materials like newspaper and magazines instead of books. It was also reported in a literacy statistics that out of 85% of Malaysians who read frequently, 77% of them choose to read newspapers, 3% magazines, 3% books and 1.6% comics. [9]. This statistics also consistent with Annamalai and Muniandy (2013), based on 119 participants which are from Polytechnic students that majority of Malaysians tends to read for entertainment purpose; read newspaper every day for at least a few times a week (68.9%), magazines (57.1%), websites (56.3%), novels (44.5%), academic books (27.1%), comic and journals [10].

A study in 2011 was conducted to find the correlation between family income and reading motivation level. The results showed there is no significant relationship between students reading motivation level and students economic background [11]. This results was supported by a study conducted in 2014, where output of the study showed no relationship between socio-economic status and online reading habits [12].

People with dyslexia in a multicultural background, particularly Malaysia, display more challenges in reading, writing and spelling since they are likely to learn more than a language [7]. Dyslexics especially students find printed data hard to understand and this might resulted in low academic performance [8][22][23].

III. READABILITY FORMULA IN MALAY LANGUAGE

Readability, comes from two words read and ability which makes an interpretation of into the ability to read. Readability has nothing to do with either legibility or literacy. Readability formula, is a mathematical formula, which measures the grade level a person must have to read and understand a text [1]. Nielsen (2015), stated that readability is used to measure the difficulty of the words and structure of sentences in a text [13].

Existing readability test tools provide a quick and easy way to test the readability of an article, besides displaying text information including number of characters, words, sentences, and average number of characters per word, syllables per word, and words per sentence [26]. Readability formulas are constructed from two basic parameters which are average words per sentence and average syllables per words. This is agreed by Doverspike (2015), where readability formulas are normally developed on simple variables such as the number of syllables, number of words and sentence length [14][15].

Hazawawi, Zakaria & Hisham (2015) found that, words that have more than three syllables and potential difficult words in Malay could be contributing factors in reading competencies, which can affect the time taken to read with the presence of those additional parameters [26] [27].

There is a mathematical readability formula in Malay language which was proposed by Khadijah Rohani in 1984, using two parameters implying on the number of words in sentences (construction of words) and the number of syllables (vocabulary) [16][25]. Below is the mathematical formula in Bahasa Melayu:

$$y_i = a + bn_i + cd_i \qquad (1)$$

where:

i	=	is the sample in the case study
n_i	=	300/S, number of words in sentence. S is the number of sentence in 300 words.
d_i	=	the number of syllables
a	=	-13.988
b	=	0.3793
c	=	0.0207

From this, we proposed an expansion of the readability formula in Malay language for people with reading difficulties. Below is the revised mathematical formula which can also assist dyslexic people for measuring the readability of a text in Malay language:

$$y_i = a + bn_i + c(d_i + k_i) \qquad (2)$$

where:

i	=	is the sample in the case study
n_i	=	300/S, number of words in sentence. S is the number of sentence in 300 words.
d_i	=	the number of syllables

k_i	=	the potential difficult words
a	=	-13.988
b	=	0.3793
c	=	0.0207

Potential difficult words in our scope are Kata Ganda,Diftong, Kata Pinjaman and Kekeliruan huruf. The k_i is calculated as below:

(Kata Ganda+Diftong+Kata Pinjaman+Kekeliruan Huruf)*5

The potential of difficult words for dyslexia is based on a study that was conducted on dyslexic's spelling aspect in Malay language, where Kata Ganda, Kekeliruan Huruf and influence of English language (Kata Pinjaman) are the main mistakes that dyslexic people tend to do when reading a text [21].

The addition of parameter = 5, is supported by a research in 1999 by University of Washington (UW), where dyslexic children tend to use nearly five times the brain area as normal children while carrying out a simple language task. This means that their brain required to work vigorously and utilize more energy than normal children. The results from the study also illustrates for the first time that there are chemical differences in the brain function of children suffering dyslexia and non-dyslexic children [17][18].

IV. READABILITY TEST TOOLS

The readability test tools are designed to quickly measure the readability level of a passage. All readability formulas work by calculating the variables that have the greatest effect on people having the ability in writing: sentence length, number of syllables per words and number of passive sentences. However, need to keep in mind that the score that we get does not tell us everything [19].

The propose tool is developed to cater for normal and dyslexic people where both readability formulas that we had mentioned earlier are implemented in the tool. The tool is based on construction of words in Malay language, such as how the syllables are count and what types of sentences will be omit in calculating the readability level. For dyslexic purpose, there are some addition features like number of potential difficult and the indicator of the level.

The developed tool is inspired by an existing readability test tool (Readability-Score.com) [20], based on some features and advantages of this tool compared to other readability test tool. Compared to other readability tools, Readability-Score.com website is flexible and easy way to test the readability of text. In addition to that, the interfaces of the website are simple even the novice user can use it comfortably. The tool will be developed in Malay language based on construction of Malay words including syllables and words count that are totally different from construction of English language.

V. EXPERIMENT PROCEDURE

In order to develop this tool, we had conducted a field study to capture the readability level of printed materials particularly

newspaper and magazines. The objectives of this experiment is to identify what readability level and suitable for people starting from what aged.
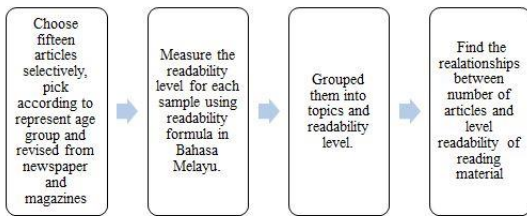

Figure 1: Block diagram of Process in Selecting and Measuring the Readability Level of Reading Sample

Block diagram in Figure 1 illustrates the process in selecting and measuring the readability level of reading sample in the study. Fifteen articles were chosen selectively from newspaper and magazines and measured using readability formula in Bahasa Melayu by Khadijah Rohani [16]. After that, we will grouped them into topics and readability level. Lastly, after we get all the data, we will compare them with readability level for each sample to see the relationship between number of articles and level readability of reading material.

VI. RESULTS

All selected articles are being selected and analyzed as part of the experiment. Table 1 shows the readability level for each reading material sampled according to their respective topics, which are Umum, Sukan, Ekonomi, Kesihatan, Rekreasi dan Pelancongan, and Sains dan Teknologi. The readability level for each reading material sample are range from readability level 7 to readability level 11. The lowest readability level is 7.6201 (Bakteria Musuh atau Sahabat) from Dewan Masyarakat and the highest readability level is 11.6563 (Tubuh Mahkamah Alam Sekitar) from newspaper Utusan. Based on the 15 articles gathered from Dewan Bahasa dan Pustaka and Berita Harian we found that public reading are scored from readability level 7 and above where it is suitable for people aged 13 years of age and above. Table 1 shows the list of articles corresponding with their readability level and topics.

When choosing and revising the material samples, we do not take into account language aspects that do not belong directly to the understanding of a sentence. This includes but not limited to scientific language and KOMSAS (*komponen sastera*) including *pantun, gurindam, seloka, syair* and *sajak*. We excluded all these components when measuring the difficulty level of text.

Table 2 shows the number of articles based on their readability levels.

Table 3 shows the readability levels that had been improved based on the existing readability level table ranging from Level 1 Level 11 that suits the academic level of Year 1 until Form 5 at school level. Should the score falls in decimal place (e.g. 1.9), then the readability level indicates that the material is stronger towards Year 2 level [16]. Level 12 and above is added to match the adult reader competencies aged 18 years

old and above. The text indicator is proposed for people with reading difficulty, where the text indicator can be shifted according to the differences between these two formulas. Each indicators indicates different type of text:

- Red - this text is difficult for people with reading difficulty
- Yellow - this text is quite difficult for people with reading difficulty
- Green - this text is suitable for all people

Table 1
The Readability Level for Each Reading Material Sample with Topics

Title	Level	Topics
Bakteria Musuh Atau Sahabat? (Dewan Masyarakat)	7.6201	Kesihatan
Misteri Lohong Hitam (Dewan Kosmik)	7.8779	Sains dan Teknologi
Jerebu (Berita Harian)	8.1071	Umum
Kediaman Kedua (Dewan Kosmik)	8.3126	Sains dan Teknologi
Bergadai Hutang Demi Anak (Dewan Masyarakat)	8.9574	Umum
Kek Lapis Sarawak Ikon Pelancongan (Berita Harian)	9.4335	Rekreasi dan Pelancongan
Cantik Tapi Terabai (Berita Harian)	9.5753	Rekreasi dan Pelancongan
Percutian Damai di Ladang Teh (Berita Harian)	9.7026	Rekreasi dan Pelancongan
Selamatkah Anak Kita? (Dewan Masyarakat)	9.7647	Umum
Muflis Muda (Dewan Masyarakat)	9.7830	Umum
Bajet 2016: Kurangkan Beban Hutang (Utusan)	10.0772	Umum
Sultan Nazrin Pelawa Maghribi Guna Kepakaran Malaysia dalam Kewangan Islam (Utusan)	10.5533	Ekonomi
Bangkitlah Bola Sepak Malaysia (Dewan Masyarakat)	10.8655	Sukan
Bantu Isteri Lawan Kanser (Utusan)	11.3209	Kesihatan
Tubuh Mahkamah Alam Sekitar (Utusan)	11.6563	Umum

Table 2
Number of Articles Based on Readability Level of Reading Sample

Readability Level of Reading Samples	Number of Articles
Level 7	2
Level 8	3
Level 9	5
Level 10	3
Level 11	2

Figure 2 shows information of text when user posted an article in Malay language in the space provided. It will provide number of total words, number of characters without spacing and number of characters with spacing.

SPIKE provides two types of readability levels: 1) Readability level for normal user and 2) Readability lever for dyslexics. All information about the reading text will be processed and appeared once the user clicks the 'Process' button. Both results of normal and dyslexia readability will be shown at the right side of the web. Aside of it, there is an indicator to indicate the difficulty of text.

Table 3
Readability level according to Year in Primary School and Form in Secondary School

Readability Level	Year/Form	Age	Text Indicator
1	Year 1	7 years old	RED
2	Year 2	8 years old	
3	Year 3	9 years old	
4	Year 4	10 years old	
5	Year 5	11 years old	YELLOW
6	Year 6	12 years old	
7	Form1	13 years old	
8	Form 2	14 years old	
9	Form 3	15 years old	GREEN
10	Form 4	16 years old	
11	Form 5	17 years old	
12 and above	Adults	18 years old and above	

Figure 2: Information of text

Figure 3 shows two types of results of readability level for normal and dyslexia. Readability level for normal considered two types of parameters which are number of words per sentences and total number of syllables

For dyslexia readability levels, there is an additional parameter used in calculation known as potential difficult words consist of *Kata Ganda, Diftong, Kata Pinjaman* and *Kekeliruan Huruf,* which can affect dyslexics when they read.

Figure 4 shows an addition features compared to the existing readability test tool. The indicator displays the type of text (easy, moderate, or difficult) and there is also a colour indicator that will change according to the readability level for normal or dyslexia readers.

VII. CONCLUSION

The readability concept has been discussed over the past decades, and until today only English readability test tool was developed targeting English native speakers. Data has shown that this tool may give inaccurate scores for other languages due to the differences in how the words are constructed. With the development of SPIKE, it is hoped that it can assist general public to better understand their reading proficiencies

particularly for people with reading difficulties. Hence, it is important now to have a readability test tool that can measure the readability level in Malay text.

Figure 3: Readability level for normal readers and for dyslexics

Figure 4: Colour indicator

REFERENCES

[1] B. Scott. Readability Formulas. [Online]. From: http://www.readabilityformulas.com/ [Accessed on 7 December 2015]

[2] Readability Statistics Feature in Microsoft Word 2013. The Windows Club [Online]. From: http://www.thewindowsclub.com/readability-statistics-feature-microsoft-word [Accessed on 20 January 2016]

[3] R. Rambo. 2013. The Reading Level of Your Writing. [Online]. From: http://www2.ivcc.edu/rambo/eng1001/reading_level.htm [Accessed on 6 December 2015]

[4] J. McGee.2010 . TOOLKIT for Making Written Material Clear and Effective SECTION 4: Special topics for writing and design. Part 7: Using readability formulas: A cautionary note. [Online]. From: https://www.cms.gov/outreach-and-education/outreach/writtenmaterialstoolkit/downloads/toolkitpart07.pdf [Accessed on 7 December 2015]

[5] Bloom, Taxonomy of Educational Objectives. Handbook I; Dave, Developing and Writing Behavioral Objectives; and Krathwohl, Bloom, and Masia, Taxonomy of Educational Objectives. Handbook II. 2010. Competencies and Learning Objectives. [Online]. From: https://sph.uth.edu/content/uploads/2012/01/Competencies-and-Learning-Objectives.pdf [Accessed on 7 December 2015]

[6] A. K. Mozihim. 2010. On Reading Habits of Malaysian Students. [Online]. From: http://www.dailyexpress.com.my/read.cfm?NewsID=370 [Accessed on 6 December 2015]

[7] C. T. B. Sung. 2013. My Bella TV Interview: Malaysia's Falling Proficiency in English- Should We Be Worried? [Online]. From: http://www.christopherteh.com/blog/2013/12/bella-english/ [Accessed on 6 December 2015]

[8] N.A.M. Hazawawi & S.Hisham. 2014. Online Dyslexia Screening Test for Malaysian Young Adults in Bahasa Melayu. Information and Communication Technology for the Muslims World (ICT4M), 2014. Kuching, Sarawak.

[9] The Star & The Malaysian Bar. 2009. Malaysians Still Not Reading Enough, Says Muhyiddin. [Online]. From: http://www.malaysianbar.org.my/general_opinions/comments/malaysians_still_not_reading_enough_says_muhyiddin.html [Accessed on 6 December 2015]

[10] S. Annamalai & B. Muniandy. 2013. Reading Habits And Attitude Among Malaysian Polytechnic Students. International Online Journal of Educational Sciences. Vol. 5, No. 1, pp. 32-41.

[11] B. Sani, M.N.W.Chik, Y. A. Nik & N.A. Raslee. 2011. The Reading Motivation and Reading Strategies Used By Undergraduates in Universiti Teknologi MARA Dungun, Terengganu. Journal of Language Teaching and Reseacrh. Vol. 2, No. 1, pp. 32-39.

[12] M.J.Z. Abidin, M. Pourmohammadi, N. A/P Varasingan & O.C. Lean. 2014. The Online Reading Habits of Malaysian Students. The Reading Matrix. Vol. 14, No. 2.

[13] J. Nielsen. 2015. Legibility, Readability and Comprehension: Making Users Read Your Words. [Online]. From: http://www.nngroup.com/articles/legibility-readability-comprehension/ [Accessed on 7 December 2015]

[14] C. Stephens. 2009. All About Readability. [Online]. From: www.plainlanguagenetwork.org/stephens/readability.html [Accessed on 7 December 2015]

[15] D. Doverspike. 2015. Readability of Assessments in a Digital Age (Part 1): Bet You Won't Read This Whole Blog. [Online]. From: http://asr.ipma-hr.org/2015/03/readability-of-assessments-in-a-digital-age-part-1-bet-you-wont-read-this-whole-blog/

[16] K. R. M. Yunus. 1982. An Assessment of Structural Variable In Malay: A Readability Formula. Unpublished manuscript, University of Miami, United States.

[17] J. Schwarz. 1999. Dyslexic Children Use Nearly Five Times The Brain Area. University Of Washington. [Online]. From: http://www.washington.edu/news/1999/10/04/dyslexic-children-use-nearly-five-times-the-brain-area/ [Accessed on 7 December 2015]

[18] News in Science. 1999. Brains Different Dyslexic Are. [Online]. From: http://www.abc.net.au/science/news/health/HealthRepublish_57074.htm [Accessed on 7 December 2015]

[19] The Writer. How Readable Is Your Writing? [Online]. From: http://www.thewriter.com/what-we-think/readability-checker/ [Accessed on 7 December 2015]

[20] Readability-Score.Com website https://readability-score.com/

[21] Z.Zubir, M.M.Daud & A. H. Ridzuan.2014.Disleksia Dalam Aspek Ejaan Bacaan Bahasa Melayu. Symposium of International Language & Knowledge (SiLK 14).

[22] K. Drummond. 2009. About Reading Disabilities, Learning Disabilities and Reading Difficulties. Reading Rockets. [Online]. From: http://www.readingrockets.org/article/about-reading-disabilities-learning-disabilities-and-reading-difficulties [Accessed on 17 January 2016]

[23] W. Herbert. 2011. Is Reading Comprehension A Hidden Disability? Association For Psychological Science. [Online]. From: http://www.psychologicalscience.org/index.php/news/full-frontal-psychology/is-reading-comprehension-the-hidden-disability.html [Accessed on 17 January 2016]

[24] K. Abdullah. 2008. Research Report: A Study on The Use of Electronic Stories in Malay Language Teaching. Jurnal Elektronik Jabatan Bahasa & Kebudayaan Melayu. Jurnal e-Utama, Vol. 1.

[25] Test Document Readability. Readability Calculator. [Online]. From: http://www.online-utility.org/english/readability_test_and_improve.jsp [Accessed on 21 January 2016]

[26] N.A.M. Hazawawi, MM.H. Zakaria & S. Hisham 2015. Developing a New Instrumnet to Measure Young Adults Dyslexics' Reading Competencies in Malaysian Language. International Conference on Special Education 2015.

[27] N.A.M. Hazawawi, M.H.Zakaria & S. Hisham 2015. Measuring Reading Competencies in Malay Language Among Young Adults With Reading Difficulties. Jurnal Teknologi. Vol.77,No.1.