Faculty of Information and Communication Technology

DEVELOPMENT OF THE READABILITY INDEX MODEL FOR YOUNG ADULTS WITH READING DIFFICULTIES IN MALAY LANGUAGE

Nur Amalina binti Mohamad Hazawawi

Master of Science in Information and Communication Technology

2019

© Universiti Teknikal Malaysia Melaka

DEVELOPMENT OF THE READABILITY INDEX MODEL FOR YOUNG ADULTS WITH READING DIFFICULTIES IN MALAY LANGUAGE

NUR AMALINA BINTI MOHAMAD HAZAWAWI

A thesis submitted
in fulfillment of the requirements for the degree of Master of Science in Information and Communication Technology

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2019

© Universiti Teknikal Malaysia Melaka

DECLARATION

I declare that this thesis entitled “Development of the Readability Index Model for Young Adults with Reading Difficulties in Malay Language” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ……

Name : ……

Date : ……..…….

© Universiti Teknikal Malaysia Melaka

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Master of Science in Information and Communication Technology.

Signature	:	…….
Supervisor Name	:	…….
Date	:	…….

© Universiti Teknikal Malaysia Melaka

DEDICATION

This thesis is dedicated to my parents, wife, family and my siblings.
For their love, support, help and encouragement.

© Universiti Teknikal Malaysia Melaka

ABSTRACT

Recent studies reported different perspectives of the relationship among human, technology and education. Reading is a fundamental part of language competency. The motivation of this study is based on the fact that there are some young adults aged 18 years old and above who still cannot read properly and are having difficulties in reading and writing. Due to the unavailability of a tool that can measures the readability levels of Malay texts, this study proposed an online readability test tool in the Malay language to include readers who have reading difficulties by using an existing conventional calculation method developed by the Malay linguists. This formula was then transformed into a computing algorithm to detect the difficult words in Malay texts. This online tool is known as *Sistem Penilaian Kebolehbacaan Bahasa Melayu* (SPIKE) aims to measure the readability levels of Malay reading materials and to focus on detecting difficult Malay words. SPICE (Software Process Improvement and Capability Determination) had been chosen as research methodology, where consists of five main ideas which are S(Situation), P(Problem), I(Investigate), C (Construct) and E(Evaluate). SPIKE is part of a continuous work done to evaluate the reading competencies among young adults with dyslexia. From the existing readability formula and recent studies, a new parameter, which is the potential difficult word was found to help the dyslexics, especially in reading. The findings reported that the additional parameters like words that have more than three syllables and potential difficult words influenced the time taken in reading the assessment. This tool will increase the accuracies in measuring reading competencies among young adults, and will be able to help the dyslexic users with reading difficulties in knowing the level of readability of each reading materials.

© Universiti Teknikal Malaysia Melaka

ABSTRAK

Akhir-akhir ini banyak kajian telah membuktikan bahawa terdapat banyak pandangan perspektif dari segi hubungan antara manusia, teknologi dan pendidikan. Seperti yang kita tahu, membaca adalah sebahagian asas kompetensi bahasa. Motivasi untuk mengkaji masalah ini disebabkan sebilangan orang dewasa berumur 18 tahun ke atas masih tidak boleh membaca dan menulis dengan lancar. Disebabkan tiada alat yang boleh mengukur tahap kebolehbacaan teks Melayu, ahli bahasa Melayu sebelum ini menggunakan kaedah pengiraan kebolehbacaan secara konvensional. Kajian ini telah mencadangkan satu alat dalam talian di dalam Bahasa Melayu untuk membantu pembaca yang mempunyai masalah membaca dan kemudiannya diubah menjadi algoritma pengkomputeran untuk mengesan perkataan sukar dalam teks Melayu. Alat dalam talian ini dikenali sebagai Sistem Penilaian Kebolehbacaan Bahasa Melayu (SPIKE), di mana bertujuan untuk mengukur tahap kebolehbacaan bahan bacaan bahasa Melayu dan memberi tumpuan kepada mengesan perkataan sukar Bahasa Melayu. SPICE (Software Process Improvement and Capability Determination) telah dipilih sebagai kaedah penyelidikan di mana mempunyai lima idea utama iaitu S(Situation), P(Problem), I(Investigate), C (Construct) and E(Evaluate). SPIKE adalah sebahagian daripada kajian yang dilakukan untuk menguji kecekapan pembacaan dalam kalangan orang dewasa yang menghidapi disleksia. Berdasarkan formula sedia ada dan kajian terdahulu, kami memperkenalkan parameter baru, di mana parameter tersebut dikenali sebagai perkataan yang berpotensi sukar dan boleh membantu orang disleksia ketika membaca. Keputusan menunjukkan parameter baru seperti perkataan yang mempunyai lebih daripada tiga suku kata dan perkataan yang berpotensi sukar memberi kesan kepada masa ketika orang dewasa membaca. Alat ini dapat membantu meningkatkan ketepatan dalam mengukur kompetensi bacaan bagi orang dewasa dan membantu pengguna disleksia yang mempunyai masalah dalam bacaan dengan mengetahui tahap kebolehbacaan setiap bahan bacaan.

© Universiti Teknikal Malaysia Melaka

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Allah for giving me the strength to finish my study. I would like to acknowledge my limitless gratitude to both of my supervisors, Dr. Mohd Hafiz Bin Zakaria and Mrs. Syariffanor Binti Hisham for their supports, expertise, generously shared knowledge, insightful criticism, and guidance throughout my research or writing. Thank you again for their time spent on me. I am deeply indebted to my beloved parent, Mohamad Hazawawi Bin Yusof and Arpah Binti Atan, for their untiring love, motivation, moral support and also financial assistance in order for me to finish my study. To my sisters, Nur Amanina Binti Mohamad Hazawawi and Nur Amalia Binti Mohamad Hazawawi, thank you for the encouragement and support. To my husband, Ahmed Yaser Bin Mohd Zabawi, thank you for your assistance and words of encouragement. It was a great help throughout the course of this research work. I humbly extend my thanks to all who have given full cooperation throughout the journey in completing my study.

© Universiti Teknikal Malaysia Melaka

TABLE OF CONTENTS

	PAGE
DECLARATION	
APPROVAL	
DEDICATION	
ABSTRACT	**i**
ABSTRAK	**ii**
ACKNOWLEDGEMENTS	**iii**
TABLE OF CONTENTS	**iv**
LIST OF TABLES	**vii**
LIST OF FIGURES	**x**
LIST OF APPENDICES	**xiii**
LIST OF ABBREVIATIONS	**xiv**
LIST OF PUBLICATIONS	**xv**
CHAPTER	
1. INTRODUCTION	**1**
1.1 Background of the study	3
1.1.1 Young adults with dyslexia	3
1.1.2 Reading by dyslexics	5
1.2 Learning disabilities in Malaysia	6
1.2.1 Literacy rates in Malaysia	6
1.3 Research problem	8
1.4 Research question	9
1.5 Research objective	9
1.6 Research contribution	9
1.7 Thesis organization	10
1.8 Summary	11
2. LITERATURE REVIEW	**12**
2.1 Introduction	12
2.2 Dyslexia	12
2.2.1 Definition of dyslexia	13
2.2.2 Symptoms and characteristics of dyslexia	18
2.3 Types of dyslexia	21
2.4 Young adults with dyslexia in Malaysia	24
2.4.1 Age range of young adults	24
2.4.2 Malay language as a nation language in Malaysia	27
2.5 Reading ability among young adults in Malaysia	28
2.6 Readability formula in Malay and English language	32
2.6.1 Definition of readability formula	32
2.6.2 Existing readability formula	33

© Universiti Teknikal Malaysia Melaka

2.6.3 Readability formula in Malay language 39
2.7 Construction of words in Malay and English language 40
2.8 Readability test tools 41
2.9 Lesson learned in these studies 49
2.10 Summary 50

3. METHODOLOGY 51
3.1 Introduction 51
3.2 Research approach 51
3.3 Research design 51
3.4 Formulation of Malay readability formula 54
3.4.1 The differences and similarities for Malay and English words 55
3.4.2 Evaluation of the existing readability formula 56
3.4.2.1 Samples of articles 57
3.4.2.2 Data collection procedure 58
3.4.3 Additional of language parameter 59
3.4.3.1 Sample of articles 59
3.4.3.2 Data collection procedure 60
3.4.4 Validate the proposed formula 63
3.4.4.1 Samples of articles 64
3.4.4.2 Data collection procedure 66
3.4.5 Implementation Malay language rules 67
3.4.5.1 Calculating syllables in Malay language 67
3.4.5.2 Calculating words in Malay language 69
3.4.5.3 Potentially difficult words 71
3.4.5.4 Aspects that were omitted in SPIKE 72
3.5 Data collection in controlled study and field study 73
3.5.1 Controlled study 74
3.5.1.1 Participants 74
3.5.2 Field study 75
3.5.2.1 Participants 76
3.5.2.2 Location of study 77
3.5.2.3 Research instruments 77
3.5.2.4 Data collection procedures 78
3.6 Summary 81

4. DEVELOPMENT OF SPIKE 82
4.1 Introduction 82
4.2 Introduction of SPIKE 82
4.3 Software requirements 84
4.4 Interfaces of SPIKE 84
4.5 Summary 95

© Universiti Teknikal Malaysia Melaka

5. RESULT AND DISCUSSION **96**
5.1 Introduction 96
5.2 Data analysis 96
5.2.1 Data analysis for evaluation of the existing readability formula 96
5.2.2 Data analysis for addition of language parameter 101
5.3 Normality test 107
5.4 Spearman correlation 109
5.5 Controlled study 110
5.6 Field study analysis 113
5.7 Descriptive statistics 114
5.8 Correlations between participants who scored 6 and above and time taken 118
5.9 Correlations between participants who scored below 5 toward time taken for each level 121
5.10 Correlation analysis of variables between score and time taken 123
5.11 Correlation analysis of variables between score and range of time taken 126
5.12 Number of syllables and number of difficult words 129
5.13 Discussion 133
5.13.1 Research instruments used in assessment 133
5.13.2 Aspects of texts in passage 134
5.13.3 Aspects of language 136
5.13.4 Reading speed 137
5.14 Summary 139

6. CONCLUSION AND FUTURE WORKS **140**
6.1 Introduction 140
6.2 Summary of research objectives 140
6.3 Summary of contributions 143
6.4 Conclusion 148
6.5 Limitations 149
6.6 Future Works 150

REFERENCES **152**
APPENDICES **173**

© Universiti Teknikal Malaysia Melaka

LIST OF TABLES

TABLE	TITLE	PAGE
1.1	Literacy rates by sex	7
1.2	Description of each chapter	10
2.1	Overview of Chapter 2	12
2.2	Summary of meaning by authors	17
2.3	The characteristics of individuals with dyslexia	19
2.4	The differences of symptoms by three articles	20
2.5	Types of dyslexia (Pierangelo, 2010)	22
2.6	Summary of definition of age range of young adults	26
2.7	Table of existing readability formulas	34
2.8	Table of suitable formulas according to respective fields	37
2.9	Readability level according to year in primary school and form in secondary school	39
2.10	Advantages and disadvantages of the readability tools	43
2.11	Readability indicators for each readability test tools	45
3.1	The differences and similarities for Malay language and English language in readability test tool	55
3.2	The readability level for each reading material sample	59
3.3	Classification for each reading sample	60
3.4	The percentage of participants according to gender	61

© Universiti Teknikal Malaysia Melaka

3.5	The percentage of participants according to races	62
3.6	The expansion of the readability level table	64
3.7	The readability level for each reading material sample with topics	65
3.8	Number of articles based on readability level of reading sample	67
3.9	Table of *pola suku kata*	69
3.10	Examples on how to count the words in Malay language	70
3.11	Rules for *Kata Pinjaman*	72
3.12	Exmples of *Kata Singkatan* and punctuations	73
3.13	Demographic details	74
3.14	Average time taken to read the articles	75
3.15	Demographic details	76
3.16	The readability level for each reading material samples	78
4.1	Description of software requirements	84
5.1	Reading material samples versus the readability level	97
5.2	The results of number of syllables versus the readability level	98
5.3	The number of potential difficult words for dyslexia versus the readability level	100
5.4	Results of questionnaires for 30 participants	102
5.5	Analysis of reading assessment for sample essay (II)	103
5.6	Analysis of reading assessment for sample essay (III)	104
5.7	Analysis of reading assessment for sample essay (IV)	105
5.8	Average time taken for each essay	106
5.9	Normality test for score over 6 and time taken for each level	107
5.10	Normality test for score below 5 and time taken for each level	108

© Universiti Teknikal Malaysia Melaka

5.11	The strength of the correlation according to the value of Spearman's correlation coefficient	109
5.12	Scale of measurement and types of statistical tests	110
5.13	Correlation matrix on time taken and range of time	111
5.14	Frequencies of participants scored for each mark	114
5.15	Frequencies of participants scored for each mark in most critical symptoms questions	116
5.16	Frequencies of participants scored for each mark in less critical symptoms questions	117
5.17	Correlations of participants who scored 6 and above towards time taken for each level	118
5.18	Correlations of participants who scored below 5 towards time taken for each level	121
5.19	Score versus time taken for each level	124
5.20	Score versus range of time taken for each level	127
5.21	Number of syllables for each level of articles	130
5.22	Number of difficult words for each level of articles	131
6.1	The expansion of the readability level table	146

© Universiti Teknikal Malaysia Melaka

LIST OF FIGURES

FIGURE	TITLE	PAGE
1.1	Statistic of students with Dyslexia in *Program Pendidikan Khas* in Malaysia *(Data pendidikan khas, pendidikan khas, Kementerian Pendidikan Malaysia)*	4
1.2	Literacy rates in Malaysia from 1995 until 2015 *(Central Intelligence Agency, 2015)*	7
2.1	Graph theory of subtype of dyslexia (Dyslexia Reading Well, 2014 and Funnell, 2013)	21
2.2	Readable.IO website	47
3.1	Research design of the study	52
3.2	Steps to formulate a formula	54
3.3	Block diagram of process in selecting and measuring the readability level of reading sample	58
3.4	Block diagram of process in selecting and modifying reading sample	60
3.5	Processes involved in selecting and measuring the readability level of reading sample	66
3.6	Process in conducting the field study	79
4.1	Overview flow of SPIKE	83
4.2	SPIKE homepage	85
4.3	Personal information	85

© Universiti Teknikal Malaysia Melaka

4.4	Programming code for personal information	86
4.5	Information of text	87
4.6	Programming code for text information	87
4.7	Programming code for text information (after click the "proses" button)	88
4.8	Programming code for 300 words	88
4.9	Readability level for normal readers	89
4.10	Programming code for readability level for normal readers	89
4.11	Readability level for dyslexia readers	90
4.12	Programming code for readability level for dyslexia readers	91
4.13	Programming code for results to be matched with readability level table	91
4.14	Text Indicator	92
4.15	Programming code for text indicator	92
4.16	Report for pre-screening dyslexia for adults	93
4.17	Report for readability test	94
5.1	Graph of total scores	115
5.2	Graph of score of most critical symptoms questions	116
5.3	Graph of score of less critical symptoms questions	117
5.4	Readability level versus number of syllables	130
5.5	Readability level versus number of difficult words	132
6.1	A transformation from readability formula statement into programming code	143

© Universiti Teknikal Malaysia Melaka

6.2 A transformation from expansion of readability formula statement into programming code 144

6.3 Process of formulation of formula 146

© Universiti Teknikal Malaysia Melaka

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Articles (Calculation Manually)	172
B	Booklet of Articles (Essay I, II,III,IV)	188
C	Questionnaires	192
D	*Senarai Semak Disleksia Orang Dewasa*	193
E	Booklet of Articles (Field Study)	194

© Universiti Teknikal Malaysia Melaka

LIST OF ABBREVIATIONS

KKC	-	Kolej Komuniti Chenderoh, Perak
KKJ	-	Kolej Komuniti Jasin, Melaka
SPICE	-	Software process improvement and capability determination
SPIKE	-	Sistem penilaian kebolehbacaan bahasa Melayu

© Universiti Teknikal Malaysia Melaka

LIST OF PUBLICATIONS

1. Hazawawi N. A. M., Zakaria M. H., Hisham S. and Sainawan S.S., 2017. An Evaluation of Proposed Readability Formula in Malay Language. *International Symposium on Research in Innovation and Sustainability 2017.*

2. Hazawawi N. A. M., Zakaria M. H. and Hisham S., 2016. SPIKE: Online Reading Competencies Measure for Malay Language. *Journal of Telecommunication and Electronic and Computer Engineering* UTeM. 8(2).

3. Hazawawi N. A. M., Zakaria M. H. and Hisham S., 2016. Measuring Reading Competencies in Malay Language among Young Adults with Reading Difficulties. *Jurnal Teknologi.* 77(19).

4. Hazawawi N. A. M., Zakaria M. H. and Hisham S., 2015. Developing a New Instrument to Measure Young Adults Dyslexics' Reading Competencies in Malaysian Language. *International Conference on Special Education (ICSE) 2015.*

© Universiti Teknikal Malaysia Melaka

CHAPTER 1

INTRODUCTION

Readability formula, a mathematical formula, was proposed since long ago to measure reading ease and grade level. Readability formulas were constructed from two basic parameters which were average words per sentence and average syllables per words (Doverspike, 2015). The Malay language readability test tool, named *Sistem Penilaian Kebolehbacaan Bahasa Melayu* (SPIKE) was developed to assist readers in knowing the score of a Malay text and acquired reading level and grade level. Readability formula in Malay language was proposed by Khadijah Rohani in 1984 and was implemented in the SPIKE. The existing readability test tools were made primarily for English text with options of either to type or to paste the text, link directly from the webpage, and the scored will be calculated based on several readability formula like Flesch Kincaid Grade Level and SMOG Index and so on. For this reason, this project was carried out to propose an expansion of readability formula that can measure reading ease in Malay language and indirectly could assist people with reading difficulties. Furthermore, readability has focused more on the reader factors which may affect understanding (Wray and Janan, 2013). The motivation of this study is based on the fact that there are some young adults aged 18 years old and above who still cannot read properly and are having difficulties in reading and writing. At the meantime, most people are not aware of the suitable reading materials that would match their reading competencies and their age level. SPIKE was introduced to cater this problem, where Information and Communication Technology (ICT) solution was used to measure level of

© Universiti Teknikal Malaysia Melaka

reading by developing an online readability test tool. The new technologies were integrated with existing technologies in use, where calculation was transformed from manual calculation into automated calculation through online test tool. In order to implement this new technology, someone has to be skillful enough to manage the challenges that go along with implementing the changes. It was observed that new and emerging technologies were being integrated with the older technologies to make ICT applications in education more effective. Overall, technology change can improve efficiency, increase quality and also enhance communication (Delaney and D'Agostino, 2015). In the last two decades, technology investment in schools show an increment, based on the assumption that technology-mediated learning environments offered chances for undergraduates to analyze information, solve problems, communicate and collaborate with each other (Lim et. al., 2013). In addition, one of the reasons that ICT literacy was included was not only technological proficiency, but also cognitive skills such as those underlying reading and problem solving, which are critical to using ICT literacy effectively (International ICT Literacy Panel, 2002). ICT literacy involve the form of literacy, where communicating data thru digital medium were as critical as reading and writing were in prior hundreds of years. ICT proficiency model encompassed 7 components which were define, access, manage, integrate, evaluate, create, and communicate, where those were within the context of cognitive, ethical and technical (Gregorian, 2002; Kartz and Macklin, 2007). The first part of this chapter started with the background information regarding potentially dyslexics among Malaysian young adults. The definition of dyslexia and the age range for young adults in Malaysia and other countries were explained in the same section. Other than that, existing readability formulas in English as well as the advantages and benefits of having this readability formula in Malay language were discussed briefly in this chapter. Furthermore, research questions and research objectives were portrayed based on our

© Universiti Teknikal Malaysia Melaka

research motivations in section 1.4 and 1.5 respectively. The research contributions were mentioned prior to this study. The last section of this chapter showed the outline of the whole thesis.

1.1 Background of the study

Identifying potential dyslexics among young adults could be a difficult task. This was due to lack of awareness about the impacts of dyslexia to young adults, and partly because of the current dyslexia awareness related activities in Malaysia that were mainly focused on the intervention program for children with dyslexia.

1.1.1 Young adults with dyslexia

Based on a report by British Dyslexia Association and Dyslexia International, in higher education, young adults complained of poor support within their sectors; their lecturers and tutors had lack of awareness regarding dyslexia. For example, the exam format can be biased to dyslexic students (British Dyslexia Association, 2012; Dyslexia International, 2016). Some of the dyslexic students maybe had overcome their reading problems but they certainly do not become a skilled reader.

Below showed a graph of statistics by *Program Pendidikan Khas* in Malaysia on students with dyslexia in primary and secondary schools from 2013 till 2017.

Figure 1.1 below showed that there were a dramatically increase in the number of students with dyslexia who enrolled in the education program from 2013 until 2017.

© Universiti Teknikal Malaysia Melaka

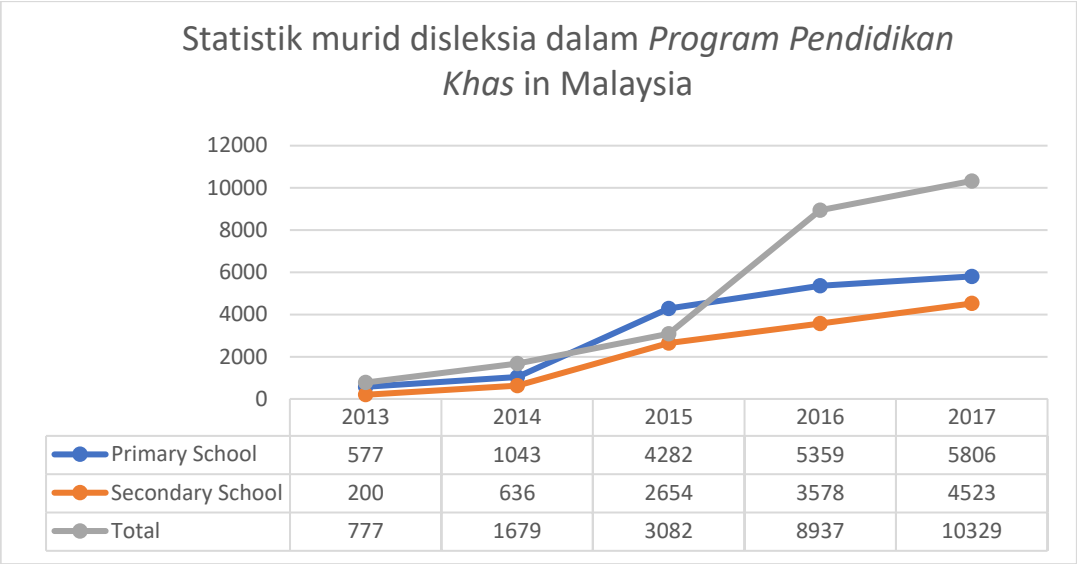

Figure 1.1: Statistic of students with Dyslexia in *Program Pendidikan Khas* in Malaysia (*Data pendidikan khas, pendidikan khas, Kementerian Pendidikan Malaysia*)

Young adults with dyslexia in Malaysia faced more challenges due to the fact that they had to master more than one language. Textual information was difficult to understand by young adults with dyslexia even some of them might have survived secondary and tertiary education (Rose, 2009). There was a crucial need to have research evidences on the prevalence of dyslexia among young adults in Malaysia. Online dyslexia screening was thought as potential instrument to identify students with dyslexia at schools (Ekhsan, Ahmad, Halim, etc, 2012). This was partly due to the fact that many young adults were not screened during their childhood. This was supported by a study that reported that many teachers had to go through the tough experience in teaching slow readers (Turner, 1997). It was observed that screening for young adults usually require a higher scale of guidance (Singleton, 2009). In Malaysia, there are still no online dyslexia screening, hence the community had to visit dyslexia centers to run-through the manual screening test which was done by the experts, and that could be timely and expensive. Yet, Turner (1997) argued that there could be some difficulties to screen for dyslexia in conditions that are hard to put it in right way and do not much impairment.

© Universiti Teknikal Malaysia Melaka