

**UNIVERSITI TEKNOLOGI MARA**

**COMPUTATIONAL TEXT ANALYSIS OF  
INTERMEDIATE AND HIGH-  
INTERMEDIATE READING PASSAGES FOR  
ESL LEARNERS**

**ANEALKA AZIZ HUSSIN**

Thesis submitted in fulfillment of the requirements  
for the degree of  
**Doctor of Philosophy**  
**Faculty of Education**

March 2010

## DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the result of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any other degree or qualification.

In the event that my thesis be found to violate the conditions mentioned above, I voluntarily waive the right of conferment of my degree to be subjected to the disciplinary rules and regulations of Universiti Teknologi MARA.

Name of Candidate	Anealka bt Aziz Hussin
Candidate's ID No	2005218443
Programme	PhD in Education
Faculty	Faculty of Education
Thesis Title	Computational Text Analysis of Intermediate and High-Intermediate Reading Passages for ESL Learners

Signature of Candidate

Date 26 March 2010

## ABSTRACT

The main concerns of the present study are on issues of matching reading materials to intended learners and standardizing reading materials difficulty level for learners with similar reading ability. This study intends to identify ways to improve the present practice of selecting and adapting reading materials for ESL learners so that the process can be done in a more objective, consistent and comprehensive manner. To do that, the study proceeds to determine text characteristics that can significantly differentiate the difficulty level of intermediate and high-intermediate reading passages (referred as IR and HIR passages respectively), determine additional predictors that can enhance the efficiency of the Flesch Reading Ease formula to estimate passage difficulty at sentence and word level and develop a set of instruments that can estimate the difficulty level of these passages in a more precise manner.

The Descriptive Correlational approach is used as the research design for the study to achieve the above objectives. The samples of the study come from IR and HIR passages. Three computational tools, the Flesch Reading Ease formula, Writer's Workbench 8.18 and Word Smith Tools 4.0, are used to extract information related to passage difficulty at text, sentence and word level respectively. The study also replicates and improves on Vogel and Washburne's (1928) process of developing readability formula. Descriptive statistics namely mean, range and standard deviation, and inferential statistics such as t-test, correlational analysis and multiple regression analysis are used to analyze the data collected.

It is found that all the text characteristics used in the study can significantly differentiate the difficulty level of reading passages between the two levels. However, not all these characteristics have significant relationships with passage difficulty level. Two of the significant ones namely the Flesch Reading Ease score (FRE) and the coverage of high frequency words (HFW), are finally selected as predictors to develop a more refined difficulty level formula. Together with the established recommended difficulty range for IR and HIR passages, a set of instruments to estimate passage difficulty level on the refined Intermediate scale are constructed.

This study is of great significance to language instructors, material writers, language learners, test-setters, publishers, researchers and self-access learners. With a more precise readability formula and its accompanying instruments, selection and adaptation of reading passages can be done in a more objective, consistent and comprehensive manner.

## ACKNOWLEDGEMENT

All praise and thanks to Allah s.w.t. who gave me strength and patience to complete this doctoral research to its completion.

My utmost appreciation goes to my supervisors: Retired Professor Datin Dr Zubaida Alsree and Assoc. Prof. Dr Chan Yuen Fook, for their continuous guidance and advice in helping me to complete my research. All your advice and assistance will not be forgotten indeed.

My gratefulness also goes to the Dean of the Academy of Language Studies, Institute of Graduate Studies, Postgraduate Committee Members at the Faculty of Education and UiTM for all the leave granted and the support given throughout my study.

My special appreciation goes to my husband, Norzairee Khalid Nordin, who provides continuous support and encouragement, and has stood by me through thick and thin in completing the thesis and to my children, Zarif Luqmani, Zafri Azizi and Zafir Husaini, who continue to love me unconditionally. My deepest thanks also goes to my parents, Aziz Hussin and Ramlah Marily, my in-laws, Khalid Nordin Mohd Nor and Zainah Mohd Sabil, my brothers and sisters, Andika, Adi Putra, Mohd Indera, Nur Atika, NurSyafiqah, Yuhana and Norzaireene, for their continuous support and love.

My gratitude also goes to my friends Boboy, Norwati, Vahid, Dzeelfa, Masturah, Baizura, Raihan, Azlina, Haliza, Ainul Rusmin, Haryati, Ikin, Aini, Asiah, and Norehan. Your love and friendship will always be remembered.

My appreciation also goes to Mastura, Azlina and Rosiah for their help as inter-raters.

My recognition also goes to my colleagues: Raihan, Baizura, Haliza, Azlina and Suryani for trying out the instruments developed by this study.

## TABLE OF CONTENTS

<b>TITLE PAGE</b>	i
<b>CANDIDATE’S DECLARATION</b>	ii
<b>ABSTRACT</b>	iii
<b>ACKNOWLEDGMENTS</b>	iv
<b>TABLE OF CONTENTS</b>	v
<b>LIST OF TABLES</b>	x
<b>LIST OF FIGURES</b>	xiii
<b>LIST OF APPENDICES</b>	xvii
<b>LIST OF ABBREVIATIONS</b>	xviii

### CHAPTER 1: INTRODUCTION

1.1 Background of the Study	1
1.1.1 Pre-tertiary Learners in Malaysia	3
1.1.2 Pre-tertiary Learners in Universiti Teknologi MARA (UiTM)	6
1.1.3 The Need for the Present Study	7
1.2 Statement of the Problem	11
1.3 Rationale of the Study	12
1.4 Purpose and Objectives of the Study	19
1.5 Research Questions	20
1.6 Conceptual Framework of the Study	21
1.7 Significance of the Study	26
1.8 Scope of the Study	29
1.9 Limitations of the Study	30
1.10 Operational Definition of Terms	32
1.10.1 Average Sentence Length (ASL)	32
1.10.2 Average Word Length (AWL)	32
1.10.3 Complex Sentences (Cx)	33