

**A CORPUS-BASED ANALYSIS OF VOCABULARY LOAD
AND DISTRIBUTION IN “*HANG OUT 4*” TEXTBOOK FOR
INDONESIAN ELEMENTARY EFL STUDENTS**

A THESIS



By:

ANGELINA DINDA ELVARA

1213019030

**ENGLISH LANGUAGE EDUCATION STUDY PROGRAM
TEACHER EDUCATION FACULTY
WIDYA MANDALA CATHOLIC UNIVERSITY SURABAYA
JUNE 2026**

**A CORPUS-BASED ANALYSIS OF VOCABULARY LOAD
AND DISTRIBUTION IN “*HANG OUT 4*” TEXTBOOK
FOR INDONESIAN ELEMENTARY EFL STUDENTS**

A THESIS

*Presented to Teacher Education Faculty
Widya Mandala Surabaya Catholic University
in partial fulfillment of the requirement of the Degree of
Sarjana Pendidikan in English Language Education*



By:

Angelina Dinda Elvara

1213019030

**ENGLISH LANGUAGE EDUCATION STUDY PROGRAM
TEACHER EDUCATION FACULTY
WIDYA MANDALA CATHOLIC UNIVERSITY SURABAYA
JUNE 2026**

APPROVAL SHEET (I)

This thesis proposal entitled “A Corpus-Based Analysis of Vocabulary Load and Distribution in “*Hang Out 4*” Textbook for Indonesian Elementary EFL Students” prepared and submitted by Angelina Dinda Elvara (1213019030) has been approved by to be examined by the Thesis Board of Examiners.



Dr. Johanes Leonardi Taloko, M.Sc.
Thesis Advisor

APPROVAL SHEET (II)

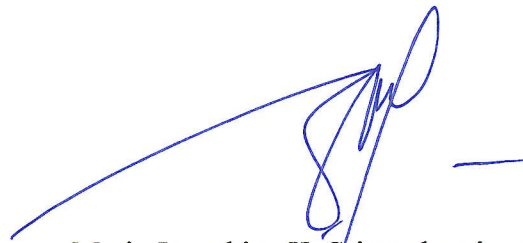
This thesis entitled “A Corpus-Based Analysis of Vocabulary Load and Distribution in “*Hang Out 4*” Textbook for Indonesian Elementary EFL Students” prepared and submitted by Angelina Dinda Elvara 1213019030 has been examined and declared PASSED by the Thesis Board of Examiners.



Drs. B. Himawan Setyo Wibowo, M.Hum.
Chairperson



Dr. Johanes Leonardi Taloko, M.Sc.
Secretary



Maria Josephine K. Sriemulyaningsih, M.Pd.
Member



Mateus Yumarnamto, Ph.D.
Dean of Teacher Education Faculty



Maria Josephine K. Sriemulyaningsih, M.Pd.
**Head of English Language Education
Study Program**

SURAT PERNYATAAN
Jalur Skripsi

Bersama in, saya:

Nama : ANGELINA DINDA ELVARA
NRP : 123019030
Program Studi : PENDIDIKAN BAHASA INGGRIS
Jurusan : BAHASA INGGRIS
Fakultas : KEGURUAN DAN ILMU PENDIDIKAN

Menyatakan dengan sesungguhnya bahwa skripsi saya yang berjudul:
A CORPUS-BASED ANALYSIS OF VOCABULARY LOAD AND
DISTRIBUTION IN "HANGOUT 4" TEXTBOOK FOR INDONESIAN
ELEMENTARY EFL STUDENTS

benar-benar merupakan hasil karya sendiri. Apabila Skripsi ini ternyata merupakan hasil *plagiarisme*, maka saya bersedia menerima sanksi berupa pembatalan kelulusan dan/atau pencabutan gelar yang telah saya peroleh.

Demikian surat pernyataan ini saya buat dengan sesungguhnya dan penuh kesadaran.

Surabaya,
Yang membuat pernyataan



ANGELINA DINDA ELVARA

Mengetahui,
Dosen Pembimbing Tunggal

Dr. JOHANNES LEONARDI TALOKO, M. Sc.
NIK. 121. 98. 0328

**SURAT PERNYATAAN
PERSETUJUAN PUBLIKASI KARYA ILMIAH**

Demi perkembangan Ilmu Pengetahuan, saya sebagai mahasiswa Universitas Katolik
Widya Mandala Surabaya:

Nama Mahasiswa : ANGELINA DINDA ELVARA
Nomor Pokok : 1213019030
Program Studi Pendidikan : BAHASA INDONESIA
Jurusan : PENDIDIKAN BAHASA INDONESIA
Fakultas : KEGURUAN DAN ILMU PENDIDIKAN
Tanggal Lulus : 24 JUNI 2026

Dengan ini SETUJU/TIDAK SETUJU*) Skripsi atau Karya Ilmiah saya,

Judul:

A CORPUS-BASED ANALYSIS OF VOCABULARY LOAD
AND DISTRIBUTION IN "HANG OUT 4" TEXTBOOK FOR
INDONESIAN EFL STUDENTS

Untuk dipublikasikan/ditampilkan di Internet atau media lain (Digital Library
Perpustakaan Universitas Katolik Widya Mandala Surabaya) untuk kepentingan
akademik sebatas sesuai undang-undang Hak Cipta yang berlaku.

Demikian surat pernyataan SETUJU/~~TIDAK SETUJU~~*) publikasi Karya Ilmiah ini saya
buat dengan sebenarnya.

Surabaya, 24 JUNI 2026

Yang membuat pernyataan,



Angela

ANGELINA DINDA ELARA
NRP. 1213019030

*)coret salah satu

STATEMENT OF AUTHENTICITY

I declare that this thesis is my own writing and its true and correct that I did not take any scholarly ideas or work from others dishonesty. All the cited works were quoted in accordance with the ethical code of academic writing I will take all the consequences if plagiarism is found in this thesis.

Surabaya, 24 Juni 2026



Angelina Dinda Elvara
NRP. 1213019030

ACKNOWLEDGEMENTS

First and foremost, the writer would like to thank Jesus Christ for His grace, blessings, and strength that allowed the writer to complete this thesis. This journey would not have been possible without the help, guidance, support, and encouragement from so many people. The writer would like to express heartfelt gratitude to:

1. The writer's beloved family, especially mom Netty, dad Pius, sis Yovita, niece Kalea for their endless patience, love, and unwavering support throughout the entire process.
2. Dr. Johanes Leonardi Taloko, M.Sc., the writer thesis advisor, for his invaluable guidance, feedback, and continuous support throughout the writing and research process.
3. Maria Josephine K.S., M.Pd., Head of the English Language Education Study Program, for facilitating both the proposal seminar and thesis defense schedule.
4. Drs. B. Himawan Setyo Wibowo, M. Hum and Maria Josephine K.S., M.Pd., the writer's thesis examiners, for their insightful feedback and helpful suggestions which significantly improved the writer's work.
5. All the lecturers and administrative staff of the Faculty of Teacher Training and Education—thank you for the help and guidance provided along the way.
6. The writer's friends—thank you for the constant support, encouragement, and prayers during this entire journey.
7. The researcher's beloved family who have prayed and supported for the researcher to be able to finish this thesis.
8. To myself, Angelina Dinda Elvara, for the commitment and responsibility shown in completing what was started. The journey was not always easy; however, through perseverance, courage, and faith, I was able to endure. The belief that every effort will ultimately bring goodness became a source of strength throughout every stage, despite the exhaustion and various challenges encountered along the way. All the struggles undertaken have

yielded tangible results, and the prayers that were offered have been answered gradually. *“For surely there is a future, and your hope will not be cut off”* served as a motivation for me to keep moving forward to the next stage. This experience has become a valuable lesson that the process itself and personal growth are things to be grateful for and deeply appreciate.

Finally, the writer truly hopes this thesis will be useful for the field of English education and for anyone who reads it.

Surabaya, 24 June 2026

A handwritten signature in black ink, appearing to read 'Dinda', with a small heart symbol above the first letter.

Angelina Dinda Elvara

NRP. 1213019030

TABLE OF CONTENTS

COVER	i
APPROVAL SHEET (I)	ii
APPROVAL SHEET (II)	iii
STATEMENT OF AUTHENTICITY	iv
SURAT PERNYATAAN (I).....	v
SURAT PERNYATAAN (II)	vi
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENTS.....	ix
LIST OF FIGURES	xii
LIST OF TABLES.....	xiii
ABSTRACT.....	xiv
CHAPTER I INTRODUCTION	1
1.1 Background of the Study	1
1.2 Statements of the Problems	6
1.3 Objectives of the Study	6
1.4 Theoretical Framework	7
1.5 The Significance of the Study	7
1.6 Scope and Limitation of the Study	7
1.7 Definition of Key-Terms	8
1.8 Organization of the Proposal.....	9
CHAPTER II LITERATURE REVIEW	11
2.1 Vocabulary	11
2.1.1. Definition of Vocabulary	11
2.1.2. The Role of Vocabulary in EFL Learning	12
2.1.3. Vocabulary Load and Vocabulary Coverage	13
2.2 Corpus Linguistics	15
2.2.1. Definition of Corpus Linguistics.....	15
2.2.2. Corpus-Based Vocabulary Analysis.....	16
2.3 Tokens, Types, and Type–Token Ratio	17
2.4 Word Frequency and Vocabulary Distribution.....	18
2.4.1. Frequency Distribution of Words in Texts.....	19

2.4.2.	Vocabulary Levels Based on Frequency	20
2.4.3.	High-Frequency Words.....	20
2.4.4.	Mid-Frequency Words	22
2.4.5.	Low-Frequency Words.....	22
2.4.6.	Vocabulary Frequency and Text Coverage.....	23
2.5	Textbooks in EFL Learning.....	24
2.5.1.	Definition of Textbooks	25
2.5.2.	The Role of Textbooks in Vocabulary Learning.....	25
2.5.3.	Textbook Evaluation in Language Teaching	27
2.6.	Previous Studies	33
CHAPTER III RESEARCH METHOD.....		37
3.1.	Research Design	37
3.2.	Data and Data	39
3.3	Source	40
3.4	Research Instruments.....	40
3.5	Methods of Data Collection.....	41
3.6	Methods of Data Analysis	41
3.7	Research Schedule.....	43
CHAPTER IV FINDINGS AND DISCUSSION		44
4.1.	Description of Vocabulary Categories	44
5.2.1.	High-Frequency Words.....	44
5.2.2.	Academic Words.....	46
5.2.3.	Low-Frequency / Off-List Words.....	48
4.2.	Vocabulary Coverage and Vocabulary Load	49
4.3.	Recapitulation of Vocabulary Distribution and Coverage	52
4.4.	Discussion of Findings	54
4.4.1.	Types of Vocabulary Presented in Hang Out! 4 Student's Book	55
4.4.2.	Vocabulary Load and Appropriateness for Elementary-Level EFL Students	60
4.4.3.	Summary of Discussion	65
CHAPTER V CONCLUSION AND SUGGESTIONS		67
5.1.	Conclusion.....	67
5.2	Suggestions	68
5.2.1.	For English Teachers	68

5.2.2. For Coursebook Writers.....	69
5.2.3. For Future Researchers	70
REFERENCES	72
APPENDICES.....	74
Appendix A: Textbook Identity.....	74
Appendix B: AntWordProfiler File Profiler Display	76
Appendix C: AntWordProfiler Statistics Result.....	78

LIST OF FIGURES

Figure 1: Cover of Hang Out! 4 Student's Book	74
Figure 2: AntWordProfiler File Profiler Display Showing Textbook Corpus Input	76
Figure 3: AntWordProfiler File Profiler Display Showing Vocabulary Items from the Textbook Corpus	76
Figure 4: AntWordProfiler File Profiler Display Showing Additional Textbook Corpus Data.....	77
Figure 5: AntWordProfiler File Profiler Display Showing Thematic Vocabulary in the Textbook Corpus	77
Figure 6: AntWordProfiler Input File Statistics Result.....	78
Figure 7: AntWordProfiler Level List Statistics Result.....	78
Figure 8: AntWordProfiler Range Statistics Result	79
Figure 9: AntWordProfiler Input File Headword Statistics Result.....	79

LIST OF TABLES

Table 1. Types of Reading and Vocabulary Coverage	14
Table 2. Example of Word Frequency Distribution in a Text (adapted from Nation, 2001).....	19
Table 3. Vocabulary Levels Based on Frequency Nation (2022)	20
Table 4. High-Frequency Vocabulary Characteristics	21
Table 5. High-Frequency Vocabulary Characteristics	21
Table 6. Characteristics of Low-Frequency Vocabulary	23
Table 7. Characteristics of Vocabulary Frequency Levels	23
Table 8. Vocabulary Size Needed for Text Coverage Nation (2022)	23
Table 9. Coverage of Vocabulary Levels in Texts.....	24
Table 10. Approaches to Selecting Textbooks.....	28
Table 11. Stages of the Textbook Evaluation Process	29
Table 12. Types of Materials Evaluation	30
Table 13. Purposes of Materials Evaluation	30
Table 14. Guidelines for Effective Textbook Evaluation	31
Table 15. Checklist for Evaluating the Organization of Textbooks.....	31
Table 16. Vocabulary Evaluation Checklist.....	32
Table 17. Vocabulary Evaluation Checklist.....	32
Table 18. Distribution of High-Frequency Words	44
Table 19. Distribution of Academic Words	47
Table 20. Distribution of Low-Frequency / Off-List Words.....	48
Table 21. Recapitulation of Vocabulary Categories	53
Table 22. Textbook Identity	74

ABSTRACT

The use of textbooks plays an important role in English as a Foreign Language (EFL) learning, particularly in providing vocabulary input for elementary-level students. This study aimed to analyze the types of vocabulary and the vocabulary load presented in *Hang Out! 4 Student's Book* for Indonesian elementary EFL students. The study employed a descriptive quantitative design using a corpus-based approach. The data consisted of all textual content in the textbook, including dialogues, reading passages, instructions, vocabulary lists, and exercises. The data were analyzed using AntWordProfiler to identify the number of tokens, types, word frequency distribution, and vocabulary coverage based on the *General Service List* (GSL) and the *Academic Word List* (AWL). The findings revealed that high-frequency words dominated the textbook, accounting for 88% of the total corpus. Academic words appeared in only 1% of the corpus, while low-frequency or off-list words accounted for 11%. These findings indicate that most of the vocabulary in the textbook consists of common English words appropriate for elementary-level learners. However, the overall vocabulary coverage was still below the 95% threshold generally recommended for effective text comprehension. Therefore, although the textbook can be considered generally appropriate for elementary EFL students, teacher support is still necessary to help students understand thematic and low-frequency vocabulary presented in the learning materials.

Keywords: *Vocabulary Load, Vocabulary Distribution, Corpus-Based Analysis, English Textbook, Elementary EFL Students.*