



Please tick (✓)
Project Paper
Masters by Research
Master by Mixed Mode
PhD

INSTITUTE OF GRADUATE STUDIES

DECLARATION OF ORIGINAL WORK

This declaration is made on the 9th day of December 2025

i. Student's Declaration:

I, RONG CHANGSHENG, P20211000130, FBK (PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled A CORPUS-BASED STUDY ON THE ALIGNMENT OF READING TEXTS WITH HIGH SCHOOL ENGLISH CURRICULUM STANDARD, TEXTBOOKS AND EXAMINATIONS IN CHINA is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

Rong Changsheng
RONG CHANGSHENG

Signature of the student

ii. Supervisor's Declaration:

I, PROFESSOR DR. MAZURA @ MASTURA BINTI MUHAMMAD (SUPERVISOR'S NAME) hereby certifies that the work entitled A CORPUS-BASED STUDY ON THE ALIGNMENT OF READING TEXTS WITH HIGH SCHOOL ENGLISH CURRICULUM STANDARD, TEXTBOOKS AND EXAMINATIONS IN CHINA (TITLE) was prepared by the above named student, and was submitted to the Institute of Graduate Studies as a *~~partial~~/full fulfillment for the conferment of DOCTOR OF PHILOSOPHY (TEACHING ENGLISH AS A SECOND LANGUAGE) (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work.

9th December 2025

Date

Signature of the Supervisor

DR. MAZURA @ MASTURA BINTI MUHAMMAD
BERKAS
FAKULTI BAHASA DAN KOMUNIKASI
UNIVERSITI PENDIDIKAN SULTAN IDRIS
35900 TANJONG MALIM, PERAK

15/12/25



INSTITUT PENGAJIAN SISWAZAH / INSTITUTE OF GRADUATE STUDIES

**BORANG PENGESAHAN PENYERAHAN TESIS/DISERTASI/LAPORAN KERTAS PROJEK
DECLARATION OF THESIS/DISSERTATION/PROJECT PAPER FORM**

Tajuk / Title: A Corpus-Based Study on the Alignment of Reading Texts with High School English Curriculum Standard, Textbooks and Examinations in China

No. Matric /
Matric's No.: P20211000130

Saya / I am: Rong Changsheng
(Nama Pelajar / Student's Name)

mengaku membenarkan Tesis/Disertasi/Laporan Kertas Projek (Kedoktoran/Sarjana) ini disimpan di Universiti Pendidikan Sultan Idris (Perpustakaan Tuanku Bainun) dengan syarat-syarat kegunaan seperti berikut: -
acknowledged that Universiti Pendidikan Sultan Idris (Tuanku Bainun Library) reserves the right as follows: -

1. Tesis/Disertasi/Laporan Kertas Projek ini adalah hak milik Universiti Pendidikan Sultan Idris.
The Thesis/Dissertation/Project Paper is the property of Universiti Pendidikan Sultan Idris.
2. Perpustakaan Tuanku Bainun dibenarkan membuat Salinan untuk tujuan rujukan dan pendidikan.
Tuanku Bainun Library has the right to make copies for the purpose of reference and education.
3. Perpustakaan dibenarkan membuat salinan Tesis/Disertasi/Laporan Kertas Projek ini sebagai bahan pertukaran antara Institusi Pengajian Tinggi.
The library has the right to make copies of the Thesis/Dissertation/Project Paper for academic exchange.
4. Sila tandakan (✓) bagi pilihan kategori di bawah / Please tick (✓) for category below: -

☒	AKSES TERBUKA OPEN ACCESS	Saya bersetuju bahawa Tesis/Disertasi/Kertas Projek boleh diterbitkan sebagai salinan fizikal atau akses terbuka dalam talian. I agree that the Thesis/Dissertation/Project Paper may be published as a physical copy or open access online.
☐	*TERHAD *LIMITED	Mengandungi maklumat terhad kepada yang ditetapkan oleh organisasi/institusi di mana penyelidikan dijalankan (Perpustakaan akan menyekat akses sehingga _____ tahun). Contains restricted information as specified by the organization/institution where the research was conducted (Library will restrict access until _____ years).
☐	*SULIT *CONFIDENTIAL	Mengandungi maklumat sulit di bawah Akta Rahsia Rahim 1972. Contains confidential information under the Official Secrets Act 1972.
☐	EMBARGO	Tesis/Disertasi/Kertas Projek ini terikat kepada tempoh embargo selama 2 tahun atas sebab-sebab berikut: The Thesis/Dissertation/Project Paper is subject to an embargo period of 2 years for the following reason:

*Nota / *Notes

1. Jika tiada pilihan yang ditanda, pilihan pertama akan dilaksanakan secara automatik.
If no option is marked, the first option will be automatically implemented.

2. Jika Tesis/Disertasi/Kertas Projek ini dikategorikan sebagai SULIT atau TERHAD, sila lampirkan surat dari pihak berkuasa/organisasi yang berkaitan dengan menyatakan alasan dan tempoh tesis ini perlu diklasifikasikan sebagai TERHAD atau SULIT. / If the Thesis/Dissertation/Project Paper is categorized as CONFIDENTIAL or LIMITED, kindly attach a letter from the relevant authority/organization stating the reason and duration for the thesis should be classified as LIMITED or CONFIDENTIAL.

Rong Changsheng

RONG CHANGSHENG

(Tandatangan Pelajar/ Signature)

Tarikh/Date:

(Tandatangan Penyelia / Signature of Supervisor)
& (Nama & Cop Rasmi / Name & Official Stamp)

Tarikh/Date:

Dr. MARYA BINTI MAZURA @ MARYA BINTI MUHAMMAD BEKAN
FAKULTI BAHASA DAN KOMUNIKASI
UNIVERSITI PENDIDIKAN SULTAN IDRIS
35900 TANJONG MALIM, PERAK

15/1/23

A CORPUS-BASED STUDY ON THE ALIGNMENT
OF READING TEXTS WITH HIGH SCHOOL
ENGLISH CURRICULUM STANDARD,
TEXTBOOKS AND EXAMINATIONS
IN CHINA

 05-4506832  pustaka.upsi.edu.my  Perpustakaan Tuanku Bainun
Kampus Sultan Abdul Jalil Shah  PustakaTBainun  ptbupsi

RONG CHANGSHENG

SULTAN IDRIS EDUCATION UNIVERSITY

2025

**A CORPUS-BASED STUDY ON THE ALIGNMENT OF READING TEXTS WITH
HIGH SCHOOL ENGLISH CURRICULUM STANDARD, TEXTBOOKS
AND EXAMINATIONS IN CHINA**

RONG CHANGSHENG

THESIS PRESENTED TO QUALIFY FOR A DOCTOR OF PHILOSOPHY

**FACULTY OF LANGUAGES AND COMMUNICATION
SULTAN IDRIS EDUCATION UNIVERSITY**

2025



Please tick (✓)

Project Paper

Masters by Research

Master by Mixed Mode

PhD

✓

INSTITUTE OF GRADUATE STUDIES

DECLARATION OF ORIGINAL WORK

This declaration is made on the9thday of ...December..2025

i. Student's Declaration:

I, RONG CHANGSHENG, P20211000130, FBK (PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled A CORPUS-BASED STUDY ON THE ALIGNMENT OF READING TEXTS WITH HIGH SCHOOL ENGLISH CURRICULUM STANDARD, TEXTBOOKS AND EXAMINATIONS IN CHINA is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

RONG CHANGSHENG

Signature of the student

ii. Supervisor's Declaration:

I PROFESSOR DR. MAZURA @ MASTURA BINTI MUHAMMAD (SUPERVISOR'S NAME) hereby certifies that the work entitled A CORPUS-BASED STUDY ON THE ALIGNMENT OF READING TEXTS WITH HIGH SCHOOL ENGLISH CURRICULUM STANDARD, TEXTBOOKS AND EXAMINATIONS IN CHINA (TITLE) was prepared by the above named student, and was submitted to the Institute of Graduate Studies as a ~~*-partial~~/full fulfillment for the conferment of DOCTOR OF PHILOSOPHY (TEACHING ENGLISH AS A SECOND LANGUAGE) (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work.

9th December, 2025

Date

Signature of the Supervisor



INSTITUT PENGAJIAN SISWAZAH / INSTITUTE OF GRADUATE STUDIES

BORANG PENGESAHAN PENYERAHAN TESIS/DISERTASI/LAPORAN KERTAS PROJEK DECLARATION OF THESIS/DISSERTATION/PROJECT PAPER FORM

Tajuk / Title: A Corpus-Based Study on the Alignment of Reading Texts with High School English Curriculum Standard, Textbooks and Examinations in China

No. Matrik / Matric's No.: P20211000130

Saya / I am: Rong Changsheng

(Nama Pelajar / Student's Name)

mengaku membenarkan Tesis/Disertasi/Laporan Kertas Projek (Kedoktoran/Sarjana) ini disimpan di Universiti Pendidikan Sultan Idris (Perpustakaan Tuanku Bainun) dengan syarat-syarat kegunaan seperti berikut: -
acknowledged that Universiti Pendidikan Sultan Idris (Tuanku Bainun Library) reserves the right as follows: -

1. Tesis/Disertasi/Laporan Kertas Projek ini adalah hak milik Universiti Pendidikan Sultan Idris.
The Thesis/Dissertation/Project Paper is the property of Universiti Pendidikan Sultan Idris.
2. Perpustakaan Tuanku Bainun dibenarkan membuat Salinan untuk tujuan rujukan dan pendidikan.
Tuanku Bainun Library has the right to make copies for the purpose of reference and education.
3. Perpustakaan dibenarkan membuat salinan Tesis/Disertasi/Laporan Kertas Projek ini sebagai bahan pertukaran antara Institusi Pengajian Tinggi.
The library has the right to make copies of the thesis/Dissertation/Project Paper for academic exchange.
4. Sila tandakan (✓) bagi pilihan kategori di bawah / Please tick (✓) for category below: -

✓	AKSES TERBUKA OPEN ACCESS	Saya bersetuju bahawa Tesis/Disertasi/Kertas Projek boleh diterbitkan sebagai salinan fizikal atau akses terbuka dalam talian. <i>I agree that the Thesis/Dissertation/Project Paper may be published as a physical copy or open access online.</i>
	*TERHAD *LIMITED	Mengandungi maklumat terhad seperti yang dinyatakan oleh organisasi/institusi di mana penyelidikan dijalankan. (Perpustakaan akan menyekat akses sehingga ____ tahun). <i>Contains restricted information as specified by the organization/institution where the research was conducted. (Library will restrict access until ____ years).</i>
	*SULIT *CONFIDENTIAL	Mengandungi maklumat sulit di bawah Akta Rahsia Rasmi 1972. <i>Contains confidential information under the Official Secrets Act 1972.</i>
	EMBARGO	Tesis/Disertasi/Kertas Projek ini tertakluk kepada tempoh embargo selama 2 tahun atas sebab-sebab berikut: This Thesis/Dissertation/Project Paper is subject to an embargo period of 2 years for the following reasons:

*Nota / *Notes:

- i. Jika tiada pilihan yang ditandai, pilihan pertama akan dilaksanakan secara automatik.
If no option is marked, the first option will be automatically implemented.
- ii. Jika Tesis/Disertasi/Kertas Projek ini dikategorikan sebagai **SULIT** atau **TERHAD**, sila lampirkan surat dari pihak berkuasa/organisasi yang berkaitan dengan menyatakan alasan dan tempoh tesis ini perlu diklasifikasikan sebagai **TERHAD** atau **SULIT**. / If the Thesis/Dissertation/Project Paper is categorized as **CONFIDENTIAL** or **LIMITED**, kindly attach a letter from the relevant authority/organization stating the reason and duration for the thesis should be classified as **LIMITED** or **CONFIDENTIAL**.

RONG CHANGSHENG

(Tandatangan Pelajar/ Signature)

(Tandatangan Penyelia / Signature of Supervisor)
& (Nama & Cop Rasmi / Name & Official Stamp)

Tarikh/Date:

Tarikh/Date:



ACKNOWLEDGMENTS

It is a pleasure to thank the many people who made this paper possible. First of all thanks to my supervisor, Dr Mazura Mastura Binti Muhammad, thanks for her help and support from beginning to end. Thanks also to all the examiners for their helpful suggestions and comments. Special thanks to all my family members for their support and encouragement.



ABSTRACT

The aim of the study is to examine the alignment among senior high school English curriculum standards, textbooks, and college entrance examination papers in Mainland China. It employs corpus linguistics software (LancsBox, Sketch Engine, Voyant Tools) and Porter's SEC Model. The study analyzes the high school English curriculum standards, high school English textbooks (seven volumes), and college entrance examination English test papers (2018-2022). Findings indicate that the high school English curriculum standard, high school English textbooks and college entrance examination English test papers are relatively consistent in terms of vocabulary scope, and the overall content scope meets the alignment requirements. The four elements related to the high school English curriculum standards, high school English textbooks, and college entrance examination English test papers in this study are (1) Keywords, (2) Lexical density, (3) Readability, and (4) Alignment index. Based on the findings it could be concluded that, corpus research approaches are able to measure the high school English curriculum standard, the vocabulary range and content writing range of high school English textbooks and college entrance examination English test papers and improves improve the accuracy and quality of English teaching material development. This research offers valuable insights for policymakers, educators, and materials developers, potentially improving the coherence of curriculum design, textbook development, and assessment creation.

ANALISIS KORPUS TERHADAP KESELARASAN STANDARD KURIKULUM, BUKU TEKS DAN PEPERIKSAAN BAHASA INGGERIS SEKOLAH MENENGAH DI CHINA

ABSTRAK

Tujuan kajian ini adalah untuk mengkaji keselarasan antara standard kurikulum Bahasa Inggeris sekolah menengah atas, kandungan buku teks dan kertas peperiksaan bagi kemasukan kolej di China. Kajian ini menggunakan perisian linguistik korpus (*LancsBox*, *Sketch Engine*, *Voyant Tools*) dan Model SEC Porter. Kajian ini menganalisis dokumen standard kurikulum Bahasa Inggeris sekolah menengah atas, tujuh jilid buku teks, dan kertas peperiksaan Bahasa Inggeris kemasukan kolej tahun 2018 hingga 2022. Dapatan kajian menunjukkan bahawa standard kurikulum Bahasa Inggeris sekolah menengah atas, buku teks Bahasa Inggeris, dan kertas peperiksaan Bahasa Inggeris kemasukan kolej adalah konsisten dari segi skop kosa kata, dan skop keseluruhan kandungan memenuhi keperluan keselarasan. Empat elemen yang berkaitan dengan standard kurikulum Bahasa Inggeris sekolah menengah atas, buku teks Bahasa Inggeris, dan kertas ujian Bahasa Inggeris kemasukan kolej dalam kajian ini ialah, yang pertama Kata kunci, kedua, Ketumpatan leksikal, ketiga Kebolehbacaan, dan keempat Indeks keselarasan. Kajian ini mendapati analisis korpus dapat menentukan standard kurikulum Bahasa Inggeris sekolah menengah, julat perbendaharaan kata dan julat penulisan kandungan buku teks serta kertas peperiksaan Bahasa Inggeris kemasukan kolej. Dapatan menunjukkan ketepatan serta kualiti pembangunan bahan pengajaran Bahasa Inggeris. Penyelidikan ini menawarkan pandangan berharga kepada pembuat dasar, pendidik, dan pembangun bahan, serta berpotensi meningkatkan keseragaman dalam reka bentuk kurikulum, pembangunan buku teks, dan pembinaan pentaksiran.

TABLE OF CONTENTS

	Page
DECLARATION OF ORIGINAL WORK	ii
DECLARATION OF THESIS	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
ABSTRAK	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	xiii
LIST OF FIGURES	xx
LIST OF ABBREVIATIONS	xxvii

CHAPTER 1 INTRODUCTION

1.1	Introduction	1
1.2	Background of the Study	2
1.2.1	A brief history of China's Gaokao (College Entrance Examination)	3
1.2.2	English Test Papers of the Gaokao in this study	6
1.2.3	A brief introduction to English Curriculum Standards in Mainland China	7
1.2.4	The high school English textbooks in this study	12
1.2.5	Reading Skill Requirements	14
1.3	Statement of Problem	17
1.3.1	The Importance of English Language Education in China	18

1.3.2	The importance of the Gaokao Examination in China	19
1.3.3	Lack of Publications on Curriculum Alignment	22
1.3.4	The importance of the curriculum alignment	25
1.4	Aims of the Study	27
1.5	Research Objectives	27
1.6	Research Questions	28
1.7	Definition of Terms	29
1.7.1	High School	29
1.7.2	English Textbooks	30
1.7.3	Reading Texts	31
1.7.4	College Entrance Examinations (Gaokao)	31
1.7.5	Test Papers	33
1.7.6	Curriculum Standards	33
1.7.7	Reading comprehension	34
1.7.8	Keywords	36
1.7.9	Lexical Density	38
1.7.10	Readability	39
1.7.11	Corpus	41
1.7.12	Corpus-based	42
1.7.13	Corpus Linguistics	43
1.7.14	Alignment	44
1.8	Limitations of Study	47
1.9	Significance of Study	51
1.10	Summary	52

CHAPTER 2 LITERATURE REVIEW

2.1	Introduction	53
2.2	Alignment Theoretical Research of this Study	54
2.2.1	Krashen's Input Hypothesis (1985)	57
2.2.2	Ellis' Frequency Theory (2002)	61
2.2.3	VanPatten's Input Processing Theory (2004)	65
2.2.4	Anderson's Cognitive Theory (1983)	68
2.3	Previous Research on Alignment	74
2.3.1	Definition of Alignment	74
2.3.2	Study Related to Alignment Approaches	77
2.3.3	Operating Process of SEC Model (Surveys of Exact Curriculum)	80
2.4	Corpus Research	82
2.4.1	Definition of the Corpus and Corpus Linguistics	83
2.4.2	The History of English corpus	85
2.4.3	The Application of Corpus Method in Language Textbooks	90
2.4.4	The Application of Corpus Method in Curriculum Standards	94
2.4.5	The Application of Corpus Method in Test Papers/Examination	96
2.4.6	Research on the Use of Corpus on Textbooks, Curriculum Standards and Examinations	98
2.4.7	Main Types of Corpora	110
2.4.8	Corpus-based Approach	112
2.5	Review of Previous Studies and Research Gaps	115
2.5.1	Previous Studies on Alignment of Curriculum, Textbooks, and Test papers	115

2.5.2	Research on English Reading Comprehension	118
2.5.3	Research on English Learning Textbooks	124
2.5.4	Research on English Language Test	135
2.5.5	Comparison of English Teaching Materials and English Test Contents	141
2.5.6	Research Gaps	143
2.6	Summary	145

CHAPTER 3 METHODOLOGY

3.1	Introduction	146
3.2	Research Framework	147
3.3	Data Description	153
3.3.1	The High School English Curriculum Standards Corpus	153
3.3.2	The China Corpus of English Textbooks (CCETB)	157
3.3.3	The China Corpus of English Test Papers	163
3.4	Data Design	164
3.5	Data Collection Instruments	167
3.5.1	The High School English Curriculum Standards Corpus	168
3.5.2	The China Corpus of English Textbooks (hereafter CCETB)	171
3.5.3	The China Corpus of English Test Papers (CCETP)	176
3.5.4	British National Corpus 1994 Baby edition (BNC 1994 Baby) Corpus --The Reference Corpus	179
3.6	Data Analysis	185
3.6.1	Data Analysis Tools and How the Data are Analyzed	185

3.6.2	Lexical Density of the Reading texts in English textbooks and Gaokao examinations	220
3.6.3	Readability Index of the Reading texts in English textbooks and Gaokao examinations	234
3.6.4	Framework of Research Methodology	268
3.7	Summary	271

CHAPTER 4 FINDINGS

4.1	Introduction	273
4.2	Research Question 1	275
4.2.1	Overall Data Description	275
4.2.2	Uploading the four corpora	276
4.2.3	Counting the top 100 high-frequency words and their proportion of parts of speech in the four corpora	282
4.3	Research Question 2	355
4.4	Research Question 3	373
4.5	Research Question 4	384
4.5.1	Calculation of the proportion of three two-dimensional tables of cognitive levels	384
4.5.2	Using Curriculum Standard as a yardstick to align the High School English Curriculum Standard, College Entrance Examination English Test Papers and the High School English Textbooks	404
4.6	Discussion	416

CHAPTER 5 CONCLUSION AND RECOMMENDATIONS

5.1	Introduction	419
5.2	Summary of Findings	422
5.3	Conclusion	428

5.4	Discussion	433
5.4.1	Research Question 1	433
5.4.2	Research Question 2	438
5.4.3	Research Question 3	442
5.4.4	Research Question 4	446
5.4.5	Summary of Research Findings	451
5.5	Theoretical Implications and Pedagogical Implications	453
5.5.1	Theoretical Implications	454
5.5.2	Pedagogical Implications	459
5.6	Limitations	461
5.7	Future Research and Recommendations	465
5.7.1	Expansion in research objects	466
5.7.2	Expansion in education levels	466
5.7.3	Expansion in geographical scope	466
5.7.4	Expansion in the depth of research	467
	REFERENCES	470
	APPENDICES	



LIST OF TABLES

Table No.		Page
1.1	Requirements for the Content of the Theme Context of English Curriculum in High School	10
1.2	Themes in Compulsory Textbooks	13
1.3	Themes in Optional Compulsory Textbooks	13
1.4	The Components of Gaokao Examination and Weighting	15
1.5	Comparison between China population and Gaokao Test Takers	21
2.1	Comparative Table Of The Alignment Analysis Models	79
2.2	List of Five-Generation English Language Corpora	86
2.3	Sixteen Tenses in English Grammar	105
2.4	Five types of clauses	105
2.5	Six Types of Phrases	106
2.6	Six Types of Sentence Structures	106
2.7	Four Types of Questions	107
2.8	Three Subjunctive Mood Sentence Patterns	107
2.9	Eight misuse of English Language in CET Band 6	108
2.10	Three Common Passive Voices in CET Band 6	108
3.1	Sample of content on reading section the high school English Curriculum Standards	156
3.2	The High School English Textbooks	157
3.3	The composition compulsory textbooks (CTB) in the CCETB	158
3.4	The composition of optional compulsory textbooks (OCTB) in the CCETB	158



3.5	The topics of reading texts in compulsory textbooks (CTB) in the CCETB	159
3.6	The topics of reading texts in Optional Compulsory Textbooks (OCTB) in the CCETB	160
3.7	The Composition of the English test papers (CCETP)	163
3.8	The Composition of compulsory textbooks (CTB) in the CCETB	174
3.9	The Composition of optional compulsory textbooks (OCTB) in the CCETB	174
3.10	The Composition of two types of textbooks in the CCETB	176
3.11	The Composition of the English test papers (CCETP)	177
3.12	The Composition of the China Corpus English test papers (CCETP)	179
3.13	The Composition of Newspaper Texts in BNC-1994baby	182
3.14	The Composition of Spontaneous Conversation in BNC-1994baby	183
3.15	The Composition of the BNC-1994 baby Corpus (four categories)	184
3.16	The Composition of the BNC 1994 baby Corpus	184
3.17	Corpus 5 - Corpus 6 keyword comparison	198
3.18	The text of the top 100 positive keywords in the Corpus 5 - Corpus 6 keyword comparison	201
3.19	Part-of-speech proportions of the top 100 keywords in the Corpus 5 - Corpus 6 keyword comparison	202
3.20	Standard Requirements	218
3.21	Test Results	218
3.22	Standard Requirements	219
3.23	Test Results	219
3.24	2018-2022 college entrance examination English test structure setting	263

3.25	Number of knowledge points in curriculum standards	264
3.26	Ratio of knowledge points in curriculum standards	265
3.27	Framework of Research Methodology	270
4.1	Basic information of the four corpora	278
4.2	The wordlist of the BNC 1994 baby (top 100 words)	283
4.3	Part-of-speech proportions of the top 100 words in the BNC 1994 baby	287
4.4	The wordlist of the High School Curriculum Standards (HSECSC)(top 100 words)	289
4.5	Part-of-speech proportions of the top 100 words in the High School Curriculum Standards (HSECSC)	293
4.6	The wordlist of the High School English Textbooks (CCETB) (top 100 words)	295
4.7	Part-of-speech proportions of the top 100 words in the High School English Textbooks (CCETB)	298
4.8	The wordlist of the College Entrance Examination (CCETP) (top 100 words)	300
4.9	Part-of-speech proportions of the top 100 words of the College Entrance Examination (CCETP)	304
4.10	The percentage of content words and function words in the top 100 of the four corpora	306
4.11	Summary of the Top 100 Wordlist.	308
4.12	Statistics of the number of different words and their text coverage	309
4.13	Corpus 1 - Corpus 2 keyword comparison - (Truncated [100])	313
4.14	The number of each part of speech of the keyword (Corpus 1 - Corpus 2 comparison)	317
4.15	Corpus 1 - Corpus 3 keyword comparison - (Truncated [100])	318
4.16	The number of each part of speech of the keyword (Corpus 1 - Corpus 3 comparison)	322

4.17	Corpus 1 - BNC1994-baby keyword comparison - (Truncated [100])	324
4.18	Compare the negative and positive keywords and lockwords in the Corpus 1 - BNC1994-baby keyword comparison - (Truncated [100])	328
4.19	The number of each part of speech of the keyword (Corpus 1 - BNC1994-baby comparison)	329
4.20	Corpus 2 - Corpus 3 keyword comparison - (Truncated [100])	330
4.21	Compare the negative and positive keywords in the Corpus 2 - Corpus 3 keyword comparison - (Truncated [100])	334
4.22	The number of each part of speech of the keyword (Corpus 2- Corpus 3 comparison)	336
4.23	Corpus 2 - BNC1994-baby keyword comparison - (Truncated [100])	338
4.24	Compare the negative and positive keywords and lockwords in the Corpus 2 - BNC 1994 baby keyword comparison - (Truncated [100])	341
4.25	The number of each part of speech of the keyword (Corpus 2- BNC1994-baby comparison)	342
4.26	Corpus 3 - BNC1994-baby keyword comparison - (Truncated [100])	344
4.27	Compare the negative and positive keywords and lockwords in the Corpus 3 - BNC 1994 baby keyword comparison - (Truncated [100])	347
4.28	The number of each part of speech of the keyword (Corpus 3- BNC 1994 baby comparison)	348
4.29	Compare the positive keywords in the Corpus 1 - BNC, Corpus 2 - BNC and Corpus 3 - BNC keyword comparison - (Truncated [100])	349
4.30	Alignment and Misalignment of Positive Keywords by Part of Speech	351
4.31	Compare the negative keywords in the Corpus 1 - BNC, Corpus 2 - BNC and Corpus 3 - BNC keyword comparison - (Truncated [100])	352

4.32	Compare the lockwords in the Corpus 1 - BNC, Corpus 2 - BNC and Corpus 3 - BNC keyword comparison - (Truncated [100])	354
4.33	The Lexical Density of Texts in the High School English Textbooks (35 texts)	355
4.34	The lexical density of the High School English textbooks (seven volumes)	359
4.35	The Lexical Density of Texts in the College Entrance Examination (2018-2022)	369
4.36	4.36 The Lexical density of the College Entrance Examination (2018-2022)	370
4.37	Compare the average Lexical density of three corpora: Corpus 1 (Curriculum Standard), Corpus 2 (Textbooks), and Corpus 3 (Test papers)	371
4.38	The Readability Index of texts in the High School English textbooks (35 texts)	373
4.39	ARI Score to Grade Level and Age Mapping	377
4.40	The Readability index of the 25 texts of the College Entrance Examination English test papers (2018-2022)	379
4.41	The average Readability index of the textbooks and test papers	382
4.42	Summary of Readability Index	383
4.43	Number of knowledge points in the High School English curriculum standard	385
4.44	Ratio of knowledge points in the High School English curriculum standard	391
4.45	The score value of each content subject in the 2018 college entrance examination English test paper	393
4.46	The score ratio of each content subject in the 2018 college entrance examination English test paper	394
4.47	The score value of each content subject in the 2019 college entrance examination English test paper	395

4.48	The score ratio of each content subject in the 2019 college entrance examination English test paper	395
4.49	The score value of each content subject in the 2020 college entrance examination English test paper	396
4.50	The score ratio of each content subject in the 2020 college entrance examination English test paper	397
4.51	The score value of each content subject in the 2021 college entrance examination English test paper	397
4.52	The score ratio of each content subject in the 2021 college entrance examination English test paper	398
4.53	The score value of each content subject in the 2022 college entrance examination English test paper	398
4.54	The score ratio of each content subject in the 2022 college entrance examination English test paper	399
4.55	Comparison of cognitive levels distribution in the reading part of the college entrance examination English test paper in 5 years (2018-2022)	399
4.56	Comparison of Course contents distribution in the reading part of the college entrance examination English test paper in 5 years (2018-2022)	400
4.57	The score value of each content theme in the High School English textbooks	402
4.58	Score ratio of each content theme in High School English textbooks	403
4.59	P value of the alignment coefficient between test papers and curriculum standard	404
4.60	Cognitive levels distribution of college entrance examination English test papers in 2018-2022	406
4.61	Comparison between the content and theme distribution of the English test papers of the 2018-2022 college entrance examination and the content and theme distribution of the curriculum standard	409
4.62	P value of the alignment coefficient between textbooks and curriculum standard	411

4.63	Cognitive Level Distribution between Textbooks and Curriculum Standard	412
4.64	The distribution of course content between textbooks and curriculum standard	414
5.1	Comparison of Keyword Alignment and Misalignment Between Corpora	437
5.2	Comparison of Lexical Density Alignment and Misalignment Between the Corpora	441
5.3	Comparison of Readability Index Alignment and Misalignment Between the Corpora	446
5.4	Comparison of Cognitive Levels: Alignment and Misalignment Between the Corpora	450
5.5	Summary of Research Findings by Research Questions	453



LIST OF FIGURES

Figure No.		Page
1.1	The high school English curriculum standard, high school English textbooks, and English test papers/examinations for college entrance examinations	2
2.1	The specific research methods of this study	55
2.2	The relationship between four theories and language acquisition	55
2.3	The Alignment Research with SEC Model	56
2.4	The relationship between four theories and language acquisition	70
2.5	Alignment of test papers and curriculum standard, textbooks	73
2.6	Corpus Research's step-by-step process	83
3.1	Theoretical assumptions of this study	149
3.2	The framework of second language learning theories	151
3.3	Sample of Wordlist from the appendix of the high school English Curriculum Standards	156
3.4	Sample of a reading text in compulsory textbooks (CTB) in the CCETB	160
3.5	Sample of a reading text in optional compulsory textbooks (OCTB) in the CCETB	162
3.6	Sample of a reading text the China Corpus of English Test Papers (CCETP)	164
3.7	The Research Design	165
3.8	Sample of Vocabulary list in the Curriculum Standard for High School	169





3.9	Step-By-Step Measures Taken to Assemble the Curriculum Standard for High School Corpus	170
3.10	Course Contents of Curriculum Standards for High School	171
3.11	Step-By-Step Measures Taken to Assemble The CCETB	173
3.12	The main production process of CCETP	178
3.13	LancsBox 6.0	186
3.14	LancsBox	188
3.15	LancsBox (corpus)	189
3.16	LancsBox (BNC 1994 baby) Download	190
3.17	LancsBox (Initializing)	190
3.18	LancsBox (BNC 1994 baby) Installed	191
3.19	LancsBox Import	192
3.20	LancsBox (reference corpus) Installed	192
3.21	(Corpora 5-8)	196
3.22	Wizard	197
3.23	Running	198
3.24	Part-of-speech proportions of the top 100 keywords in the Corpus 5 - Corpus 6 keyword comparison	203
3.25	Sketch Engine	206
3.26	Sketch Engine (POS)	207
3.27	Sketch Engine (Corpus)	207
3.28	Sketch Engine (Compile)	208
3.29	Sketch Engine (I have my own texts)	209
3.30	Sketch Engine (upload files)	209
3.31	Sketch Engine (choose files)	210



3.32	Sketch Engine (corpus content)	211
3.33	Sketch Engine (BNC 1994 baby is compiled)	211
3.34	Sketch Engine (BNC 1994 baby is compiled and ready to use)	212
3.35	Sketch Engine (BNC 1994 baby ready to use)	213
3.36	Sketch Engine (BNC 1994 baby -Wordlist)	213
3.37	Sketch Engine (BNC 1994 baby - Wordlist -POS- noun list)	214
3.38	Voyant tools	215
3.39	Online-Utility	216
3.40	Voyant tools	222
3.41	Voyant tools-2.1.txt	223
3.42	Voyant tools-2-2.txt	224
3.43	Voyant tools	226
3.44	Voyant tools-1.1.txt	227
3.45	Voyant tools-1.2.txt	229
3.46	Voyant tools-1.3.txt	230
3.47	Voyant tools-1.4.txt	232
3.48	Voyant tools-1.5.txt	234
3.49	Online-Utility.txt	236
3.50	Online-Utility Process text.txt	238
3.51	Online-Utility 2-3.txt	239
3.52	Online-Utility 2-4.txt	241
3.53	Online-Utility Optional 1-1.txt	244

3.54	Online-Utility Optional 1-2.txt	246
3.55	Online-Utility Optional 1-3.txt	247
3.56	Online-Utility Optional 1-4.txt	249
3.57	Online-Utility Optional 1-5.txt	251
3.58	Online-Utility 2022 A.txt	254
3.59	Online-Utility 2022 B.txt	256
3.60	Online-Utility 2022 C.txt	258
3.61	Online-Utility 2022 D.txt	260
3.62	Online-Utility 2022 E.txt	261
3.63	The main steps and processes studied in this section	266
3.64	The framework for the alignment analysis	267
3.65	The production process of CCETB and CCETP	268
4.1	Summary of the Four (4) Corpora	276
4.2	The research report has been created	277
4.3	Go to the report-findings	278
4.4	Corpus 1 (Curriculum Standards) and Corpus 2 (Textbooks) Comparison of Corpus Size	279
4.5	Corpus 2 (Textbooks) and Corpus 3 (Test papers) Comparison of Corpus Size	280
4.6	Corpus 1 (Curriculum Standard) and Corpus 3 (Test papers) Comparison Size	281
4.7	Part-of-speech proportions of the top 100 words in the BNC 1994 baby	288
4.8	Part-of-speech proportions of the top 100 words in the High School Curriculum Standard (HSECSC)	293
4.9	Part-of-speech proportions of the top 100 words in the High School English Textbooks (CCETB)	299

4.10	Part-of-speech proportions of the top 100 words of the College Entrance Examination (CCETP)	305
4.11	Statistics of the relationship between the growth of words and their growth in text coverage	311
4.12	The number of each part of speech of the keywords (Corpus 1 - Corpus 2 comparison)	317
4.13	The number of each part of speech of the keyword (Corpus 1 - Corpus 3 comparison)	322
4.14	Compare the negative and positive keywords and lockwords in the Corpus 1 - BNC1994-baby keyword comparison - (Truncated [100])	328
4.15	The number of each part of speech of the keyword (Corpus 1 - BNC1994-baby comparison)	330
4.16	Compare the negative and positive keywords in the Corpus 2 - Corpus 3 keyword comparison - (Truncated [100])	334
4.17	The number of each part of speech of the keyword (Corpus 2-Corpus 3 comparison)	336
4.18	Compare the negative and positive keywords and lockwords in the Corpus 2 - BNC 1994 baby keyword comparison - (Truncated [100])	341
4.19	The number of each part of speech of the keyword (Corpus 2-BNC1994-baby comparison)	343
4.20	Compare the negative and positive keywords and lockwords in the Corpus 3 - BNC 1994 baby keyword comparison - (Truncated [100])	347
4.21	The number of each part of speech of the keyword (Corpus 3- BNC 1994 baby comparison)	348
4.22	Compare the positive keywords in the Corpus 1 - BNC, Corpus 2 - BNC and Corpus 3 - BNC keyword comparison - (Truncated [100])	350
4.23	Compare the negative keywords in the Corpus 1 - BNC, Corpus 2 - BNC and Corpus 3 - BNC keyword comparison - (Truncated [100])	353



4.24	Compare the lock keywords in the Corpus 1 - BNC, Corpus 2 - BNC and Corpus 3 - BNC keyword comparison - (Truncated [100])	354
4.25	The Lexical Density of Texts in the High School English Textbooks (35 texts)	358
4.26	The lexical density of the High School English textbooks (seven volumes)	360
4.27	Voyant tools-2018 A.txt	363
4.28	Voyant tools-2018 B.txt	365
4.29	Voyant tools-2018 C.txt	366
4.30	Voyant tools-2018 D.txt	367
4.31	Voyant tools-2018 E.txt	368
4.32	The Lexical Density of Texts in the College Entrance Examination (2018-2022)	370
4.33	The Lexical density of the College Entrance Examination (2018-2022)	371
4.34	Compare the average Lexical density of three corpora: Corpus 1 (Curriculum Standard), Corpus 2 (Textbooks), and Corpus 3 (Test papers)	372
4.35	The Readability Index of texts in the High School English textbooks (35 texts)	375
4.36	The Readability index of the texts of the College Entrance Examination English test papers (2018-2022)	380
4.37	The process of making three two-dimensional tables	384
4.38	Integrated icons for the six major elements of course content	388
4.39	Text types and their three levels of content requirements	389
4.40	Comparison of cognitive levels distribution in the reading part of the college entrance examination English test paper in 5 years (2018-2022)	400



4.41	Comparison of Course contents distribution in the reading part of the college entrance examination English test paper in 5 years (2018-2022)	401
4.42	2018-2022 College Entrance Examination English test papers and curriculum standards alignment coefficient P value	405
4.43	Proportion distribution of cognitive level of English test papers of college entrance examination in 2018-2022	407
4.44	Comparison between the course content of the 2018-2022 College Entrance Examination English Test Paper and the Course Content Distribution of the Curriculum Standard	409
4.45	P value of alignment coefficient between English textbooks and curriculum standard	411
4.46	P value bar chart of alignment coefficient Textbooks and Curriculum Standard	412
4.47	P value of the distribution of course content between textbooks and curriculum standard	414

LIST OF ABBREVIATIONS

ACT	Anderson’s Cognitive Theory
ANOVA	Analysis of Variance
AWL570	Academic Word List
BISE	Board of Intermediate and Secondary Education
BNC	British National Corpus
BoE	Bank of English
CCETB	China Corpus of English Textbooks
CCETP	China Corpus of English Test papers
CCSS	Common Core State Standards
CEM	Corpus for English majors
CET	College English Test
CHILDES	Child Language Data Exchange System
CLI	Content and Language Integration
CLT	Communicative Language Teaching
COVID-19 pandemic	a global pandemic of coronavirus disease 2019
CTB	compulsory textbooks
D-level analyzer	syntactic developmental level analyzer
FLTRP	Foreign Language Teaching and Research Press
GSL2000	General Service List 2000
IELTS	International English Language Test System
L1	First Language

L2	Second Language
L2SCA	Second Language Syntactic Complexity Analyzer
LAD	Language Acquisition Device
LDOCE	Longman Dictionary of Contemporary English
MANOVA	Multivariate Analysis of Variance
MEXT	Japanese Ministry of Education, Culture, Sports, Science and Technology
ML	Modern Language
NETS	National English Test System
NMET	National Matriculation English Test
OCTB	optional compulsory textbooks
PEP	People's Education Press
PETS	Public English Test System
SEC Model	Survey of Exact Curriculum Alignment Analysis Method
SLA	The Second Language Acquisition
SLOs	students' learning outcomes
SPSS	Statistical Package for the Social Sciences
TAACO	Tool For The Automatic Analysis Of Cohesion
TAALED	Tool for the Automatic Analysis of Lexical Diversity
TAALES	Tool for the Automatic Analysis of Lexical Sophistication
TAASSC	Tool for the Automatic Analysis of Syntactic Sophistication and Complexity
TEM	Test for English Majors
TOEFL	Test of English as a Foreign Language

US

United States

USA

the United States of America



CHAPTER 1

INTRODUCTION



1.1 Introduction

This chapter describes the research background, problem statement, research objectives, research questions, definitions of terms, limitations of research as well as the significance of the research which are the fundamental elements underpinning the current study.

The next section highlights the background of the study.



1.2 Background of the Study

This study was conducted to provide solutions for improving students' learning efficiency, specifically, improving their English scores in the college entrance examination and their ability to use English in practice. Three (3) main objects of the current study are the high school English curriculum standard, high school English textbooks, and English test papers for college entrance examinations, informally known as the Gaokao (refer to Figure 1.1). Hence, it would be apt that this section paints the background of these three objects. In addition, it will also explore reading requirements which will also be the focal point of the current study.

Figure 1.1

The high school English curriculum standard, high school English textbooks, and English test papers/examinations for college entrance examinations



Brief history of China's college entrance examination (henceforth Gaokao), the number of English language learners from China, English language policy in China,

the High School English Curriculum Standards and the high school English textbooks and college entrance examination English test papers will be highlighted in this segment. These subtopics will provide an overall picture that embodies the current study.

1.2.1 A brief history of China's Gaokao (College Entrance Examination)

This section offers an overview of the “Gaodeng Jiaoyu Kaoshi,” which translates to Higher Education Examination. It underlines how the Gaokao examination structure has evolved in addressing the needs of China for a more standardized examination to measure the academic achievement of the students, while also highlighting the prominence of English as a core subject in the Gaokao examination.

The Gaokao examination was officially introduced in the 1950s following the establishment of the People’s Republic of China. It was initially designed to unify university admissions in replace of decentralized assessment models that varied across the provinces. By implementing the Gaokao examination, the government of China was able to streamline all candidate selection for higher education (Ministry of Education of the People’s Republic of China [MOE], 2021).

Unfortunately, amid the Cultural Revolution (1966–1976), the implementation of the Gaokao examination came to a halt due to the inevitable shift of political and social priorities. It was reinstated on October 21, 1977 and this significantly marked a turning point in Chinese education policy where it reaffirms the role of standardized evaluations for higher education admissions (Zhao & Ding, 2020). Over the decades, adjustments have been made to the structure, content, and administration of the Gaokao examination to align with the rapid economic and social changes occurring in China. Recent reformations included more flexible subject choices, diversified testing formats, and greater emphasis on analytical and problem-solving abilities rather than overemphasizing the traditional rote learning (Gu & Lin, 2019).

For example, in the 1990s, the examination had evolved into a “3+2” model. three (3) signifies three (3) core subjects (Chinese, Mathematics, and a foreign language, typically English); while two (2) indicates of electives. Later, the “3+2” model was expanded to the “3+X” model. This model has a range of electives, which were introduced to accommodate a broader spectrum of students’ academic interests (Gu & Lin, 2019). In addition to this transformation, the provincial authorities were gradually given the flexibility to introduce pilot programs featuring alternative testing formats, selective weighting of subjects, and supplementary assessments including interviews and project-based tasks (Zhao & Ding, 2020).

Currently, the Gaokao examination has firmly positioned itself as the most extensive and competitive standardized examination in China. Despite these transformations, the English subject has remained as a mandatory or main foreign language for the Gaokao examination, and this reflects its status as a vital skill for learners to master in order to engage with the global community.

However, through the years many provinces have gradually increased the proportion of marks assigned to the listening skills, and in some experimental programs, speaking sections are also included. For instance, the province of Shanghai incorporated an oral English assessment as part of its local Gaokao format since 2002, which aims to evaluate students' practical language skills rather than relying solely on their reading and writing abilities (Zhao & Ding, 2020). These transformations aim to foster a more comprehensive assessment of students' language proficiency beyond mere rote grammar knowledge (MOE, 2021).

The emphasis of the English subject in the Gaokao examination underscores China's emphasis on international communication and global competitiveness, as policy makers continue to adapt the Gaokao to produce graduates equipped for the demands of a modern, interconnected world (MOE, 2021).

This section will further be segmented into the Gaokao English test papers.

1.2.2 English Test Papers of the Gaokao in this study

As aforementioned, The Gaokao examination of the People's Republic of China is mainly divided into two stages: 1949-1977 and 1977-2025.

The first stage occurred prior to the Cultural Revolution era. The examination was established after the inception of the People's Republic of China and its chief objective was to promote a centralized merit-based University admission to students of China. The Gaokao examination had gone through several adjustments; from individual enrollment to joint enrollment and then to unified enrollment. However, during the Cultural Revolution (1966–1977) the Gaokao was suspended. And during the revolution, university admissions became highly politicized. The second stage occurred after the aforementioned historical event. These two stages vividly illustrate the journey of the Gaokao examination from its early establishment in 1949, through complete suspension during the Cultural Revolution (1966-1977), to its revival post 1977 and at present, the Gaokao examination plays the role as the key pillar of higher education system in China.

The English test in this study refers to the college entrance examination English test. Since the Mainland China implements the mixed proposition system of the Ministry of Education and provincial education departments, the examination papers are divided into national papers and provincial independent papers. The national



examination paper is the college entrance examination paper organized and made by the Examination Center of the Ministry of Education, which applies to most provinces and regions in the country. The provincial independent test paper mainly involves five (5) provincial units, Beijing, Tianjin, Shanghai, Zhejiang, and Hainan. At present, there are four (4) types of college entrance examination papers, namely, the national first volume, the national second volume, the national third volume, and the independent proposition paper. Among them, an autonomous proposition can be divided into a completely autonomous proposition and a partially autonomous proposition.



1.2.3 A brief introduction to English Curriculum Standards in Mainland China

The English Curriculum Standard in Mainland China has undergone four (4) major phases of development since its initial publication in 2001, reflecting the evolving aims of China's education system and specifically the language teaching and learning in China.

The first phase of development is **the Senior English Curriculum Standards (Experiment) which was the 2003 Edition**. It represented the government of China's early attempt to standardize English instruction at the senior secondary level in all provinces. This edition of English Curriculum Standards was confined to high





school content, and it provided the standards needed by teachers and educators for a more coherent and unified approach to classroom instructions. The overarching goal behind the English Curriculum Standards was to establish a consistent framework that would guide classroom practices across different provinces, ensuring that all students received an equal quality of English education. Central to these standards was the emphasis on reinforcing students' foundational (basic) language skills in reading, writing, listening, and speaking. These competencies are essential to elevate the English proficiency of the students which was in line with China's long-term educational agenda (Ministry of Education of the People's Republic of China [MOE], 2021).



The second phase of development involved the revised edition of **the English Curriculum Standards for Full-time Compulsory Education**. Compared to the 2003 edition, the revised edition of the English Curriculum Standards represented a major expansion of language instruction to both primary and junior secondary levels. Students were exposed to a systematic English instruction from an early age through secondary schooling, and this iteration promoted a cohesive set of objectives and learning experiences aimed at strengthening foundational skills and fostering higher-order language proficiency amongst students. Such an integrated approach reflects a shift toward a more comprehensive progression of English instructions (Ministry of Education of the People's Republic of China [MOE], 2021).





The third phase involved **the English Curriculum Standard for General High School (2017 Edition)**, which was divided into seven (7) main parts with each serving a distinct purpose. The first section outlined the fundamental nature of the curriculum while defining the basic concepts that guided English instruction at this level. This also established the theoretical framework underpinnings the language teaching and learning in China. The second section identified a roadmap for skills development and learners' expected learning outcomes. The third section laid out the structure of the curriculum. It also clearly explained how specific modules, topics, and courses should be integrated. The fourth section focused on the content that teachers and educators should teach their students. This section set guidelines for the sequence and depth of language skills mastery i.e. reading, writing, listening, and speaking skills. Building on this, the fifth section stipulated standards for academic quality and described benchmarks that students should meet to demonstrate meaningful language learning progress. The second last segment offered suggestions for effective implementation which enabled teachers and educators to adapt the curriculum's principles to local contexts, teaching methodologies, and student needs. The final section clarified how schools could tailor language instructions and how students could navigate their personal learning experiences to chart their own academic pathways. This English Curriculum Standard edition was introduced to align high school English education across the provinces and ultimately preparing students for success in an increasingly demanding world (Ministry of Education of the People's Republic of China [MOE], 2017).



The final developmental phase is **the English Curriculum Standards for General High School** (2020) which was expanded from the 2017 framework by adjusting its academic objectives to align with rapidly changing social and global landscapes. Key improvements to this developmental phase include a broader focus on learner autonomy. Learners are encouraged to undertake an active role in shaping their respective learning paths. In addition, the **English Curriculum Standards** place renewed emphasis on integrated language skills, challenging traditional practices that tend to isolate listening, speaking, reading, and writing skill in classroom settings. This edition advocates a more real-life application of English instructions in both academic and everyday contexts, and it also aims to nurture a new generation of competent, self-directed language learners who are able to adapt to various communicative demands (Ministry of Education of the People's Republic of China [MOE], 2020) (refer to Table 1.1).

Table 1.1

Requirements for the Content of the Theme Context of English Curriculum in High School

Theme context	Theme community	Content requirements
People and self	Life and Learning To be and to do	Personal, family, community, and school life Healthy lifestyle and positive attitude towards life Know yourself, enrich yourself and improve yourself Willingness to learn, good at learning, lifelong learning Rules and methods of language learning



Theme context	Theme community	Content requirements
People society	and Social services and interpersonal communication Literature, Art, and Sports History, Society, and Culture Science and Technology;	Excellent conduct, correct attitude towards life, civic duty, and social responsibility The meaning and value of life Future career development trends, personal career tendencies, future planning, Opinions on innovation and entrepreneurship; 1. Good interpersonal relationship and social interaction 2. Public welfare undertakings and voluntary services 3. Cross-cultural communication, tolerance, and cooperation 4. Novels, dramas, poems, biographies, brief history of literature, classic speeches, literary masterpieces, 5. Overview and development of painting, architecture, and other fields 6. Overview and development of film, television, music, and other fields 7. Sports activities, large-scale sports events, sports and health, and sports spirit 8. Cultural customs and traditional festivals of different nationalities 9. People who have made outstanding contributions to society 10. Important international organizations and public welfare institutions 11. Common sense of law and awareness of 12. Material and intangible cultural heritage 13. Social hot issues 14. Major political, historical events and cultural origins 15. Social progress and human civilization 16. Scientific and technological development and information technology innovation, scientific spirit, and





Theme context	Theme community	Content requirements
Man and nature	Natural ecology Environmental protection Disaster prevention To explore the universe	information security; 1. Geographical profiles of 2. Protection of the natural environment and natural heritage 3. Man and environment, man and plants and animals 4. Natural disasters and prevention, safety knowledge, and self-protection 5. The relationship between human survival, social development, and environment 6. Research achievements in natural sciences 7. Exploration of the mysteries of the Earth and the universe;

The English Curriculum Standards for General High School (2020) divides courses into three (3) categories: compulsory, optional compulsory, and elective. Compulsory courses are established to promote the comprehensive development of students, with all studied textbooks subject to formal assessment. Optional compulsory courses, on the other hand, cater to individual learners' interests and entrance examination requirements in which teaching materials are selected specifically for both study and evaluation. Meanwhile, elective courses are designed and offered by high schools based on their unique academic interest, allowing students the flexibility to learn independently without being bound by examinations.

1.2.4 The high school English textbooks in this study

The high school English textbooks were prepared according to the High School





English Curriculum Standards issued by the Ministry of Education. In line with the English Curriculum Standards for General High School (2020), English textbooks in high schools are also divided into compulsory courses, optional compulsory courses, and elective courses. High School English Textbooks published by the People's Education Press (henceforth, PEP) are widely used in the Mainland China and they are the top popular textbooks used by Chinese students (Huang and Tian, 2022). There are seven (7) English textbooks with three (3) compulsory and four (4) optional compulsory, each of which has five (5) themes. Tables 1.2 and 1.3 show the themes in Compulsory and Optional Compulsory Textbooks.

Table 1.2*Themes in Compulsory Textbooks*

Themes	Volume I	Volume II	Volume III
Theme 1	Teenage life	Cultural heritage	Festivals
Theme 2	Traveling around	Wildlife protection	Morals and virtues
Theme 3	Sports and Fitness	The Internet	Diverse cultures
Theme 4	Natural Disasters	History and traditions	Space exploration
Theme 5	Languages around the world	Music	The value of Money

Table 1.3*Themes in Optional Compulsory Textbooks*

Themes	Volume I	Volume II	Volume III	Volume IV
Theme 1	People of achievement	Science and scientists	Art	Science fiction
Theme 2	Looking into the future	Bridging cultures	Healthy lifestyle	Iconic attractions
Theme 3	Fascinating	Food and Culture	Environmental	Sea





Themes	Volume I	Volume II	Volume III	Volume IV
	parks		protection	exploration
Theme 4	Body language	Journey across a vast land	Adversity and courage	Sharing
Theme 5	Working the land	First aid	Poems	Launching your career

Each of the seven (7) English textbooks – three (3) compulsory and four (4) optional compulsory – organizes its material around five (5) core themes that serve as fundamental contexts for language teaching and learning in the classrooms. These themes involve the following broad areas – personal growth and life skills; society and culture; science, technology and innovation; environment and sustainability and global perspective and communication. Collectively, these five (5) themes will encourage a holistic language learning experience amongst learners by integrating real-life knowledge and practical skills. These skills not only assist learners strengthen their language abilities (reading, writing, listening, and speaking skills) but also broaden their understanding of personal, social, and global issues. Indirectly, these themes will prepare learners for deeper engagement in both academic and everyday settings.

1.2.5 Reading Skill Requirements

Reading skills occupy a core position in the Gaokao examination, high school English textbooks, and the English Curriculum Standards for General High School



(2020). This is reflected in the score distribution (refer to Table 1.4).

The Gaokao English exam typically assesses four (4) core language skills: listening, reading, writing, and language knowledge/application. A new trend is that more provinces are incorporating speaking sections into the exam. The score distribution is summarized below (Table 1.4):

Table 1.4

The Components of Gaokao Examination and Weighting

Component	Weighting
Listening Skills	20–30%
Reading Skills	40–50%
Language Knowledge/Usage	10–15%
Writing Skills	15–20%
Speaking Skills	Experimental in Some Provinces

Hence, reading comprehension often account for the largest portions of the English exam's total score; performance in these skills can heavily influence an examinee's overall result in Gaokao examination (Ministry of Education of the People's Republic of China [MOE], 2017).

In the English Curriculum Standards for General High School (2020), the reading skills is the key component that underpins language mastery. The main purpose of the curriculum standards is to develop the ability to handle various text types amongst the learners and lessons should be aligned with real-life

communicative needs and academic interests of the learners. The curriculum standards serve as guidelines to teachers and educators to ensure that critical reading skills are fostered amongst the learners. The national curriculum highlights the commitment of the government of China in shaping the learners into competent readers and equipped with analytical thinking skills.

Textbooks approved by the MoE of China also place significant emphasis on reading passages that are weaved with authentic materials to mirror real-world contexts. The text types used in these textbooks span multiple genres – such as narratives, descriptive and informational essays. This offers a multifaceted approach to develop the reading proficiency amongst the learners. Subjects from textbooks help students develop a variety of abilities, such as understanding narratives, evaluating arguments, or extracting facts for problem-solving. In other words, the reading components across textbooks will equip Chinese learners with the adaptability and critical thinking skills needed to survive both their academic settings and everyday life.

The current study explores the consistency of reading skills and comprehension among the English Curriculum Standards for General High School (2020), high school English textbooks, and the Gaokao examination. Specifically, the current study explores the consistency in terms of reading skills and comprehension ability among three sources that are the English Curriculum Standards for General High

School (2020), high school English textbooks, and the Gaokao examination.

Such coherence is crucial to reduce any mismatch between classroom instruction and test requirements, promoting a more unified approach to teaching and learning of English in China. By aligning instructional materials, curriculum standards, and exam requirements, the academic growth and the overall performance of the students will be improved.

The next section highlights the statement of problem.

1.3 The Statement of Problem

As aforementioned, the aim of the study is to ascertain how and to what extent the high school English curriculum standards, textbooks, and college entrance examination materials are aligned. The Statement of Problem section will highlight the importance of English in China, the importance of the Gaokao examination, the lack of research on curriculum alignment in China and the importance of the curriculum alignment.



1.3.1 The Importance of English Language Education in China

Inevitably, English serves as the main medium of communication in international business, science, technology and other spheres (Grin & Arcand, 2013). In China, the significance of fostering English language proficiency amongst the students and the general population of China is unarguably inseparable from the country's socio-economic ambitions and the current demand for more international engagements (Zhao, 2020). In fact, in China the mastery in English is requisite for academic success and for Chinese professionals; it is the foundational skills used to impress when engaging with foreign counterparts such as foreign corporations, institutions, and market (Bolton & Botha, 2015).



Despite its importance, the proportion of proficient English speakers amongst China students are relatively small (Xiang and Wang, 2024). Ploscaru (2024) explains that about 200-400 million out of the total population of China are learning English, however the number of those who are able to fluently speak the language is relatively small. In other words, less than 10 million Chinese or an estimation of 1% of the total China population can speak English (Ploscaru, 2024). In fact, in 2023 English Proficiency Index, China was ranked 82nd place out of 113 countries, indicating that the English language proficiency in China is of low level (British Council, 2023)





Due to these challenges, the government of China had reformed its educational policies and English language was reinforced as the primary foreign language in schools. In addition, English was established as one of three (3) core components in the Gaokao examination. Dockx et al. (2020) emphasize that even though English was introduced in Chinese schools, classroom instructions often lack consistency. There are noticeable gaps between classroom content and the competencies tested in high-stakes exams such as the Gaokao (Dockx et al., 2020). It is, therefore, imperative to align between the curriculum standards, textbooks, and Gaokao examination cohesively. This alignment can bridge the gaps between classroom instructions, approved high school textbooks and the language competencies evaluated in this examination which thereby will improve both academic success and English language proficiency of the students.

1.3.2 The importance of the Gaokao Examination in China

The predecessor of Gaokao was the imperial examination system. The imperial examination system was founded in the Sui Dynasty (605 AD), perfected in the Tang Dynasty, developed in the Song Dynasty, fell in the Yuan Dynasty, flourished in the Ming Dynasty, and ended in the Qing Dynasty (abolished in 1905) after 1,300 years. The emergence of the imperial examination system was a huge progress in history. It adhered to the principles of "*free registration, open examinations, equal competition,*





and merit-based appointment", which gave people at the lower levels of society the opportunity to participate in the operation of society and change their own destiny. The imperial examination system broke the monopoly of cultural capital by the privileged strata of society. It is said that "if a good scholar is a common man, his son will become a public minister; if he is not a good scholar, the son of a public minister will be a common man." The implementation of the imperial examination system ensured the orderly flow of social talents to the greatest extent and provided institutional equality of opportunities for human development (Elman, 2013).

It has been 120 years since the Chinese central government abolished the imperial examination system in 1905 (late Qing Dynasty) and established a Western school examination system. In other words, Gaokao is the modern version of the imperial examination system in China, because it undertakes the function of selecting talents for the country. It is also a mean to strength social management, and it affects the future career of ordinary Chinese students - achieving social class elevations, such as students entering prestigious universities study, and have decent jobs and higher salary upon graduation.

Table 1.5 shows a comparison between the population of China and the number of test takers of Gaokao examination.



Table 1.5

Comparison between China population and Gaokao Test Takers

Year	Population of China	Gaokao Test Takers (Approx.)
2020	1411.78 million	10.71 million
2021	1412.60 million	10.78 million
2022	1411.75 million	11.93 million
2023	1410.50 million (est.)	12.91 million

Based on the statistics above, a simple calculation can be done to determine the ratio between the population of China and the number of students who took the Gaokao in the same year. Let's take the year 2022 as an example.

- The estimation of China's population in 2022 - 1.41175 billion
- The number of Gaokao test takers in 2022 - 11.93 million

This indicates that there is one (1) Gaokao test taker for every 109 people in the total population of China. Thus, in 2022 it could be estimated that the ratio was 1:109 or about 1% of the total population of China who had sat for the Gaokao examination. The sheer scale of annual Gaokao test taker indicates that the Gaokao examination is the largest standardised test in China and this serves as one of the motivations of undertaking the current research.

The Gaokao examination (college entrance examination) results determine the level of university that the examinees will attend and their future work income level.



This examination system is involved to a considerable extent in the allocation of social resources, hence this system is vital in China.

1.3.3 Lack of Publications on Curriculum Alignment

Research on the curriculum alignment found that the main focus on these researches were the inconsistency between; first, testing content and teaching content; second, testing vocabulary phrases and teaching vocabulary phrases; third, testing text genre and teaching text genre; and finally, testing vocabulary frequency and teaching vocabulary frequency. In these studies, it was determined that these problems are similar in various countries (Gedik and Kolsal, 2022; Tai and Chen, 2015; Wang, 2018). It should be emphasized that none of these studies focus on the alignment between curriculum standards, teaching content and testing content, particularly the reading skills.

There is evidence of incomparability between English textbooks and English examinations (Gedik and Kolsal, 2022; Tai and Chen, 2015; Wang, 2018). In the process of practical teaching, people (including parents and teachers) have found that teaching and examination content can potentially be out of line. This phenomenon is also common in many parts of the world, such as Turkey, Iraq, Sweden and other places (Gedik and Kolsal, 2022; Tai and Chen, 2015; Wang, 2018). Scholars who





have conducted research on the connection between curriculum standards, English textbooks and English exams are still scarce. However, the following research focused on the alignment of lexical phrases between curriculum standards and test content and structure.

The results indicated that inconsistencies between vocabulary and phrases in the college entrance examination and textbooks often occur. For example, Wang (2018) studied the vocabulary of forty (40) college entrance examination reading comprehension texts in Jiangsu Province from 2008 to 2017. It was found that the propositional personnel need to further optimize the structure of the test paper so that the layout of the reading text is more reasonable, and the genres and themes are more comprehensive. In addition, Wang (2018) reiterated that high school English teachers should improve the effectiveness of teaching of reading in terms of genre, vocabulary width, vocabulary block, and function words. Another example is the research by Tai and Chen (2015). Tai and Chen (2015) conducted a comparative corpus-based analysis of the English test items in high School Entrance Exams and curriculum-based English textbooks to provide more valid information for syllabus design and materials development.

Meanwhile, Gedik and Kolsal (2022) highlighted issues on gaps between curriculum standards, textbooks and tests are major concerns and more research is needed. They emphasized that Turkish high school students who were taught English





with official textbooks were only able to tackle low-frequency and unable to tackle more sophisticated words at a higher level of syntactic complexity when they take the nationwide exam (Gedik and Kolsal, 2022).

There are some studies that highlighted the inconsistencies between the compilation of textbooks and the testing of exams, which affect the enthusiasm of students, the teaching effect of teachers, and the purpose of the policies issued by the education department. In the theoretical framework of literary genre, Wang et al. (2022) analyzed the literary genre features of the reading comprehension part of the English Test Paper 2021 (New High-Test Paper I) from the macro and micro dimensions, at the stylistic, lexical and syntactic levels. The results show that the lexical complexity of expository texts is significantly higher than that of pragmatic and narrative texts, and expository texts contain more complex sentences and have higher syntactic complexity than practical and narrative texts. Yua and Renandyab (2021) posited that although the study indicates that some words are distributed favorably, the textbooks fail to provide sufficient coverage and repetition of the words required by the 2017 English Curriculum Standard. Meanwhile, Cheng and Chang (2022) wrote that the reading passages in the English tests generally have lower overall readability, lower narrativity, and higher syntactic complexity than those in the textbooks. Therefore, it is very important to solve the problem of consistency between teaching and examination.



It should be highlighted that studies that look into the alignment between the curriculum standards, teaching content and testing content are still limited. This serves as the motivation and the necessity for conducting the current research.

Ensuring this alignment is integral so that the intended educational objectives, textbooks and examination are coherently integrated. With the alignment of curriculum standards, instructional content, and assessments, first, instructional practices; second, students learning and academic success of the students; and the education policy of China will be improved. Therefore, addressing the current research gap is imperative.

1.3.4 The importance of the curriculum alignment

The current study aims to explore the alignment of the English Curriculum Standards for General High School (2020), the high school English textbooks and the Gaokao Examination. Abate and Mishore (2024) explain that research on alignment studies serve as guidelines to teachers and educators on how to refine their instructional strategies to effectively meet the intended course objectives stipulated in the curriculum standards.



According to Biggs (2003, p.2), “What and how one teaches must lead to where the student is intended to go.”. Biggs (2003 and 2014) further explains that teachers and educators should ensure that what they teach (the curriculum standards), how they teach (instructions and textbooks), and how they assess (examinations) are cohesively toward the same intended learning outcomes. This is further reiterated by Ornstein & Hunkins (2017, p.102), quoted as saying “Curriculum alignment is about ensuring that all elements in the educational process – curriculum, instruction, and assessment – are working harmoniously toward achieving the intended learning outcomes.” Additionally, Zhao et al. (2023) emphasize that alignment studies provide invaluable insights to the MoE of China on the consistency within these major components (i.e., the English Curriculum Standards for General High School (2020)), the high school English textbooks and the Gaokao Examination). In other word, this alignment promotes consistency.

Hence, this study examines three core components: English Curriculum Standards for General High School (2020), high school English textbooks, and the Gaokao examination particularly how reading skills and comprehension. This will ensure that the curriculum standards, teaching practices and assessments are well coordinated to support the learning needs of the students in Mainland China.





1.4 Aims of the Study

The aims of this study are twofold: first, to find out whether the content of teaching and the content of the tests complement each other, that is, whether students are actually tested based on the curriculum standards and the high school textbooks. Second, to gather the findings from the alignment study which then can be used to form a comprehensive guideline of curriculum standards that ensure there is a fairness or balance in testing among the three elements (Curriculum Standard, Textbooks and Exam papers) share the same essential characteristics.



1.5 Research Objectives

The objectives of the research are as follows:

- a) to determine the similarities and differences of the lists of Keywords of the Reading text section between High School English Curriculum Standard, High School English textbooks, and the College Entrance Examination in Mainland China.
- b) to gauge the similarities and differences in lexical density of the Reading text section between High School English Curriculum Standard, high school English textbooks, and the College Entrance Examination in the Mainland



China.

- c) to explore the similarities and differences of readability used in the Reading text section between High School English Curriculum Standard, high school English textbooks and the College Entrance Examination in the Mainland China.

The overarching research objective of the current study is:

- d) to determine the alignment of Reading text between High School English Curriculum Standard, high school English textbooks, and the College Entrance Examination in the Mainland China.

1.6 Research Questions

Based on the research objectives, the current study aims to answer the following research questions:

- a) What are the similarities and differences in the lists of Keywords of the Reading text section between High School English Curriculum Standard, high school English textbooks, and the College Entrance Examination in the Mainland China?



- b) What are the similarities and differences in lexical density of the Reading text section between High School English Curriculum Standard, high school English textbooks, and the College Entrance Examination in the Mainland China?
- c) What are the similarities and differences in Readability used in the Reading text section between High School English Curriculum Standard, high school English textbooks, and the College Entrance Examination in the Mainland China?

The overarching research question of the current study is:

- d) To what extent does the Reading text section between High School English Curriculum Standard, high school English textbooks, and the College Entrance Examination aligned in the Mainland China?



1.7 Definition of Terms

1.7.1 High school

The pre-university education system currently implemented in China generally consists of six (6) years of primary school, three (3) years of junior high school, and three (3) years of high school (Ministry of Education of the People's Republic of China, 2021). The first nine (9) years of schooling are considered compulsory for all students in Mainland China.





In this study, the term High School refers to students who are in Grade Ten (10), Grade Eleven (11), and Grade Twelve (12) and more importantly, these students have completed nine (9) years of compulsory education. In other words, they are undergoing the three-year pre-university education system.

1.7.2 English Textbooks

The inclusion of the English subject into the education system of China began during the Qing Dynasty and with the establishment of the Tongwen Guan (School of Combined Learning) in 1862. This inclusion significantly marked the initial step towards incorporating Western knowledge into the Chinese education system.

Today, English textbooks must be approved by the MoE of China in making sure that consistency of instructional materials across the provinces is upheld. Published by the People's Education Press in China, its content is diverse and it incorporates digital enhancement for the purpose of optimized learning.

In the current study, English textbooks refer to high school English textbooks utilized in mainland China and published by the People's Education Press. These textbooks include three (3) volumes of compulsory courses and four (4) volumes of optional compulsory courses. However, this study only focuses on the reading





sections aligned in these textbooks.

1.7.3 Reading texts

As mentioned above, this study focuses on English textbooks approved by the Ministry of Education and published by the China People's Education Press. Furthermore, this study focuses only on the reading sections of these textbooks. In the Gaokao examination study, the Reading component consists of 40-50% of the English examination. It is considered as the biggest skill component in the English Gaokao examination.



Hence, reading texts refer to the 35 reading texts in high school English textbooks and the 25 reading materials in the college entrance examination English test papers.

1.7.4 College Entrance Examinations (Gaokao)

The examination targets full-time high school graduates and citizens of the People's Republic of China with equivalent degrees. The admission is divided into two (2) categories: 1) industry, agriculture, and medicine (including physical education), and





2) literature and history (including foreign languages and arts). The compiling institution of test contents is the Ministry of Education or the provincial education administration department. The examination subjects are Chinese, mathematics, foreign language, and general (liberal arts or science). Foreign languages can be any one of six of the following: English, Japanese, German, French, Russian, Spanish. Comprehensive is divided into liberal arts comprehensive (including Politics, History, and Geography) and science comprehensive (including Physics, Chemistry, Biology, and Technology) (Ministry of Education of the People's Republic of China, 2021).

In this study, the college entrance examination (English subject) refers to the English language test in the Nationwide Unified Examination for Admissions to General Universities and Colleges (Gaokao). It is divided into two (2) parts: language knowledge and language application according to the 2019 National Unified Examination Program (English) for Admissions to General Universities and Colleges (GaoKao). Language knowledge requires candidates to master and be able to use the basic knowledge of English pronunciation, vocabulary, and grammar, as well as the learned functional ideas and topics.

In the Gaokao English language test, students are required to master about 3,500 English words which were selected by the MoE of China. These 3,500 words covers the vast spectrum of the English language, including basic vocabulary, academic terms and functional language (Ministry of Education of the People's Republic of





China, 2021).

1.7.5 Test papers

As mentioned in Section 1.2, three (3) main objects of the current study are the high school English curriculum standards, high school English textbooks, and English test papers/examinations of the English subject of the annual college entrance examination in mainland China from the year 2018 to 2022 (5 years). The current study aims to determine how and to what extent the high school English curriculum standards, textbooks, and college entrance examination materials are aligned.



Specifically, the current study focuses only on the Reading component of the Gaokao English examination. The test paper analysis encompasses test papers from 2018 to 2022, consisting of 25 reading passages. Full text content of the reading section was examined to determine the coherence between curriculum standards, instructional materials, and assessment practices.

1.7.6 Curriculum Standards

Curriculum standards are defined as educational guidelines that stipulate the



knowledge and skills students should master at each grade level. In other words, it serves as a framework for teachers and educators to design instructional materials and assessments to ensure consistency and quality in education are upheld (Schwartz, 2023).

In the current study, Curriculum Standards refer to the English Curriculum Standards for General High School (2020) that is currently implemented in high schools across the mainland China. This English curriculum standards are published by the MoE of China and this document includes an appendix consisting of the English word glossary.

1.7.7 Reading comprehension

Reading comprehension is the ability to read a text, process it, and understand its meaning. It relies on two interconnected abilities: word reading (being able to decode the symbols on the page) and language comprehension (being able to understand the meaning of the words and sentences) (Oakhill, et al, 2014).

When comprehending a text, students do not merely recall the exact words and phrases they have read. Instead, a mental model was constructed and this is done by integrating the meaning of words and sentences in the text into a coherent



whole (Oakhill et al., 2014). However, comprehension is influenced by the main factors, including vocabulary knowledge, lexical complexity, and overall text readability.

Ludewig et al. (2023) explain that vocabulary knowledge is fundamental to reading comprehension and it refers to the extent of words a student understands and is able use. Possessing an extensive vocabulary knowledge will enable the student to effectively grasp the meaning of the text. Arya et al. (2011) emphasize that lexical complexity refers to the variety of words utilized in a reading text. This includes several factors such as the length of a word, its frequency and more importantly, specialized terminology. It should be noted that increased lexical complexity can reduce reading comprehension. Shanahan (2013) clarifies that text readability refers to how easy it is for a student to comprehend a text, and this is influenced by factors such as sentence length, syntactic structure, and lexical complexity.

Hence, in the current study, reading comprehension specifically refers to the articles and paragraphs in the reading comprehension section of high school English textbooks and English test papers, focusing on three aspects: Keywords, Lexical Density, and Readability.





1.7.8 Keywords

Notomo (2023) rightly states that the notion of keyword has long defied a precise definition, and quotes Boyce et al. (1994) for the general definition “a surrogate that represents the topic or content of a document”, which makes sense in the context they referred to (libraries and information science). Notomo distinguishes four senses or roles: 1) terminology: specialized lexical items from a particular domain; 2) topics: terms and labels that are part of a systematic concept system, such as Wikipedia category names; 3) index terms: terms indicating major concepts, events, or people, including named entities, and 4) summary terms: words or phrases meant to serve as a quick description of the content.



Of these four roles or senses of the term, it is probably the last one that is most often thought of. In this sense, keywords can be loosely defined as words that somehow encapsulate the topics discussed in a document or collection of documents; in other words, what those documents “are about”. This is what most authors in corpus linguistics studies agree on: keywords are meant to capture the aboutness of a document or set of documents, and make up their ontology (Scott and Tribble 2006; Mahlberg 2007; Bondi 2010; Marchi 2018).

Scott (1997, p. 236) stated that a key word may be defined as a word which occurs with unusual frequency in a given text. This does not mean high frequency





but unusual frequency, by comparison with a reference corpus of some kind.

Keywords have also been referred to as “pointers” that “merit chasing up and tracking down” (Scott 2010, 55–56). Similarly, Baker (2006, p.137) states that keywords “act as signposts to the underlying discourses”, while Hunt & Harvey (2015, 139) point out that keywords “serve as indicators of expression and style as well as content to provide a sense of the ‘aboutness’ of a language variety”. Bondi (2010, p.1) qualifies keywords as “markers of the aboutness and the style of a text”. Stubbs (2010, p.25) compares keywords to the “tips of icebergs: pointers to complex lexical objects which represent the shared beliefs and values of a culture”.



In summary, keywords in corpus linguistics are regarded as words and phrases that act as pointers, markers, indicators, or signposts to the contents, style and discourse of a corpus.

Keywords are calculated by first making a list of words for the corpus and a list of words for the "reference" corpus, then comparing the frequency of each word in the two lists. A reference corpus is any corpus chosen as the standard against which the target corpus is compared. The reference corpus usually has to be fairly large and of the right type for keywords work. If the word occurs, for example, 5% of the time in the small wordlist and 6% of the time in the reference corpus, then it doesn't appear to be "critical", but if the scores are 25% and 6%, then the first one will be very





"critical"(https://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/clc_top.htm).

In this study, it refers to the comparison of two wordlists, which come from English curriculum standards, English textbooks, English test papers and one reference corpus.

1.7.9 Lexical Density

Lexical density is the proportion of content words to function words in a text. The higher the proportion of content words, the greater the lexical density. Lexical density, lexical diversity or lexical richness (Daller, van Hout and Treffers-Daller, 2003) are terms which refer to statistical measures that gauge the lexical richness of texts and may also be used to assess students' overall progress. The lexical richness of a text accounts for how many different words are used in a text, while lexical density provides a measure of the proportion of lexical items (i.e., nouns, verbs, adjectives and some adverbs) in the text (Johansson, 2008). Both measures have been applied in computer analyses of corpus data. As a rule, texts with a lower density are more easily understood and spoken texts have lower lexical density levels than written texts (Ure, 1971; Halliday, 1985). However, as argued by Johansson (2008), a text may have high lexical diversity (i.e., contains many different word types), but low lexical density (i.e., contain many pronouns and





auxiliaries rather than nouns and lexical verbs) or vice versa.

In this study, coefficient is used as a comparison standard to measure the high school English curriculum standard and the college entrance examination English test papers. It can compare numerically how easy it is to distinguish between different texts.

1.7.10 Readability

Understanding what makes a text difficult to read is an important element of literacy instruction, materials development, and research into text readability. The Common Core State Standards (Common Core State Standards Initiative, 2010), often used in the United States, detail three factors that influence text readability: reader and task factors, qualitative text complexity factors, and quantitative text complexity factors.

Knowing the effect of language features on text readability, researchers have developed hundreds of readability formulas over the past 80 years (Crossley, Skalicky, et al., 2017). These readability formulas quantify linguistic features in the text that act as proxies for language features known to influence the reading process. These features are then used in statistical models to predict text readability. For instance, Flesch Reading Ease (Flesch, 1948) is an early readability formula that





counted the number of syllables in words as a proxy for word complexity (i.e., lexical sophistication) and the number of words per sentence as a proxy for sentence complexity (i.e., syntactic complexity). Readability formulas have played an important role in a variety of learning environments by allowing teachers, administrators, and materials developers to better match readers with texts that complement their reading level, which is a critical component of learning to read (Allington, 2005; Stanovich, 1985; Wolfe et al., 1998).

Readability refers to the ease of reading or understanding certain reading materials (Dale & Chall, 1948; Klare, 1963; Mc Laughlin, 1969). The study of readability focusing on texts' expressive style indicates how vocabulary and sentence structure affect readers' understanding of texts (Dale & Chall, 1948).

Readability metrics (readability formulas) currently tend to use simple measures such as word length, sentence length, word frequency. Common formulas are: Flesch–Kincaid readability tests; Dale–Chall readability formula; Gunning fog index; Fry readability formula; McLaughlin's SMOG formula; and The FORCAST formula. The readability of a text is mainly affected by the complexity of its words and grammar. Readability, hence, refers to the ease with which a written text is understood by readers, and various researchers have used various factors to measure readability.



In this study it refers to the readability of reading texts in English textbooks and reading materials in college entrance examinations (2018-2022).

1.7.11 Corpus

There are many ways to define a corpus (e.g., Francis 1992: 17; Atkins, Clear and Ostler 1992: 1), but there is an increasing consensus that a corpus is a collection of (1) *machine-readable* (2) *authentic* texts (including transcripts of spoken data) which is (3) *sampled* to be (4) *representative* of a particular language or language variety.

While all scholars agree upon the first two qualities, there are differing opinions regarding what can be counted as representative.

There are many different expressions about the definition of a corpus, but its essence content is approximately the same. Corpus comes from Latin and originally means body. It refers to a collection of texts stored in an electronic database. It is an electronic text set that is collected according to certain sampling standards and can represent a language or a variant or genre of a language.

In this study, it refers to the reference corpus (BNC), as well as the corpus compiled by the researchers themselves based on the materials (including Curriculum Standard corpus, Textbook corpus, and Test paper corpus).



1.7.12 Corpus-based

There are currently two popular views on this issue. The first viewpoint is a comparison between corpus-based and intuition-based approaches. Broadly speaking, compared with the more traditional intuition-based approach, which rejected or largely ignored corpus data, the corpus-based approach can offer the linguist improved reliability because it does not go to the extreme of rejecting intuition while attaching importance to empirical data. The key to using corpus data is to find the balance between the use of corpus data and the use of one's intuition.



In the second viewpoint, it is the comparison between corpus-based and corpus-driven approaches. There are four basic differences between the corpus-based vs. corpus-driven approaches: types of corpora used, attitudes towards existing theories and intuitions, and focuses of research.

- (1) Regarding the type of corpus data used, there are three issues – representativeness, corpus size and annotation.
- (2) attitudes towards existing theories and intuitions.
- (3) The third important difference between the corpus-driven and corpus-based approaches is their different research focuses.
- (4) the corpus-based approach is not as radical as the corpus-driven approach.



The corpus-based approach is a method that uses an underlying corpus as an inventory of language data. From this repository, appropriate material is extracted to support intuitive knowledge, verify expectations, allow linguistic phenomena to be quantified, find proof for existing theories or retrieve illustrative samples. It is a method where the corpus is interrogated, and data is used to confirm linguistic pre-set explanations and assumptions. It acts, therefore, as additional supporting material (Storjohann, 2005).

In this research, adopting the above perspective, appropriate material is extracted from this repository to support intuitive knowledge, verify expectations, quantify linguistic phenomena, find evidence for existing theories, or retrieve illustrative samples.

1.7.13 Corpus Linguistics

We have, so far, assumed that corpus linguistics is a methodology rather than an independent branch of linguistics. This view, however, is not shared by all scholars. For example, it has been argued that corpus linguistics ‘goes well beyond this methodological role’ and has become an independent ‘discipline’ (Tognini-Bonelli 2001: 1).



One view is that corpus linguistics is an independent discipline. It has its theoretical system and operating methods. As corpus linguistics is based on a large number of real language data, the conclusions drawn from systematic and exhaustive observation and generalization of the corpus have unparalleled innovative significance for the construction of language theory.

In the view of other researchers, corpus linguistics is a research method. This method is based on a large number of real languages and can be used to answer questions that are difficult to answer by other means, thus greatly enriching existing research methods. Corpus linguistics takes a large number of carefully collected authentic texts as research materials and concludes mainly through the method of probability statistics. Therefore, corpus linguistics is empirical in nature (Liang, Li and Xu, 2010).

In this study, corpus linguistics is considered as both a method and a discipline.

1.7.14 Alignment

Science education reform emphasizes the importance of alignment (Jin et al., 2019). After U.S. national normative examinations failed to assess students' knowledge and skills, Bloom focused on aligning academic assessment tasks with behavioral goals





in the 1960s. Alignment was originally used to investigate the assessment and instructional goals. Later, it was included in educational reforms by the American Association for the Advancement of Science (1993) and the NGSS Lead States (2013). Webb (1999) defines alignment as the degree of conceptual, procedural, and methodological congruence among the elements of a curriculum system. Webb explains that alignment analysis can be conducted on all curriculum system elements. State Standards and State Assessment Systems: A Guide to Alignment (Columbia, USA) defines alignment as two or more items aligning. Alignment also means combining pieces into a suitable whole (Marca et al., 2000). All of these statements clarify that two or more elements can be involved in congruent matching.



Achieving alignment in the classroom is a critical aspect of standards-based curriculum reform (Yang, 2020). The purpose of the curriculum is to enhance learning outcomes (Wiggins & McTighe, 2005), and the alignment of teachers' teaching objectives with the instructional activities students engage in, assists students in achieving their learning objectives (Cui & Lei, 2015; Duis et al., 2013). The alignment of teaching objectives with classroom assessments helps reveal important aspects of student's learning process, understand students' learning, and promote the achievement of teaching objectives (Cui & Lei, 2015; Duncan & Hmelo-Silver, 2009). Aligning instructional activities with classroom assessments exposes teachers' instruction and helps them give students feedback that helps students meet instructional goals (Cui & Lei, 2015; Hall, 2002). When teachers'





instructional goals match student activities and classroom assessments, teaching and learning are maximized and student learning is promoted (Biggs, 1996; English & Larson, 1996; Krajcik et al., 2007; Liu, 2009). In addition to guiding instructional implementation, curriculum standards also define the scope of academic-level examinations (PRC, 2022). Thus, aligning curriculum standards, instructional objectives, classroom teaching activities, and classroom assessments helps implement curriculum standards, improve teaching, and improve student achievement (Krajcik et al., 2007; Porter, 1997).

Webb (1997) believed that alignment refers to the extent to which standards and evaluation are agreed and cooperate with each other to guide students to learn what they expect to know and do. This definition is considered to be aligned in a broad sense. Bhola, Impara and Buckendahl (2005) defined alignment as the alignment between content standards and assessment in a specific subject area of a state. The assessment here is to measure students' achievements in content standards. Therefore, alignment is considered as the link between standards and assessment. La Marca et al., (2000) stressed that students should show their knowledge and skills in the assessment to meet the preset of the curriculum standards, and then make appropriate explanations according to their performance. This is to say that the assessment should reflect the relevant information of the students, that is, whether the students' performance meets the expectations of the curriculum standards.





There are also scholars in China who have conducted such research, such as Zhou (2018), Tian (2018), and Peng (2018) of Chongqing Normal University, who have studied the specific performance of the alignment of classroom teaching, teaching materials and tests of high school English reading, listening and speaking with English curriculum standards, respectively, referring to the provisions of high school English curriculum standards on "reading", "listening" and "speaking". This kind of research is also called content validity research by many scholars.

In this study, "alignment" in this paper refers to the matching level and degree of alignment between the college entrance test English reading comprehension questions and the curriculum standards in the content theme dimension and cognitive level dimension. This study belongs to the second type described above.

1.8 Limitations of Study

Objectively, it mainly refers to the restrictive conditions of research that cannot be freely selected by researchers, mainly due to the restrictions of educational system and policy. In addition, during the global COVID-19 pandemic, especially in Mainland China, population mobility is still restricted. In order to reduce the cross-infection rate of social workers and reduce the expenditure of research funds, this study uses secondary data analysis. Both the reading part of the English test





papers and the reading part of the textbooks are smaller than the actual scope.

In terms of college entrance examination English test questions, the test questions include four parts: listening (30 points), reading comprehension (40 points), language knowledge application (including cloze test and grammar test, 45 points in total), and writing (including 10 points for correcting passages and 25 points for written expression, 35 points in total). It should be emphasized that the reading comprehension part in the test consists of three parts, with a total score of 95 points: reading comprehension (40 points), language knowledge application (45 points) and short passage correction (10 points in the writing part of the test paper).

In addition, listening consists of 30 points and writing 25 points. However, in order to facilitate statistical comparative research and consider the teaching habits of English education in China, this study did not include the correction of short passages in the scope of the study. The reading and writing parts of the textbooks are presented in front of teachers and students in the classroom in the form of reading texts in the actual preparation and teaching process, as students' writing answers are not at the same level. More importantly, the reality is that it is difficult to obtain the actual writing answers from students who take the English test. Therefore, this study does not list English writing model texts as actual reading materials, although this is not consistent with the actual teaching process.





At the same time, studies on geographic scope constraint costs do not tell the whole story. First of all, this study only selects the English education stage of high school. English education begins in elementary school and goes through middle school (middle school and high school) to university (including associate degree, bachelor's degree, master's degree, and Doctoral degree). According to the Notice of the Ministry of Education on Printing and Distributing the Pilot Program of Compulsory Education Curriculum, in mainland China, English education is provided from the third grade of primary school to university. At the same time, the regulation makes detailed rules on school hours.



Secondly, there are many versions of English education courses in mainland China after the publication policy is revised. Only the edition of People's Education Press is selected here because this edition is more widely used and has been used for the longest time. Furthermore, the reading passages selected were from the commonly used set of textbooks in the largest province. Other versions of the textbooks appeared late in the policy, and their influence is relatively small, and the accompanying high-test questions are limited to a single province or municipality, which is not representative.

Third, this study only focuses on the reading part of high school English textbooks. At the same time, this study involves the reading part of the English college entrance examination paper. Other parts of the book such as listening,





after-school questions, and exercises were not covered in this study. A similar scope of the study was used in the examination paper section. The reason for this is the lack of an English communication environment in mainland China. In fact, in many economically underdeveloped areas, it is more suitable for students to focus their teaching time (62 minutes per week) on an intensive reading of texts. In this way, the reading part plays the multiple functions of listening to the text, writing model essays, and oral practice.

Fourth, there are many kinds of English test papers for the college entrance examination in mainland China. This is in line with the current Gaokao policy. This study only involves the English-related papers of the college entrance examination that matches the high school English textbooks of the People's Education Press and does not involve other papers. Fifth, compared with CET-4 and CET-6 (an English proficiency test system developed in mainland China), the English College Entrance Examination is more of a selective (competitive) test than a test of language proficiency. This may have a lot to do with the uneven distribution of educational resources in mainland China, here refers to the national key universities and famous universities. These schools have received key support from the state in terms of teacher resources and scientific research funds. Unfortunately, given the focus of the study, this part of the discussion was decisively abandoned in this study to prevent the unconstrained expansion of the study.





In addition, this study does not consider the current situation of English education in Hong Kong, Macao, and Taiwan. According to the relevant laws of the People's Republic of China, these three regions are integral parts of the country. However, this study only covers areas under the jurisdiction of the Ministry of Education of mainland China. Finally, in terms of research technology route, in terms of corpus analysis, this study only focuses on keywords, readability, and lexical identity, but there are many aspects not covered, such as N-grams and semantic prosody.

1.9 Significance of Study



There are four important aspects to this research. First, education policymakers can better compile and adjust the latest education policies and continuously improve the guidance of policies. They can combine the latest foreign linguistic research results according to the continuous progress of Chinese students' learning skills, improve the actual effect of education policy implementation, reduce policy misjudgment and misuse, and actively learn from the advanced policies and excellent results of foreign English teaching. Secondly, editors can modify the original textbook according to the textbook compilation guide and constantly improve the content of the textbook according to the actual needs of national education development.





Thirdly, teachers can highlight the teaching focus in a short time and improve the students' learning effect immediately. English teachers in school and off-campus language training departments can use the latest corpus research methods combined with the latest computer technology to improve the pertinence of teaching and training, especially for language items such as synonyms, antonyms, key high-frequency words, and common phrases. Fourth, students can use the corpus to pay attention to the key high-frequency words and common phrases in their major, to do a purposeful review when reviewing and preparing for the exam. Utilizing computers and corpus, researchers can make up for the lack of an English language environment in China, focusing on professional listening and speaking training.



1.10 Summary

The first chapter introduces the brief history of English education in China, background and problem statement of this study; Research objectives, research questions, research significance, definition of terms and organization of this research.

The next chapter introduces four sections, alignment theoretical research, the corpus research, the theoretical framework of this study, and the review of previous studies and research gaps.

