

**LINKING COLLEGE ENGLISH READING
COMPREHENSION TESTS BAND FOUR TO THE
COMMON EUROPEAN FRAMEWORK OF REFERENCE
FOR LANGUAGES (CEFR)**



**A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Master of Arts in English Language Studies**

Suranaree University of Technology

Academic Year 2019

การเทียบเคียงข้อสอบการอ่านเพื่อความเข้าใจของการสอบภาษาอังกฤษสำหรับ
นักศึกษาที่ไม่ได้เรียนวิชาเอกภาษาอังกฤษระดับ 4 กับกรอบมาตรฐาน
ความสามารถทางภาษาของประเทศในกลุ่มสหภาพยุโรป



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต
สาขาวิชาภาษาอังกฤษศึกษา
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2562

**LINKING COLLEGE ENGLISH READING COMPREHENSION
TESTS BAND FOUR TO THE COMMON EUROPEAN FRAMEWORK
OF REFERENCE FOR LANGUAGES (CEFR)**

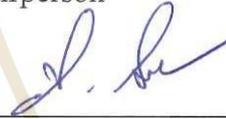
Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for a Master's Degree.

Thesis Examining Committee



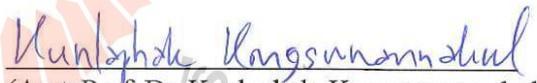
(Asst. Prof. Dr. Adcharawan Buripakdi)

Chairperson



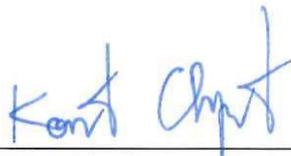
(Dr. Sirinthorn Seepho)

Member (Thesis Advisor)



(Asst. Prof. Dr. Kunlaphak Kongsuwannakul)

Member



(Assoc. Prof. Ft. Lt. Dr. Kontorn Chamniprasart)

Vice Rector for Academic Affairs
and Internationalization



(Asst. Prof. Dr. Thara Angskun)

Dean of Institute of Social Technology

ตั้ง ชิงลี่ : การเทียบเคียงข้อสอบการอ่านเพื่อความเข้าใจของการสอบภาษาอังกฤษสำหรับ
นักศึกษาที่ไม่ได้เรียนวิชาเอกภาษาอังกฤษระดับ 4 กับกรอบมาตรฐานความสามารถทาง
ภาษาของประเทศในกลุ่มสหภาพยุโรป (LINKING COLLEGE ENGLISH READING
COMPREHENSION TESTS BAND FOUR TO THE COMMON EUROPEAN
FRAMEWORK OF REFERENCE FOR LANGUAGES (CEFR)) อาจารย์ที่ปรึกษา :
ดร.สิรินทร ศรีโพธิ์, 176 หน้า

เนื่องจากอิทธิพลของกรอบมาตรฐานความสามารถทางภาษาของประเทศในกลุ่มสหภาพ
ยุโรป (CEFR) ที่ส่งผลต่อการเรียน การสอนและการวัดผลความสามารถในการใช้ภาษาต่างประเทศ
ของผู้เรียนภาษาอังกฤษทั่วโลก การศึกษานี้จึงมีวัตถุประสงค์เพื่อเทียบเคียงข้อสอบการอ่านเพื่อ
ความเข้าใจของการสอบภาษาอังกฤษ สำหรับนักศึกษาชาวจีนที่ไม่ได้เรียนวิชาเอกภาษาอังกฤษ
ระดับ 4 (CET - 4) กับกรอบมาตรฐานดังกล่าว เพื่อให้ทราบถึงระดับความสามารถด้านการอ่านและ
ให้แนะนำเพื่อการพัฒนาข้อสอบการอ่านเพื่อความเข้าใจของข้อสอบ CET - 4

ข้อสอบการอ่านเพื่อความเข้าใจของการสอบภาษาอังกฤษ สำหรับนักศึกษาที่ไม่ได้เรียน
วิชาเอกภาษาอังกฤษระดับ 4 จำนวน 3 ชุด ได้ถูกนำมาวิเคราะห์เพื่อหาคุณลักษณะของข้อความและ
บทความที่ใช้ในข้อสอบการอ่านเพื่อความเข้าใจ โดยใช้ตารางเปรียบเทียบ (DUTCH CEFR GRID)
และทำตามขั้นตอนที่กำหนดไว้ในคู่มือที่เกี่ยวกับการเทียบเคียงการสอบอื่น ๆ กับกรอบ CEFR

ผลของการเทียบเคียงพบว่าข้อสอบการอ่านเพื่อความเข้าใจของ CET - 4 เทียบได้ในระดับ
B1 ของ CEFR และคุณภาพของข้อสอบดังกล่าวสามารถปรับปรุงให้เป็นไปตามกรอบ CEFR ได้
โดยการเพิ่มความถี่เกี่ยวกับเรื่องส่วนบุคคลมากขึ้น ควบคุมความยาวรวมถึงความยากของ
บทความที่ใช้สอบให้มีความเหมาะสมอย่างเป็นระบบ

สาขาวิชาภาษาต่างประเทศ
ปีการศึกษา 2562

ลายมือชื่อนักศึกษา Beigun Li
ลายมือชื่ออาจารย์ที่ปรึกษา A. Lu

DENG XINLI : LINKING COLLEGE ENGLISH READING
COMPREHENSION TESTS BAND FOUR TO THE COMMON
EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES
(CEFR). THESIS ADVISOR : SIRINTHORN SEEPHO, Ph.D., 176 PP.

CET-4 READING COMPREHENSION TESTS / THE CEFR / THE MANUAL /
THE DUTCH CEFR GRID / LINKING STUDY

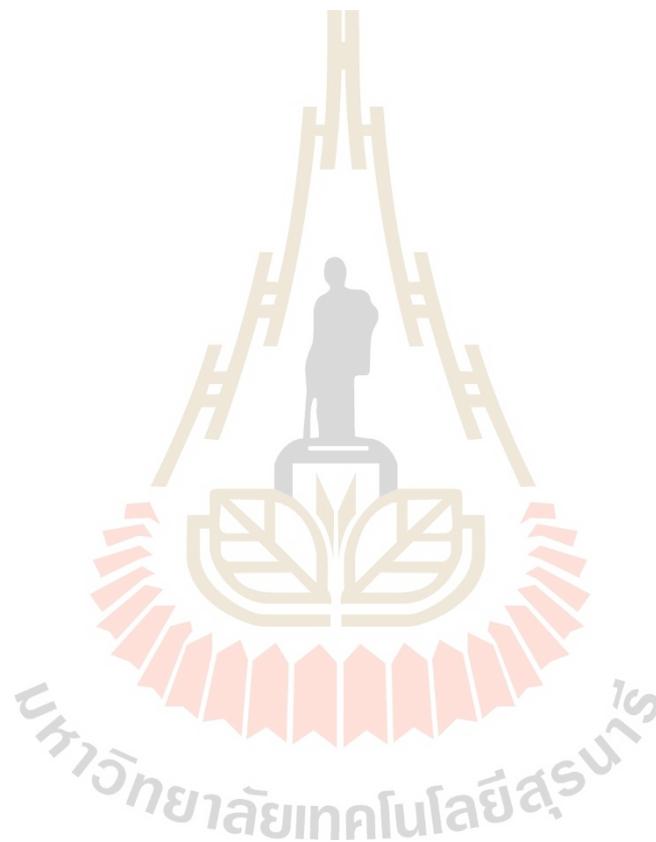
With the great impact of the Common European Framework of Reference for Languages: Teaching, Learning and Assessment (CEFR) on English language education worldwide, this study deals with linking the College English Test Band four (CET-4) Reading Comprehension Tests in mainland China to the CEFR. It aims at seeking the interpretation of the CET-4 test takers' language competence in terms of reading and providing suggestions for improvements to the CET-4 Reading Comprehension Tests.

Three sets of CET-4 Reading Comprehension Tests were analysed in order to find the characteristics of texts and items of CET-4 Reading Comprehension Test and to claim a link between the tests and the CEFR by following the procedures proposed in the CEFR Manual designed by the Council of Europe to relate tests to the CEFR. while analysing the content of the three sets of CET-4 Reading Comprehension Tests, the Dutch CEFR Grid was adopted at the second procedure.

The results indicated that CET-4 Reading Comprehension Tests generally correspond with the B1 level of the CEFR. In addition, in order to improve the CET-4 Reading Comprehension Tests, it is hope that test developers should put more

emphasis on the personal domain of texts. At the same time, text length of CET-4 Reading Comprehension Tests should be considered as well. To improve this factor, shortening the text length or lowering the difficulty level should be investigated.

In conclusion, the findings of the current study may shed light on the language competency of the CET-4 test takers in terms of reading and provide some suggestions for the further improvement of CET-4 Reading Comprehension Tests.



School of Foreign Languages

Academic Year 2019

Student's Signature Deng Xin Li

Advisor's Signature S. Li

ACKNOWLEDGEMENTS

I would like to show my sincere gratitude for all of those who have provided support for my study, research, and thesis here.

First and foremost, I would like to express my sincerest thanks to my supervisor Dr. Sirinthorn Seepho. During my study with her, not only she taught me the knowledge of professional courses and gave me crucial instructions on my academic writing, but also gave me the greatest encouragement and support in all aspects. Her patience, academic experience, and inspiration helped frame a core argument of my own to lead the thesis. Through numerous talks and discussions with her, I have learned to be more logical, precise, and concise in my writing. Without her help and support, technically and intellectually, I would not have completed my thesis with this quality and efficiency.

At the same time, I would also like to thank my thesis examination committee: Asst. Prof. Dr. Adcharawan Buripakdi (who served as the chair) and Asst. Prof. Dr. Kunlaphak Kongsuwannakul. They read my work carefully and gave me valuable comments and feedback to help me improve the quality of my work. Besides, I would like to express sincere gratitude to the teachers who helped me collect the data as analysts. Their names are not appeared in the acknowledgment page due to their concern of confidentiality. However, I do feel grateful for their friendly support. Without their kind participation, I would not be able to complete my research.

Deng Xinli

TABLE OF CONTENTS

	Page
ABSTRACT (THAI)	I
ABSTRACT (ENGLISH)	III
ACKNOWLEDGEMENTS	V
TABLE OF CONTENTS	VI
LIST OF TABLES	X
LIST OF FIGURES	XII
LIST OF ABBREVIATIONS	XIII
CHAPTER	
1. INTRODUCTION	1
1.1 Background of the study.....	1
1.1.1 English language assessment in China	1
1.1.2 Brief introduction to CET	3
1.1.3 Brief introduction to the CEFR.....	5
1.1.4 Gaps to be filled.....	7
1.2 Rationale of the study	8
1.3 Purposes of the study.....	11
1.4 Research questions	11
1.5 Significance of the study	11
1.6 Definitions of key terms	13
1.7 Summary of the chapter.....	14

TABLE OF CONTENTS (Continued)

	Page
2. LITERATURE REVIEW	15
2.1 Global English language tests	15
2.1.1 The International English Language Testing System (IELTS)	15
2.1.2 Test of English as a Foreign Language (TOEFL).....	17
2.2 Local English language tests in Mainland China	18
2.2.1 Test for English Major (TEM).....	18
2.2.2 College English Test (CET).....	19
2.3 The CEFR.....	30
2.3.1 Origin of the CEFR.....	30
2.3.2 Content of the CEFR	31
2.3.3 Impact of the CEFR.....	33
2.4 The Manual.....	37
2.4.1 Origin of the Manual	37
2.4.2 Content of the Manual	38
2.4.3 Impact of the Manual.....	41
2.5 Review of research on CEFR linking projects	42
2.5.1 Linking projects outside Asia	42
2.5.2 Linking projects in Asia.....	45
2.5.3 Linking project of CET with the CEFR.....	49
2.6 Summary of the chapter.....	50

TABLE OF CONTENTS (Continued)

	Page
3. METHODOLOGY	51
3.1 Research design	51
3.2 Research questions	52
3.3 Research Procedures	52
3.3.1 Familiarization	53
3.3.2 Specification	59
3.4 Research Context and Participants	61
3.5 Materials	64
3.6 Instruments	67
3.6.1 Descriptor-sorting activities	67
3.6.2 The Dutch CEFR Grid	67
3.6.3 Questionnaire	70
3.6.4 Semi-structured interview	72
3.7 Data collection	73
3.7.1 Qualitative data	73
3.7.2 Quantitative data	73
3.8 Data analysis	74
3.8.1 Qualitative data analysis	74
3.8.2 Quantitative data analysis	75
3.9 Summary of the chapter	76

TABLE OF CONTENTS (Continued)

	Page
4. RESULTS	77
4.1 Results of Familiarization Workshop	77
4.2 Results of Specification workshop	79
4.3 Summary of this chapter.....	90
5. DISCUSSION, IMPLICATIONS AND CONCLUSION	91
5.1 Summary of findings	91
5.2 Discussion	92
5.2.1 The Effects of Familiarization Workshop on Linking Consistency.....	92
5.2.2 Discussion of Specification workshop.....	94
5.2.3 Discussion of final results.....	96
5.3 Implications	101
5.3.1 Implications for CET-4 test developers	101
5.3.2 Implications for college English teachers.....	102
5.4 Limitations of the study and recommendations for future research	103
5.5 Conclusion.....	105
REFERENCES	106
APPDENDICES	115
CURRICULUM VITAE	176

LIST OF TABLES

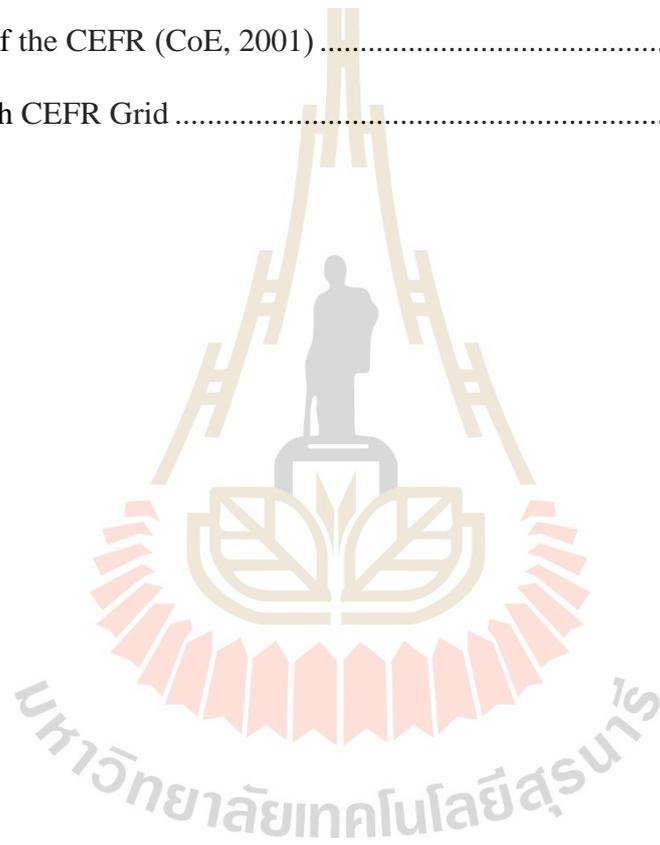
Table	Page
2.1 Test format of CET-4.....	23
3.1 Linking procedures	53
3.2 Preparatory pack for self-study.....	55
3.3 Estimated time for Familiarization activities.....	55
3.4 Estimated time for Specification workshop.....	59
3.5 Time and participants in the whole process.....	64
3.6 Texts and test items of CET-4 Reading Comprehension Tests	66
3.7 Vocabulary of the original Dutch CEFR Grid.....	68
3.8 Grammar of the original Dutch CEFR Grid.....	69
3.9 Operations of the original Dutch CEFR Grid.....	70
3.10 Operations of the revised Dutch CEFR Grid.....	70
3.11 Purposes of instruments.....	72
3.12 Data collection and analysis methods.....	76
4.1 Intra-rater reliability of descriptor-sorting activities.....	78
4.2 Inter-rater reliability of descriptor-sorting activities.....	78
4.3 Text sources of CET-4 reading texts	79
4.4 Authenticity of CET-4 reading texts.....	80
4.5 Discourse type of CET-4 reading texts.....	81
4.6 Domain of CET-4 reading texts.....	82
4.7 Topic of CET-4 reading texts	83

LIST OF TABLES (Continued)

Table	Page
4.8 Nature of content of CET-4 reading texts.....	83
4.9 Text length for CET-4 reading comprehension texts.....	84
4.10 CEFR level of vocabulary in CET-4 reading texts	85
4.11 CEFR level of grammar in CET-4 reading texts	86
4.12 Levels of texts estimated for CET-4 reading texts.....	87
4.13 Item type of CET-4 reading tests	88
4.14 Operations of CET-4 reading tests items	89
4.15 Estimated level of Items.....	90
5.1 Findings of the first research questions	91
5.2 Findings of the second research questions.....	92
5.3 A comparison of findings in the present study and with those of Huang and Jia’s study	96
5.4 Text length, vocabulaty and grammar of the second texts in three sets of test papers	99
5.5 B1 level of reading in the CEFR.....	100

LIST OF FIGURES

Figure	Page
1.1 Components of CET The CEFR levels (CoE, 2001).....	4
1.2 The CEFR levels (CoE, 2001).....	5
2.1 9 levels of the CEFR (CoE, 2001).....	32
2.2 The Dutch CEFR Grid.....	40



LIST OF ABBREVIATIONS

ALTE	Association of Language Testers in Europe
ACTFL	American Council on the Teaching of Foreign Languages
CEFR	Common European Framework of Reference for Languages: Learning, Teaching, Assessment
CET	The College English Test
CET-SET	CET-Spoken English Test
CFL	Chinese as a Foreign Language
CLARe	Center for Language Assessment Research
CLB	Canadian Language Benchmarks
COE	Council of Europe
CSE	China Standards of English
DLI	Defense Language Institute
EF EPI	EF English Proficiency Index
ELF	English as a Lingua Franca
EU	European Union
EFL	English as a Foreign Language
ETS	Educational Testing Service
FCE	First Certificate in English
FSI	Foreign Service Institute
HSK	Hanyu Shuiping Kaoshi

LIST OF ABBREVIATIONS (Cont.)

IB-CET	Internet-based CET
IELTS	International English Language Testing System
ISLPR	International Second Language Proficiency Rating
ILR	The Interagency Language Roundtable scale
MOE	Ministry of Education
MOET	Ministry of Education and Training
NCETS	National College English Teaching Syllabuses
NEEA	National Education Examinations Authority
NFL	The National Foreign Language
NMET	National Matriculation English Test
TEM	Test for English Majors
TestDaF	Test of German as a Foreign Language
TOEFL	Test of English as Foreign Language
TOEIC	Test of English for International Communication

CHAPTER 1

INTRODUCTION

The present study aims at linking the College English Test Band 4 (CET-4) Reading Comprehension Tests to the Common European Framework of Reference for Languages (CEFR) and investigating the characteristics of the CET-4 reading comprehension texts and test items by using the Dutch CEFR Grid. This chapter offers an introduction to the study by presenting the background of the study, and detailing the rationale, purposes, research questions and significance of the study. Lastly, definitions of the key terms are given.

1.1 Background of the study

1.1.1 English language assessment in China

Tests, in general, serve a large multitude of functions in society, such as encouraging higher levels of competence and knowledge, and measuring as well as improving the effectiveness of teachers and schools (Eckstein & Noah, 1993). Language tests have an impact on various aspects of language education, such as teaching and learning, which have been illustrated by a considerable number of research articles (Shohamy, Donitsa-Schmidt, & Ferman, 1996; Ferman, 2004; Cheng, 2005). Bachman (1990) stated that one of the most prevalent uses of language tests is for the purpose of evaluation in education programs, and this view is also accepted by Hughes (2007) who sees language tests as a means of measuring language proficiency. Admittedly,

countries possess different language tests owing to diverse test purposes, educational background and language policy.

China, which is the world's most populous country, possesses a long history of testing and examinations which can be traced back to the imperial period nearly two thousand years ago with the Han Dynasty (206 BC to AD 220). However, it was not until the late 1970s that modern testing theory and practice were developed in China (Fan, Ji, & Song, 2014). Since the 1970s, following the resumption of the College Entrance Examination (known as *gaokao* in China) after an interruption of 10 years, China has witnessed a rise of interest in English language education and a massive growth in the number of English learners because of reforms and an opening-up policy. More than 225 million students take part in formal education at primary, secondary and tertiary undergraduate levels in which English language study is compulsory, which makes China the country with the largest number of English language learners and test takers throughout the world currently (Kunnan, 2014). The educational experiences of students in China involve numerous high-stakes tests of English in order to enter educational programs and/or obtain employment, each of which has far-reaching influences on their prospects for advancement, studies and careers (Cheng & Crutis, 2010). Thus, English language assessment is a subject of great importance in China.

This context has facilitated some excellent research to establish a series of large-scale, domestically developed and administered high-stakes English language assessments in China, such as the National Matriculation English Test (NMET), the College English Test (CET), the Test for English Majors (TEM), the Graduate School Entrance English Examination (GSEEE), the Public English Testing System (PETS), and the Cambridge Business English Certificate (BEC) to name a few. Of these, the

CET is a large-scale standardized test which has the most intense impact on college English education, on individuals, and on society (Sun, 2016).

1.1.2 Brief introduction to CET

CET is administered nationwide by the National College English Testing Committee on behalf of the Higher Education Department of the Ministry of Education (MoE). The test is used as a means for examining and promoting the implementation of the National College English Teaching Syllabus (NCETS) which was the first syllabus in Mainland China to improve the level of English language education in tertiary institutions.

As the number of people taking the test has increased rapidly, the test has become the largest EFL test throughout the world (Jin, 2005; Zheng & Cheng, 2008). The test plays a crucial role in Chinese college English teaching and in the academic success of thousands of graduates, their careers and even their personal development. For example, the CET pass rate is considered to be one of the most essential criteria for evaluating the teaching quality of higher education institutions (Ma, 2014). Consequently, it exerts a huge impact on Chinese college English teaching and learning. Furthermore, the CET-4 was mandatory for college graduates for a long time, and only those who had passed CET-4 were able to obtain their diplomas. In addition, the impact of the test on the career development of college students was of great importance because of its wide recognition by employers and even society as a valid instrument to access college education levels (Jin, 2010). For example, college graduates with CET certificates are more highly preferred by companies or institutions (Ma, 2014). For some companies especially multinational companies, foreign companies, organizations or institutions, CET has become one of the primary prerequisites for recruiting applicants

and even for promotion in state-owned work places in Mainland China.

CET consists of two types of tests, namely the CET paper-based test and the CET-Spoken English Test (CET-SET) as shown in Figure 1.1. The CET paper-based test comprises the College English Test Band 4 (CET-4) and the College English Test Band 6 (CET-6). Similarly, CET-SET also comprises CET-SET 4 and CET-SET 6 (Figure 1.1). Based on the National College English Test Syllabus (2016), CET is administered twice a year, on the third Saturday in June and the first Saturday in December, respectively. The CET-SET, on the other hand, is administered in May and in November (Wu, 2014). The CET-4 has a core of four tests, which are compulsory for all students in their second year of university study.

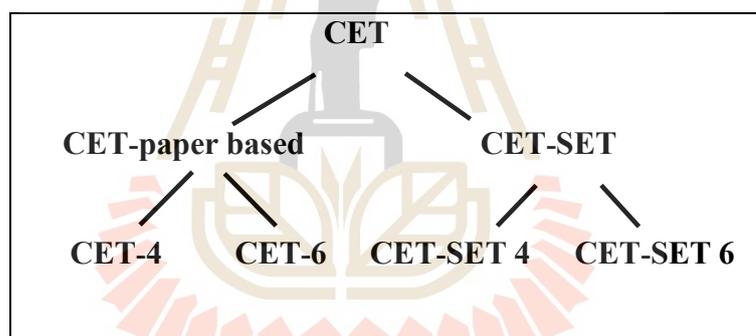


Figure 1.1 Components of CET

During the past 30 years, the test has been reformed many times to meet the English language needs of society in Mainland China. Furthermore, Ma (2014) advocated the idea of substantially modifying CET to be better aligned with prevailing international standards. Even though a lot of effort has been constantly exerted to improve the quality and acceptability of the test, it is still difficult to assess students' English language ability based simply on their test scores. Consequently, it has become a matter of both necessity and urgency to assess the test scores and improve the

transparency of CET certificates through taking action which will lead to wider international recognition.

1.1.3 Brief introduction to the CEFR

In 2001, an internationally influential document in language education entitled the Common European Framework of Reference for Languages: Learning, Teaching, and Assessment (CEFR) was published by the Council of Europe (CoE). The CEFR consists of a descriptive scheme in terms of language use and learning and it comprehensively elaborates the language competency that learners require for the purpose of using language to communicate and the necessary knowledge as well as the skills they need in order to be able to act effectively (CoE, 2001). In this framework, language learning is divided into six levels from A1 to C2 arranged in three broad bands - A1 and A2 (Basic User), B1 and B2 (Independent User), C1 and C2 (Proficient User) as shown in Figure 1.1.

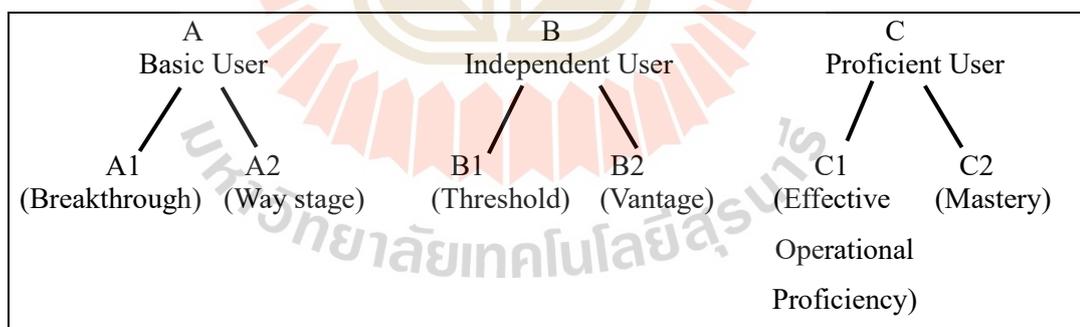


Figure 1.2 The CEFR levels (CoE, 2001)

One of the objectives of the CEFR is to promote as well as to strengthen cooperation among different educational institutions in countries by providing a common reference standard for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc., across Europe” (CoE, 2001, p.1). Another

objective that the CEFR is specifically concerned with refers to testing and examinations. It aims at “helping describe the levels of proficiency required by existing standards, tests and examinations so as to facilitate comparisons between different systems of qualifications” (CoE, 2001, p.21). Over nearly two decades, this framework has had a dramatically significant influence on language education, particularly in the field of testing both in Europe and beyond. As Little (2006) described, the impact of the CEFR on language testing far exceeds its impact on other aspects of language education such as curriculum design and pedagogy.

Because of the rapidly increasing influence of the CEFR, some institutions and test providers started to relate their examinations to the CEFR to achieve transparency in their examinations and to improve their conformity to this framework. However, there was the problem raising of how to ensure concordant interpretations of the levels in disparate contexts. In response to the need to assist test providers in linking their certificates and diplomas to the CEFR in a reliable way, the CoE developed a Manual two years after the publication of the CEFR. With several years of study and practice, based on the pilot version, the Manual of Relating Language Examinations to the Common European Framework of Reference for Language: Learning, Teaching, Assessment (Manual) was officially published by the CoE in 2009.

In order to help examination providers to situate their examinations in relation to the CEFR, a set of five clear and coherent procedures linking examinations or tests were clearly provided in the Manual. For example, the first procedure (Familiarization) aims at providing some activities for familiarizing analysts with the CEFR, and the second procedure (Specification) offers guidance on describing the examination coverage and relating results reported from the examination to the CEFR levels.

Not surprisingly, some institutions and test providers started to relate their examinations, such as Test of German as a Foreign Language (TestDaf), Graded Examinations in Spoken English (GESE) and Integrated Skills in English (ISE) to the CEFR. More importantly, linkage with the CEFR was also established between those tests which have profound international influence such as the International English Language Testing System (IELTS), the Test of English as a Foreign Language (TOEFL), the Test of English for International Communication (TOEIC), and the *Hanyu Shuiping Kaoshi* (HSK) Chinese Proficiency Test.

1.1.4 Gaps to be filled

As explained above, the CEFR has had a great impact on language education in numerous countries. A large number of examinations and tests have been linked to it to establish the validity of the language certificates. Interestingly, a relatively small number of research studies have been conducted in Mainland China in this respect, particularly concerning the development of a linkage between large-scale local English tests and the CEFR. More specifically, little has been done so far on linking CET to the CEFR although the researcher has reviewed related studies in this regard. A study by Huang and Jia in 2012 investigated the feasibility of linking CET with the CEFR. However, in this study, the specification procedure outlined in the Manual was adopted. The results of this present study reveal that the CET content is basically consistent with the CEFR, and also that the CEFR is applicable in a Chinese foreign language testing context. Thus, CET can be linked to the CEFR. Furthermore, the textual origin, types, dimension, and topics in the CEFR and CET are highly consistent.

In the study conducted by Huang and Jia (2012), the two most serious limitations are, on the one hand, the lack of the first procedure provided in the Manual

to link an examination or test to the CEFR, namely, the Familiarization procedure. As the Manual states in several places such as on pages 7, 9, 10, and 17 (CoE, 2009) familiarization with the CEFR and the relevant scales serves a fundamental and indispensable stage before carrying out the subsequent procedures. It provides researchers with an in-depth and comprehensive knowledge of the CEFR, not just a superficial knowledge on a global scale. On the other hand, the methodology section in Liu and Jia's study seems problematic, because the details about the participants and each procedure have not been clearly described. Neither the number of the participants nor the specific procedure followed concerning the linkage was not described in their study. As suggested in the Manual (CoE, 2009), the results provided by the specification phase will be more reliable and trustworthy if more than one person is involved. The results obtained from this study thus seem to be deficient and need to be further verified.

In order to bridge the gaps in Huang and Jia's study, the current study, therefore, aims at analyzing the content of the CET-4 Reading Comprehension Tests and then linking them to the CEFR by following the procedures specified in the CEFR Manual.

The major reason for selecting the reading comprehension section in this study is it consists of 35% of the total score in CET-4 and it also accounts for one of the largest proportion of college English teaching. Additionally, reading, as a receptive skill, is an essential and important aspect of English teaching and learning.

1.2 Rationale of the study

One of the main reasons to conduct the present study is the critical importance of CET in the Chinese context and the great impact and far-reaching influence of the

CEFR on language learning, teaching and particularly assessment worldwide. Numerous countries have adopted or adapted this framework to their own language education contexts and positive impacts have been demonstrated by a considerable number of research studies (North, 2000; North, 2007; Hulstijn et al. 2010; Fu, 2011). Furthermore, a series of language tests possessing significant international influence such as TOEFL, IELTS, TOEIC and HSK etc., have been linked to the CEFR levels. Thus, by reviewing the quality of the CET-4 Reading Comprehension Tests and exploring the level of the CET-4 Reading Comprehension Tests with regard to the CEFR, the present research project aims at linking the CET-4 Reading Comprehension Tests to the CEFR by following the guidelines of the CEFR Manual. After reviewing the academic literature on the study of the CEFR, some reasons for research and some research gaps have been found:

On the one hand, studies on CEFR outside Mainland China not only focus on the discussion of its origin, content and impact (Little, 2006; Figueras, 2012), challenges and limitations (Weir, 2005; Alderson, 2007; Westhoff, 2007; North, 2014), but also deal with the alignment of different kind of tests with the CEFR levels (Khalifa & French, 2009; Khalifa, Frenchman & Clamour 2010; O'Sullivan 2010; Papergirl, 2010; Green, 2019; Fleckenstein et al, 2020). However, little has been done in Mainland China to link CET to the CEFR (Zou et al., 2015). As described above, the first attempt on this topic was made by Huang and Jia in 2012, which serves as a preliminary study for linking a language test to the CEFR in Mainland China. In the same year, Liu (2012) conducted a study to align TEM with the CEFR. Furthermore, similar research in terms of linking TEM to the CEFR was conducted by Wang in 2015 by taking Reading Comprehension Test as samples. In 2016, the NMET was aligned with the CEFR levels

by Zou who followed the four procedures specified in the Manual. As Zou et al. (2015) stated, research on linking language tests to the CEFR for Chinese scholars is still in the exploratory stage and more research studies are urgently needed to back up the results reported from these studies. The present study employed Huang and Jia's study as a conceptual framework for investigating the feasibility of linking CET to the CEFR. The similarities and differences between these two studies are described and discussed in later chapters.

On the other hand, researchers in Mainland China have pointed that the CEFR provides experience and reference for language education. For example, Wang (2012) stated that the CEFR can be regarded as a reference for Chinese foreign language tests and it can play a vital role in the improvement of the quality of foreign language tests. Some researchers revealed in their study (Yang & Gui, 2007; Yan, 2008) that using the CEFR as criteria to assess Chinese undergraduates' English proficiency would be helpful for English education in Mainland China to reach an international standard. Similarly, Zou et al. (2015) underlined the importance of the CEFR in Chinese foreign language education by expressing the need for conducting thorough research on the CEFR. Zou et al. (2015) pointed out that it is the mission of language educators and EFL teachers to apply the CEFR into Chinese foreign language education innovatively in order to be in line with international standards in Mainland China, and one of the applications of the CEFR refers to linking large-scale standard tests to this framework. As discussed in 1.1, the huge number of CET test-takers makes it the largest EFL test in the world, and the test itself has been widely and vigorously discussed among academics in Mainland China (Zheng & Cheng, 2008). Some scholars pointed out the limitations of the test and they have called for its reform. Thus, it is urgent to conduct

the present study with the purpose of improving the quality of the test to reach international standards.

1.3 Purposes of the study

With regard to the problems addressed in 1.1, this study aims at linking the content of the CET-4 Reading Comprehension Tests to the CEFR levels with the help of the procedures provided in the Manual. In detail, the purposes of this study are as follows:

1.3.1 To explore the level of the CET-4 reading comprehension texts in the CEFR and the major characteristics of these reading texts.

1.3.2 To explore the level of the CET-4 reading comprehension test items in the CEFR and the major characteristics of the test items.

1.4 Research questions

Based on the purposes explained above, the following questions are formulated:

1.4.1 Which level are the CET-4 reading comprehension texts in the CEFR and what are the major characteristics of the reading texts?

1.4.2 Which level are the CET-4 reading comprehension test items in the CEFR and what are the major characteristics of the test items?

1.5 Significance of the study

Since the launch of CET in Mainland China, it has undergone constant reform and adjustment with regard to content, format, and test scores. Recent reforms of CET-4 were introduced in 2013 and 2016, respectively, which has led to some changes in the test content and format. However, few researchers have examined or studied the quality

of the newly reformed CET-4 in terms of its validity, reliability and test constructs. As described in the Manual, the procedures provided in the Manual offer the means not only to analyze the content of the test and provide evidence of the quality of the tests in a detailed way, but also to provide guidance for item writers and evidence of the relation between the test and the CEFR. Therefore, the present study offers an opportunity to review the quality of the newly reformed CET-4 Reading Comprehension Tests and then to provide some suggestions for its further improvement. Thus, the results of the present study will enable the stakeholders who are responsible for the development of CET-4 to consider the need to improve as well as upgrade the CET-4 test so that it will be internationally recognized and meet international standards.

On the other hand, it is generally accepted in China that the function of the test serves as an impartial indicator of students' academic success (Cheng, 2005). CET, as a high-stake test, measures the English language ability of college students in accordance with the NCETS. Consequently, it has a profound effect on university or college English teaching and learning as well as assessment (Cheng, 2005). Therefore, by linking the CET-4 Reading Comprehension Tests to the CEFR, this paper will provide Chinese students and EFL teachers with the opportunity to understand what a particular CEFR level means in respect of learners' reading proficiency and how a CET-4 test-taker is able to perform in terms of the standards detailed in the CEFR levels. Consequently, teachers will be able to gradually consolidate the theoretical criteria with the teaching and assessment practices they employ in the classroom.

1.6 Definitions of key terms

The followings are the definitions of two key terms in the present study: Familiarization and Specification.

1) CEFR: The Common European Framework of Reference for Languages:

Learning, Teaching, and Assessment, abbreviated in English as CEFR or CEF or CEFRL was published by the Council of Europe in 2001. It is a common framework of reference for language learning, teaching and assessment and a guideline used to describe the achievements of learners of foreign languages across Europe and, increasingly, in other countries. The framework comprises a descriptive scheme for language use and learning and describes language ability at six levels arranged in three bands-A1 and A2 (Basic User), B1 and B2 (Independent User), and C1 and C2 (Proficient User). It describes what knowledge and skills language learners have to develop at each level. The framework has been now translated into more than 40 languages and adopted or adapted in numerous countries, and an increasing number of countries are using it for reference or adoption for different language education contexts.

2) CET: The full name of CET is the College English Test. It is a large-scale standardized test administered nationwide by the National College English Testing Committee on behalf of the Higher Education Department of the MoE in China. CET was first administered in 1987 to promote the implementation of the National English Teaching Syllabus. Additionally, the test also aims at measuring the English proficiency of college or university undergraduate students in accordance with the National English Teaching Syllabus. This test is administered twice a year, in January and June, respectively. It consists of two types of tests, namely the college English paper-based test and the CET-Spoken English Test (CET-SET). The college English paper-based test

comprises College English Test Band 4 (CET-4) and College English Test Band 6 (CET-6). Similarly, CET-SET comprises CET-SET 4 and CET-SET 6. CET plays a critical role in the achievement of students and in obtaining a job and career development at numerous Chinese colleges or universities.

3) Familiarization: Familiarization refers to selecting the training activities designed to train those who wish to take part in the linking activities to have an in-depth understanding of the CEFR. There are a lot of activities for Familiarization such as introductory activities, qualitative analysis of the CEFR scales, preparing for rating and conclusion, all of those are clearly described in chapter 3 of this study. However, some of the activities selected might differ in terms of the level and purpose of the linking activities.

4) Specification: Specification refers to guidance on describing the examination coverage and relating the results reported from the examination to the CEFR levels.

1.7 Summary of the chapter

This chapter gives an overall description of the present study. Firstly, the considerable worldwide impact of CEFR was addressed. Then, the rationale of the study, purposes of the study, research questions, and significance of the study were described. Lastly, definitions of the key terms in the current study are given. In the next chapter, the origin, content and the impact of CEFR and the Manual are discussed. Additionally, relevant research studies both in Mainland China and abroad are also presented. Finally, CET-4 in China and related studies concerning test validity, reliability and linking activities between CET-4 and the CEFR are described.

CHAPTER 2

LITERATURE REVIEW

This chapter reviews the related literature on CEFR research and locates the present study in the context of previous studies. It consists of four sections. The first section is mainly concerned with a review of well-known global language tests including TOEFL and IELTS. The second section details two local English language tests in Mainland China, namely, TEM and CET. The third and fourth sections elaborate the origin, the content as well as the impact of the CEFR and its Manual. The last section presents some previous studies on CEFR, particularly linking tests to CEFR both outside and inside Asia.

2.1 Global English language tests

In the world of foreign language education, the two most famous international standardized tests of English are TOEFL and IELTS. These two tests have a strong influence on second and foreign language learners who would like to measure their language proficiency especially in an academic context. Many local or institutional tests have tried to equate their scores with these two international tests.

2.1.1 The International English Language Testing System (IELTS)

The International English Language Testing System (IELTS), is one of the prime English-language tests in the world and it is an international standardized test of English language proficiency for non-native English language speakers. It was

launched jointly in 1980 by Cambridge English Language Assessment (then known as UCLES) and the British Council. The test is accepted by different academic institutions as well as various professional organizations in a large number of countries such as Australia, Britain, and the United States, as a standard for study, work and immigration.

IELTS consists of two main versions including IELTS Academic and IELTS General Training. IELTS Academic is intended for those wanting to enroll in universities or other institutions of higher education and for professionals. IELTS General Training is designed for those who plan to undertake non-academic training or to gain work experience, or for immigration purposes. These two types of tests are completely identical in terms of their listening and speaking sections. The differences lie in IELTS Reading and IELTS Writing. To be more specific, IELTS General Training reading passages are shorter on average and more varied, covering more “work and living” subject matter. As for writing, the task of IELTS General Training is far less academic, such as writing a letter instead of summarizing a scholarly source.

There are four skills tested in IELTS including speaking, listening, reading and writing. The exam is divided into two major sections and the first section is a paper exam in which listening, reading and writing are completed in one sitting. The speaking test may be taken on the same day or up to seven days before or after the other tests. To pass the test, no minimum score is required. Band scores are used for each language sub-skill (Listening, Reading, Writing, and Speaking) and the band scale ranges from 0 (“Did not attempt the test”) to 9 (“Expert User”). Different institutions or organizations set their thresholds according to their requirements. In addition, institutions are advised not to accept a report older than two years to be valid, unless the test takers demonstrate that they have worked to maintain their levels.

The number of IELTS test takers is extremely large and continues to increase each year. In 2017, more than 3 million tests were taken in over 140 countries, up from 2 million tests in 2012, 1.7 million tests in 2011 and 1.4 million tests in 2009. In 2007, IELTS administered more than one million tests in a single 12-month period for the first time ever, making it the world's most popular English language test for higher education and immigration.

2.1.2 Test of English as a Foreign Language (TOEFL)

Apart from IELTS, the Test of English as a Foreign Language (TOEFL) is one of the major English-language tests in the world as well and these two types of tests share a range of similarities which mainly lie in their format and acceptance. To be more specific, TOEFL is a standardized test aiming at measuring the English language ability of non-native speakers who intend to enroll in English-speaking universities. Similarly, the test is accepted by many English-speaking academic and professional institutions.

The test was first administered and designed by the Education Testing Service (ETS), an organization that specializes in standardized tests for higher education in the USA. Similarly to IELTS, the format of TOEFL can be divided into four sections, namely, speaking, reading, writing and listening. Each section is scored between 0-30 and then these are added together to obtain a total score of between 0-120. ETS issues official score reports which are sent independently to institutions and are valid for two years following the test.

There are three types of tests including computer-based tests (CBT), paper-based tests (PBT) and the TOEFL Internet-based Test (iBT). Currently, CBT and PBT have gradually been replaced by the iBT format after its introduction in late 2005. Even though paper-based testing is still utilized in some regions, the TOEFL iBT test has

been introduced in phases, with some countries, such as Canada, France, Germany, and the United States in 2005 and the rest of the world in 2006, with test centers added regularly. In September 2006, the CBT was discontinued and the scores are no longer valid. TOEFL scores are still required for university admission in the US.

Though these two tests are accepted worldwide, ETS, Cambridge English Language Assessment and the British Council still conduct studies to set equivalent cut off scores for CEFR to align them.

2.2 Local English language tests in Mainland China

2.2.1 Test for English Major (TEM)

The Test for English Major (TEM) is a national undergraduate examination which aims at measuring the English proficiency of Chinese university undergraduates who major in English Language and Literature as well as examining the National College English Teaching Syllabus for English Majors (Syllabus, NACFLT, 2000). The Syllabus divides the four-year English major undergraduate programme into the foundation stage (the first and second year) and the advanced stage (the third and fourth year). The TEM test battery correspondingly consists of TEM-4 and TEM-4 Oral, and TEM-8 and TEM-8 Oral. TEM-4 and TEM-4 Oral assess students' English proficiency at the end of the foundation stage, and TEM-8 and TEM-8 Oral assess s proficiency at the end of the advanced stage.

The TEM is administered by the National Advisory Committee for Foreign Language Teaching (NACFLT) on behalf of the Higher Education Department, Ministry of Education, People's Republic of China. The four tests in the battery are all administered once a year with TEM-4 in April, TEM-8 in March, TEM-4 Oral in May

and TEM-8 Oral in December. The total test time is 135 minutes for TEM-4 and 195 minutes for TEM-8. Each oral test takes around 25 minutes to complete.

The TEM test scores are reported to the Academic Affairs Office of the participating universities. In the case of TEM-4 and TEM-8, individual test takers scoring 60 or above receive a certificate from the NACFLT on which their level of performance is reported, including 'excellent' (score 80 or above), 'good' (score between 70 and 79) and 'pass' (score between 60 and 69). Neither composite scores nor section scores are reported to test takers. However, the test takers can check their composite scores through the Academic Affairs Office of their university. For the two oral tests, test takers who pass the tests are awarded a separate certificate from the NACFLT on which the same three levels are reported: 'excellent', 'good' and 'pass'. The levels are converted from the average of the total raw scores awarded by two TEM authorized oral examiners.

2.2.2 College English Test (CET)

Description of CET

a) Launch of CET

The launch of CET can be dated back to the 1970s after the implementation of China's reform and opening-up policy. Since then, with the entry of many foreign companies and the steady growth of China's foreign trade, batches of qualified personnel especially those who are proficient in English have been urgently required in China (Ma, 2014). Internationally, with regard to the prestigious status of English all over the world, the central government thus started to take actions in the fields of economics as well as education to accelerate its integration with the rest of the world. Additionally, given the fact that college graduates usually serve as the essential

labor force driving the substantial economic and technological development of the country, the State Education Commission (former MOE) began attaching great importance to English education at the higher levels. Therefore, the NCETS (Science and Technology) in 1985 and NCETS (Arts and Science) in 1986 were gradually implemented after a large-scale longitudinal investigation by a team of senior English scholars (Jin, 2010). The move played a critical role in the development of college English education by setting teaching aims and requirements for college English education. The Syllabuses divided college English courses into 6 bands, and each band was equal to one semester's formal college English study. Among the 6 bands, 1 to 4 are compulsory for all non-English majors and bands 5 to 6 are elective. In order to approve the Syllabus by the State Education Commission, standardized tests were administered to students when they finished Band 4 or 6. Consequently, CET-4 was initially launched in 1987 and CET-6 in 1989 in order to check how well the two syllabuses were implemented and to provide feedback for College English teaching at the same time (Ma, 2014).

The series of CET tests are intended for students of all academic disciplines, especially for those who have completed corresponding courses in higher education institutions, except in English language and literature at tertiary-level education in Mainland China. In general, second-year students can take CET-4 as long as they have completed College English Courses Bands 1 to 4, and students who have completed CET-4 are eligible for CET-SET 4. To take CET-6, students need to complete the College English Courses Bands 5 to 6 as well as passing CET-4, and similarly CET-SET 6 is only eligible to those who have finished taking CET-6. If students fail their tests the first time, they can continue taking them before graduation.

b) Reform process of CET

Although CET in Mainland China is well established, there remain some limitations to the test according to the results reported by some studies. For example, Han (2002), pointed out a series of problems existing in college English teaching for undergraduates who had passed CET, such as students taking a long time consuming and students with high scores but low ability. He recommended that one of the most important measures for promoting the quality of college English teaching was the radical reform of CET. Furthermore, Han, Dai and Yang (2004) reported in their survey that most of the English teachers (77.9%), with a total number of 1194 from 40 universities and colleges in Mainland China, did not consider that there was a correlation between the CET certificate and learners' actual language ability. Therefore, as a result of these problems and limitations revealed by numerous studies, improvements and reforms have been constantly introduced on CET in order to improve its quality and to better promote the implementation of the College English Teaching Syllabus.

So far, CET has passed through several important stages of development in terms of changes in its content and format. CET was developed in 1987 with no speaking part, and too much importance was placed on reading. In 1997, the test format was changed to include dictation for listening and a translation part from English to Chinese for reading. In 1999, with increasing demand for improved oral English communication, the CET-SET was initiated. In 2005, CET was reformed once again with substantial changes including test content, formats and a scoring system. The maximum score of CET was 100 points before 2005. Those college students who passed CET-4 or CET-6 were issued a certificate by NCETC on behalf of the Higher Education

Department of the MOE. More specifically, two kinds of certificates were issued. Out of 100 points, a test score that is higher than 85 is regarded as "Excellent" and higher than 60 but below 85 is regarded as "Pass" in the certificate. The reported score of the CET shows whether a candidate has met the requirements of the College English Teaching Syllabus. In the reform of 2005, the 100 score system was changed to a norm-referenced system with scores ranging from 220 to 710. On the other hand, the pass score and qualification certificate were eliminated, and a detailed score report was sent to candidates. In addition, the percentage of the listening part was increased from 20% to 35%. In 2007, the Internet-based CET (IB-CET) was introduced. In 2013, changes were made again including the elimination of the Cloze part. In 2012, three sets of tests were administered to candidates in one examination. In 2016, the CET-SET was also reformed several years after its introduction. Only those who had either passed CET-4 with a score of more than 550 (80 before 2005) or passed CET-6 with a score of at least 520 (75 before 2005) were eligible for the test before 2016. However, the newly reformed CET-SET indicates that all the graduates who have taken CET-4 or CET-6 are eligible to take the corresponding CET-SET if they wish. In other words, college students who have taken CET-4 are eligible for CET-SET 4, and those who have finished taking CET-6 can register for CET-SET 6. The CET-SET score is reported according to a letter grade scale: A, B, C and D, with A representing the highest grade level and D the lowest, and only those who have achieved A, B and C levels are awarded certificates.

Format and content of CET-4

Since the present study is concerned with linking CET-4 Reading Comprehension Tests to the CEFR, only the CET-4 format and content are presented.

The current test content and format in CET-4 include 1) Writing 2) Listening 3) Reading and 4) Translation as shown in Table 2.1.

Table 2.1 Test format of CET-4

Structure	Content	Question types	Question number	%	Test time
Writing	writing	a short essay	1	15%	30 mins
Listening comprehension	short conversations	multiple choice	7	7%	25 mins
	long conversations	multiple choice	8	8%	
	passage comprehension	multiple choice	10	20%	
Reading comprehension	banked cloze	cloze	10	5%	40 mins
	long passage comprehension	matching	10	10%	
	reading in depth	multiple choice	10	20%	
Translation	from Chinese to English	passage translation	1	15%	30 mins
Total			57	100%	125 mins

Note: % represents score percentage of CET-4

a) Writing

This section aims to assess students' ability to communicate ideas using written English (College English Test Syllabus, 2016). Prompts including outline, scenarios, pictures or charts are provided to students and then they are required to write an essay or paragraph of no less than 120 words based on the input material provided. This section constitutes 15% of the total score and the time is 30 minutes.

b) Listening

The listening section aims to assess students' abilities to understand oral conversations or passages (College English Test Syllabus, 2016). It comprises three parts including Short Conversation, Long Conversations and Passages Comprehension. There are three Short Conversations (450-500 words in all), two Long Conversations (240-280 words per conversation) and three Conversation Passages (220-240 words per passage). For Short Conversations, 2-3 questions (7 questions in total) are presented at the end of each news item. For Long Conversations, 4 questions for each (8 questions

in total) are presented. For Conversation Passages, 3-4 questions for each (10 questions in total) are formulated. Typescripts are based on standard American English or standard British English and the speed is approximately 120-140 words per minute (wpm). Both news reports and the questions are spoken only once and then students will have 15 seconds for each question. Multiple choice is used in this section. The listening comprehension contributes 35% to the total score, of which 7% is for Short Conversations, 8% is for Long Conversations, and 20% is for Conversation Passages. The time for this section is 25 minutes.

c) Reading

The reading section aims to assess students' ability to understand English in a written format (College English Test Syllabus, 2016). It comprises four passages including one Banked Cloze (200-250 words), one Long Passage Comprehension (approximate 1000 words) and two Reading Passages in depth (300-350 words). This section accounts for 35% of the test score, of which 5% is for the Banked Cloze test, and 10% is for the Long Passage Comprehension, and 20% is for Reading in depth. The time is 40 minutes for this part.

The Banked Cloze test aims at examining candidates' ability to understand and use words in a language context. In this section, a passage with ten blanks is provided to the candidates. Based on their understanding of the passage, candidates are required to select one word for each blank from a list of 15 choices given in a word bank following the passage and then mark the corresponding letter on an Answer Sheet.

The Long Passage Comprehension aims at examining candidates' ability to obtain information through Scanning and Skimming. Paragraph matching is

used in this part. Skimming requires candidates to extract the main idea from a text and Scanning requires candidates to find specific information at a high rate of speed in a text. Candidates are required to read a passage with at least ten sentences. Each sentence contains information given in one of the paragraphs. Candidates must choose the paragraph from which the information is derived and then mark the corresponding letter on an Answer Sheet. It is worth noting that some paragraphs can be chosen more than once.

Reading in depth aims at examining candidates' reading comprehension ability at different levels including understanding the gist of the passage, important details, a comprehensive analysis, and guessing the meaning of words according to context. This section comprises 2 passages and each passage is followed by some questions or unfinished statements (5 questions for each passage). Multiple choice is used in this section and candidates have to select the best choice and then mark it on an Answer Sheet.

d) Translation

Translation aims to assess students' ability to apply appropriate strategies and language knowledge to express a Chinese paragraph in English. Students are asked to translate a paragraph from which unfamiliar terminology and words are excluded from Chinese into English within the stipulated time. The paragraph generally refers to Chinese culture, history, geography or economic development. This section takes 30 minutes and it contributes 15% to the total score.

e) CET-SET 4

CET-SET, an optional component of CET, is administered independently of CET-4 and CET-6. The test uses a computerized-based testing form

to assess students' interactional competence. Analogue examiner and examination questions are presented on the screen, and the examination questions include texts or figures (e.g. drawings, pictures, diagrams.).

Candidates work in pairs, which are generated randomly by the computer system, and interact with the examiner and the other candidate. The test is composed of a warming-up activity followed by four tasks and it takes 15 minutes in total.

The warming-up activity aims to prepare candidates for the following four tasks. The two candidates who are arranged in a group have approximately 20 seconds each to make a brief self-introduction in turn. This task takes one minute.

Task one is Read Aloud. The candidates are given 1 minute to read a 120-word passage after preparing for 45 seconds. The preparation task is conducted at the same time for both candidates and it takes two minutes.

Task two is Short-answer Questions. The candidates are given 40 seconds to answer two questions based on the passage they have just read. This task is conducted at the same time for both candidates and it takes one minute.

Task three refers to a personal statement. The candidates are given 1 minute each to make a personal statement based on the prompts after preparing for 45 seconds. As for task one and task two, this is also conducted at the same time for two both. This task takes two minutes.

Task four refers to double interaction. The two candidates are given 3 minutes to talk to each other about the given scenarios and tasks after preparing for 1 minute. This task also takes 4 minutes.

The evaluation of the students' performances of CET-SET depends on

six criteria:

1) Accuracy in pronunciation, stress/intonation and use of grammar and vocabulary; 2) Vocabulary and grammatical structures; 3) Each individual candidate's contribution to the group discussion; 4) Discourse management; 5) Flexibility in dealing with different situations and topics; 6) Appropriate use of linguistic resources according to context (National College English Testing Committee, 2006).

Research studies on CET-4 in Mainland China

An overview of the previous studies on CET-4, reveals that the examination board, test developers, and college English teachers and students are mainly concerned about 1) its nature, status and impact, especially the washback effect on college English teaching (Jin, 2005; Qiu, 2006; Shao, 2006; Zheng, & Cheng, 2008; Wu, 2014; Fan & Song, 2015; Sun, 2016), 2) its disadvantages and calls for reforms to the test (Han, 2002; Huang & Wu, 2004; Jin, 2005; Hou, 2010; Ma, 2014), 3) its reliability and validity (Yang & Weir, 2001; Zhong, 2007; Sun & Xie, 2015), and 4) a comparative analysis of the four sub-skills between IELTS and TOEFL (Qiu, 2006; Li, 2008; Gao, 2011). Recently, as a result of the international impact of CEFR, researchers have also attempted to investigate the feasibility of aligning CET-4 with CEFR (Huang & Jia, 2012). However, compared with other research studies about CET-4, studies in this regard are relatively rare. Furthermore, CET-4 was reformed in 2016 and changes were made to the content and format in the reading section. No research study has been found about linking the newly reformed CET-4 to CEFR.

Validity of CET-4

Validity and reliability are the most important factors to take into consideration as these two elements are the test-internal constructs when designing a

test. Validity refers to “the extent to which references made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment” (Gronlund, 1998). Validity of an assessment is the degree to which it measures what it is supposed to measure. According to Bachman (1990, p.289), “Validity is the most important quality to consider in the development, interpretation, and use of language tests” and it has been considered as a unitary concept with regard to the appropriateness and adequacy of the way people interpret and use test scores.

The issue of test validity of CET has drawn a fair amount of attention from Chinese language testing researchers ever since its launch and development in China. Among many researchers embarking on this question, research studies (Yao, 2015; Cheng, 2016) found that CET possessed high validity. For example, with the reform of CET in 2013, some researchers focused on studying the validity of the newly reformed CET. For example, Xu (2015) studied the content validity of CET-4 reading comprehension in the first test paper reform of 2013 in order to determine the future direction for college English curriculum reform as well as improve college English teaching. The results reported by Xu demonstrated that the design of the newly reformed CET-4 reading comprehension conformed to the standards of CET-4 Test Syllabus. The CET-4 reading comprehension had a high validity on the whole. Hao, Feng, Li and Yang (2015) also investigated the content validity of the CET-4 reading comprehension in 2013 (the first test paper after the reform). The results demonstrated that the newly reformed CET was generally in accordance with the standards of the Test Syllabus. However, the researcher in this study suggests that further improvements need to be made on the content validity of the CET-4 reading comprehension.

Reliability of CET-4

Reliability refers to the consistency of the scores obtained, i.e., consistent scores for each individual from a comparison of one instrument with another or from one set of items with another set (Subong, 2005). Reliability consists of two aspects: internal reliability and external reliability. Internal reliability refers to “the consistency of the results obtained from a piece of research” (Nunan, 1992, p.14). External reliability refers to “the extent to which independent researchers can reproduce a study and obtain results similar to those obtained in the original study” (ibid). Reliability is an indispensable condition for validity, in the sense that “test scores which are not reliable cannot provide a basis for valid interpretation and use” (Bachman, 1990, p. 290).

A review of the studies on the reliability of CET-4 reading comprehension, showed that most studies were conducted before 2013. Only limited studies were found on investigating the reliability of CET-4 reading comprehension. In 2012, more than one sets of tests were administered to candidates in one examination. Cao and Wang (2015) conducted a study to investigate the consistency of the reliability of the CET-4 reading section from each set of test papers administered simultaneously in one examination by contrastive analysis. The Reading sections from three sets of papers in 2014 were selected as materials. Through a contrastive analysis, the researchers reported that the reliability of the CET-4 reading section from each set of paper was generally high and consistent in the light of the College English Test Syllabus. On the contrary, with regard to the Requirements for college English Teaching, the reliability needed to be further improved.

With respect to previous studies concerning the validity and reliability

of CET-4 reading comprehension, the studies which could be found have proved that the validity and reliability of CET-4 reading comprehension were generally high while suggesting that further improvements should be made on CET-4. However, only a limited number of studies in this regard could be found after the reform of CET-4 in 2016. More evidence is needed to back up these claims.

2.3 The CEFR

2.3.1 Origin of the CEFR

In 1991, an intergovernmental symposium entitled “Transparency and Coherence in Language Learning in Europe: Objectives, Evaluation, Certification” was held by the Swiss Federal Authorities in the Swiss municipality of Rüşchlikon. At this symposium, the idea of constructing a common European framework for languages was launched. This framework was mainly used to “promote and facilitate co-operation among educational institutions in different countries, provide a sound basis for the mutual recognition of language qualifications and assist learners, teachers, course designers, examining bodies and educational administrators to situate and co-ordinate their efforts” (CoE, 2001, p.5-6). One year after the symposium, an international working party as well as an authoring group were set up to be responsible for developing this framework. Through a long process of reviewing, studying, analyzing, incorporating and validating, the first version produced of CEFR was examined by 2000 experts in 1996 and the results were announced after the conference in Strasbourg in 1997. After four more years of piloting, the official versions of CEFR in French and English were published in January 2001. Now this framework has been translated into more than 40 languages.

2.3.2 Content of the CEFR

As described by North (2007), the core of the CEFR is a set of common reference levels and a descriptive scheme used to analyze L2 learning objectives, guide the design and development of L2 teaching and learning materials and activities, as well as assessment of L2 outcomes. The descriptive scheme covers two main dimensions: vertical and horizontal. The vertical dimension describes language learners' ability in terms of listening, reading, speaking and writing at an ascending six levels from A1 to C2 with three broad bands as mentioned in chapter one. Communicative language activities in the CEFR include four aspects including reception, production, interaction and mediation. As a consequence, the scales of the language proficiency of these skills including listening, reading, spoken and written production, and spoken and written interaction are clarified. In order to facilitate non-specialist users and also provide teachers and curriculum planners with key points, a global scale (Appendix A) was presented in the form of a set of Common Reference Levels. Apart from a global scale, the Common Reference Levels also comprise a Self-assessment grid (CoE, 2001, p.26-28) and a Rating Scale of Qualitative Aspects of Spoken Language Use (CoE, 2001, p.28-29). The Self-Assessment Grid can assist learners to profile their language abilities and identify their own proficiency level. A Rating Scale of Qualitative Aspects of Spoken Language Use was aimed to assess learners' spoken performance, describe qualitative aspects of spoken language at each level, and provide criteria for range, accuracy, fluency, interaction and coherence. In fact, the suggested 9 levels were developed by the Swiss research project which eventually developed the CEFR levels. These 9 levels are A1, A2, A2+, B1, B1 +, B2, B2+, C1, and C2. 'Plus' levels are given in the scales, in the upper half of the cells labelled A2, B1 and B2 (see Figure 2.1).

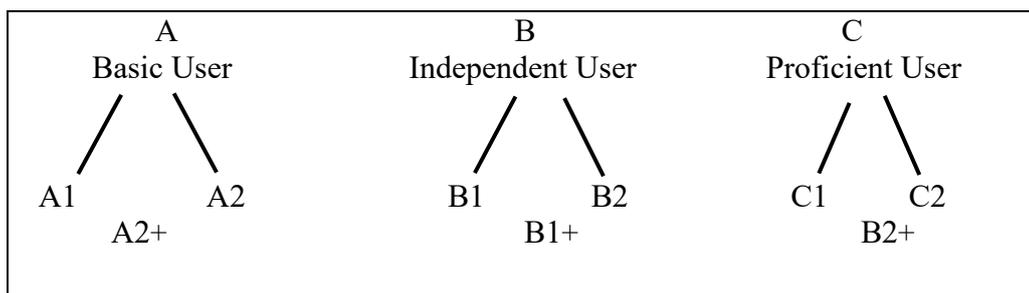


Figure 2.1 9 levels of the CEFR (CoE, 2001)

In accordance with the framework, what a learner is expected to be able to do and how well she or he performs are described at each level by using the positive ‘can do’ statement. As Figueras (2012) stated, even though countries, programs, and classrooms differ in various aspects of language education, it is common to specify what students ‘can do’ and can be observed to do rather than what they cannot do. Jin, et al. (2017) also stated, can-do descriptors describe performances typical of a proficiency level, which are observable, measurable and describable. An example of a ‘can do’ statement is presented below (CoE, 2001, p. 24):

Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

The horizontal dimension of the CEFR refers to the classification of the contexts of language use into different aspects such as domains, situations, constraints, mental contexts, communication themes, purposes, and communicative tasks. It also deals with the learners’ communicative language competences and the strategies acting as a hinge between these competences (linguistic resources) and communicative activities (real ability).

2.3.3 Impact of the CEFR

As a result of the wide international impact and positive comments received from numerous researchers, a great number of countries outside Europe started to adapt the CEFR or even adopt it to fit their language education contexts. For example, in Taiwan, the CEFR was adopted by the MoE as a means to establish a standard of English proficiency in 2005. Following this move, students and English teachers were required to demonstrate their level of English proficiency with reference to the CEFR framework by taking an external English language test. At the same time, major test providers in Taiwan were required to calibrate their tests against the CEFR levels to help users choose a test that they deemed suitable for themselves (Wu, 2012). With the support of the government, a score from an external English test has thus become important and competitive in Taiwan. Consequently, the number of English language test takers dramatically increased. They can choose tests such as IELTS, TOEFL, as well as TOEIC. However, the General English Proficiency Test (GEPT, see www.lttc.ntu.edu.tw) which was developed locally is the most widely used test. The GEPT is a five-level criterion-referenced testing system in which there are four components of assessment including listening, reading, speaking and writing at each level. The five levels of the GEPT are Elementary, Intermediate, Level, Advanced and Superior. After finishing the GEPT-CEFR Calibration Project, the MoE in Taiwan urged English teachers, students and government employees to achieve a pass at a certain level of the GEPT (Jin et al., 2017). For example, English teachers are required pass the GEPT-High Intermediate Level, which is equivalent to CEFR-B2, and government employees are required to achieve a pass at the GEPT-Elementary Level, equivalent to CEFR-A2. Those with a better command of English have a greater chance

for promotion. As Wu and Wu (2007) stated, the introduction of the CEFR may simulate positive development in EFL education in Taiwan even though some problems have arisen.

In Japan, the Japan Language Testing Association (JLTA) tried to adapt the CEFR to the local language education context. The CEFR-J, a Japanese version of the CEFR, was firstly developed in 2008 (www.cefr-j.org). Before the CEFR was developed in Japan, the distribution of Japanese learners' English proficiency was virtually unknown with regard to the CEFR levels. With the adaptation of the CEFR in the Japanese language education context, a survey on the CEFR levels of Japanese learners' English proficiency was conducted. This study revealed that more than 80% of Japanese EFL learners were non/basic users (A1 or A2), less than 20% were independent users (B1 and B2) and almost none were proficient users (C1 and C2) (Negishi, 2012). Thus, some adjustments were made concerning the use of the CEFR especially regarding the lower level of language proficiency in order to make the CEFR-J more suitable for the Japanese language education context. For example, an additional pre-A was added and then A1 was divided into three levels: A1.1, A1.2, and A1.3. Furthermore, A2, B1, and B2 were divided into two levels: A2.1, A2.2, B1.1, B1.2, B2.1, and B2.2. No changes were made to the C levels in the CEFR-J. Therefore, the resultant CEFR-J has a level A with six sub-levels (Pre-A1, A1.1, A1.2, A1.3, A2.1, and A2.2), a level B with four sub-levels (B1.1, B1.2, B2.1 and B2.2), and a level C with two sub-levels (C1 and C2) (Negishi & Tono, 2014). On the other hand, the can-do descriptors are adapted to the Japanese context. The final version of the CEFR-J was published in 2012.

Similarly, the impact of the CEFR in Vietnam was remarkable. In 2008, a project was launched by the Ministry of Education and Training of Vietnam (MOET)

to improve the foreign language proficiency of the Vietnamese. This is a 12-year project called NFL 2020 (The National Foreign Language 2020) aiming at renovating foreign language teaching and learning within the national education system (Socialist Republic of Vietnam, 2008), on the one hand, and particularly to develop their English language capacity to engage in a globalized world, on the other hand. In order to obtain this objective, new approaches were explored by NFL 2020 to make English language teaching and learning in Vietnam more relevant, efficient, and productive, so that most Vietnamese graduate students would be able to communicate in English. One of the approaches to improve English proficiency in Vietnam was the adoption of the CEFR, which was designated as the basis for “developing foreign language curricula, textbooks, teaching plans and assessment criteria at all levels of education to ensure their continuity” (Vietnamese Government, 2008, p. 2). Furthermore, the CEFR was also utilized for teacher training in Vietnam. From 2011-2012, a survey was conducted to evaluate Vietnamese teachers’ English proficiency. The result revealed that 80-90% of primary, secondary and upper secondary school teachers are under-qualified to teach English. A survey also showed that 319 college English teachers (approximately 45%) were under-qualified to teach English. Therefore, teacher proficiency standards were set and teachers were retrained to meet the expected standard which was measured by the CEFR. For example, primary and lower secondary school teachers are required to reach B2 level in the CEFR, while upper secondary school teachers are expected to have language skills at the C1 level. Additionally, the English curriculum was matched with the CEFR and proficiency benchmarks were set for students to improve their English proficiency. According to Hung (2015), the CEFR provides an ideal standard to adopt for foreign language education in Vietnam and English proficiency level is

increasing. Vietnam was classified as moderate in proficiency and ranked 28th in the EF English Proficiency Index (EF EPI 2013) with the adoption of the CEFR.

In addition to the impact of CEFR on the above-mentioned countries, its effect on Thailand has also been profound. In 2014, the CEFR was introduced to the basic education system by the Office of the Basic Education Commission, a department under the Ministry of Education, suggesting that the CEFR in combination with Communicative Language Teaching (CLT) should be employed as its latest policy to improve the standard of English learning and teaching in Thailand's schools. The objective of implementing CEFR in the Thai English learning context aims at setting an achievement benchmark for Thai students, indicating that students graduating from grade 6 are supposed to reach the CEFR level of at least A1, grade 9 of A2, and grade 12 and vocational college of B1 (Wudthayagorn, 2018). In 2015, the Office of the Higher Education Commission, also a department under the Ministry of Education, adopted the CEFR into higher education and every higher education institution was required to assess students' English language proficiency upon graduation with reference to CEFR or other similar standards (Office of the Higher Education Commission, 2015). In Thailand, the CEFR is also used to measure the English language ability of Thai English teachers. In 2015, the Ministry tested all civil servant English teachers in a CEFR-referenced online placement test. Consequently, 94% of the teachers failed to reach the targeted proficiency level of B2 (Franz & Teo, 2018).

2.4 The Manual

2.4.1 Origin of the Manual

As mentioned in 1.1, the CEFR has had a significant impact on testing until now. Some researchers or institutions set about relating their examinations to the CEFR. However, it was still difficult to discuss the academic levels in Thailand with reference to the CEFR levels in the late 1990s. According to North (2007), the majority of claims of a relationship to the CEFR was basically derived from self-declaration, personal opinions or small groups of experts. Therefore, the CoE had to take an active role in assisting examination providers to situate their examinations within the CEFR, and simultaneously to validate the claim of a linkage between language examinations and the CEFR (Figueras, et al., 2005). In response to this need, the Finnish authorities hosted a 3-day seminar in Helsinki in July 2002 entitled 'Relating language examinations to the Common European Framework of Reference for Languages: Learning, Teaching, and Assessment (CEFR)'. Following this meeting, the first version of the Manual was produced by a group of five authors who were supported by a number of consultants drawn from participants at the Helsinki seminar, and then the first draft was published in 2003 in which four methodological procedures (Familiarization, Specification, Standardization, and Empirical validation) were provided for users to develop a linkage between examinations and the CEFR. Based on the pilot version, the final version of the Manual was officially published by the CoE in 2009 in which the original four procedures (Familiarization, Specification, Standardization, and Validation) of linking an examination to the CEFR were changed into five procedures (Familiarization, Specification, Standardization Training/Benchmarking, Standard setting, and Validation).

2.4.2 Content of the Manual

The final version of the Manual was published in 2009, which was a further improvement on the first version in 2003. The content of the Manual is described below and the linking procedures of the present study are based on the final version published in 2009.

The Manual published in 2009 provides five inter-related sets of procedures for users to design a linking scheme in terms of a set of required activities. Nevertheless, projects might be diverse in terms of the extent to which they can be linked and also dependent on the types of assessment (CoE, 2009). Linking generally starts with the first procedure which is Familiarization, and is then followed by Specification, Standardization Training/Benchmarking, Standard setting and Validation. The major aim of Familiarization is to ensure participants have an in-depth knowledge of the CEFR. Specification is about mapping the coverage of the examination on to the categories of the CEFR (CoE, 2009). Standardization Training and Benchmarking aim at facilitating understanding of the CEFR. Standard setting aims to decide on the cut-off scores for test-takers' borderline performances. Validation of tests through evidence establishes the reliability of the tests for the test users.

During the Specification procedure, the Dutch CEFR Grid for listening and reading can be used to analyze the listening and reading examinations to relate them to the CEFR. The 'Dutch CEFR Grid' is also known as the 'CEFR Content Analysis Grid' for Reading/Listening, and is often informally referred to as "The Dutch Grid". This on-line tool allows test developers to analyze the Reading and Listening tests, for the purpose of relating them to the CEFR. The Grid was produced by a working group consisting of the Project coordinator (Charles Alderson) from Lancaster University and

other team members with the financial support of the Dutch Ministry of Education. Later, with further funding from the Dutch Ministry of Education, a computerized version was produced by the team.

See <http://www.lancaster.ac.uk/fss/projects/grid/grid.php> as shown in Figure 2.2.



The Dutch CEFR Grid Reading / Listening

Using the Grid - a brief guide

A Reading or Listening test is made up of one or more Tasks. Each Task is made up of one or more input Texts and one or more test items. Thus Text(s) plus item(s) constitute a Task. A test may include several tasks, but each task should be analysed separately. Therefore, when analysing tests according to the GRID, you need to analyse separately

- each text on which any one task is based,
- each individual item constituting the task,
- and the CEFR level of the task itself.

Each test has a Summary Page which indicates how many Tasks there are, for each Task how many Texts it contains, and how many items there are per Task.

Once you are fully familiarised with the GRID, we strongly recommend that this GRID be completed by analysts working in a team, discussing their analysis as they proceed, and amending it in the light of discussions and the ongoing process. It may be more practical for each individual analyst first to do their own analysis of the test(s) in question, but thereafter research has shown that it is essential that the team of analysts

- meet face to face,
- examine and discuss their individual analyses,
- resolve disagreements through discussion (for which adequate time needs to be ensured - at least half an hour per task, depending how many items and texts there are per task),
- and then enter into the GRID their agreed analyses.

To help you understand this process, we have developed

- a [Sample Grid](#) which you can explore to see how individual test tasks, texts and items have been analysed (also accessible using the buttons on the right),
- and
- a [Training Module](#) which enables you first to analyse one or more test tasks (with associated texts(s) and items) on your own, and then to compare your analysis with that of experts who are fully familiar with the CEFR and with this GRID (also accessible using the buttons on the right).

Enjoy!

Disclaimer: Neither the Dutch CEFR Construct Project Group nor Lancaster University accept any responsibility for the functioning of this Grid and specifically for the security of any data that is input into the Grid. Therefore it is highly recommended that users should print out copies of the Summary Pages for reference and as back-up, in case failures of the system should occur and data is lost

Examples and training

Click the button below to examine a pre-analysed test designed to show you how the system works

[Sample Grid](#)

Click the button below to start a training module. This is a Grid which allows you to analyse a test and then compare your analysis with one made by experts.

[Training Module](#)

Test analysis

Use the menu below to analyse a test which has already been set up in the system (password protected). You will need to have been given an ID and password by the test administrator).

CET 4
CET-4
Chapter 2 the Dare
Charlie Collier
.....

Submit

In the admin area you can view tests that you have set up, add new analysts, modify details such as passwords, and enter new tests for analysis.

[Admin Area](#)

Figure 2.2 The Dutch CEFR Grid

According to the project, a necessary part of the linking process is to characterize the content of the test items and tasks. The Project Team states that this linking process can be facilitated by using the Grid. Details about the Dutch CEFR Grid will be discussed in the next chapter.

It should be noted that analysts or Grid users are supposed to be thoroughly familiar with the CEFR so that the Grid can be used effectively.

2.4.3 Impact of the Manual

As stated in previously, the procedures provided in the Manual have been widely employed by many researchers to relate various tests and examinations to the CEFR. The Manual helps test providers to look through the test specifications and test items and understand thoroughly the test quality, which is completely central to the improvement or reform of tests or examinations in order to reach an international standard. Furthermore, the Manual also offers guidance on relating results reported from examinations or tests to the CEFR and provides evidence of the procedures followed (CoE, 2009). Huang and Jia (2012) contend that the Manual provides test developers and organizations with a scientific and systematic approach for the efficient alignment of tests with the CEFR. Wu and Wu (2007) also describe how the Manual offers a detailed report on the quality of the tests and also provided a useful instrument for the re-examination of language tests. Zou (2016) similarly reports that the procedures provided in the Manual are effective in assessing the content of examinations or tests, setting cut-off scores and confirming the CEFR levels.

2.5 Review of research on CEFR linking projects

2.5.1 Linking projects outside Asia

Research on linking with the CEFR has been conducted by numerous researchers in different countries. For example, Kecher and Eckes (2010) conducted a study to examine the relationship between TestDaF (Test of German as a Foreign Language) and the CEFR levels by linking all four skills (listening, speaking, reading and writing) of the test to the CEFR levels. This study adopted the four methodological procedures (Familiarization, Specification, Standardization, Empirical validation) outlined in the preliminary pilot version of the Manual published in 2003. Staff members of the TestDaF Institute took part in the Familiarization phase. While conducting their 1.5-hour-long familiarization activities, staff encountered some difficulties during the process, for example, lack of time for familiarization with the CEFR and too many scales. Thus, a second familiarization workshop was held again with an extension of the time from 1.5 hours to 2 hours 15 minutes. While verifying the empirical validation, they used the German section of DIALANG (see more at [http://www.celelc.org/projects/Past Projects/DIALANG/index.html](http://www.celelc.org/projects/Past%20Projects/DIALANG/index.html)) as an external criteria measure for the receptive skills data, and assessment from teachers to students was collected to validate the data on productive skills. The results of this study showed that there was a tendency in the reading and listening items towards level B1+ of the CEFR although the deviation represented only half a level below the targeted level of B2.

O'Sullivan (2010) reported a study which aims at establishing empirical evidence of a linkage between Communicator (a comprehensive test of reading, listening and writing skills in English) and CEFR level B2. This was a joint project

conducted by the City & Guilds of London (the examination developers) and the Center for Language Assessment Research (CLARe) located in Roehampton University, London. This study also adopted the four inter-related steps provided in the preliminary pilot version of the Manual. 2 leaders and 9 group participants from City & Guilds took part in Familiarization activities which included sorting the CEFR descriptors, self-assessment of language levels and benchmarking samples of candidates' writing to the CEFR levels. At the Specification stage, the test specification was re-written by using Weir's (2005) validation framework in order to insure the clearly described theory of validity and validation running through the linking project before completing the forms. Specification forms then were firstly completed by 3 participants from City & Guilds and then reviewed by 1 member of the CLARe staff. At the Standardization stage, 4 judges from CLARe and 9 judges from City & Guilds and 24 external judges were involved. A modified Angoff method was utilized for Listening and Reading papers as a standard setting method. Firstly, the judges defined the least competent learners at B2 (both for listening and reading skills), and then they predicted how such individuals would response to each item. The results of this study revealed that this linking project has promoted the systematic reform of Communicator. Furthermore, City & Guilds decided to extend this project to include all examinations in the Spoken ESOL and IESOL suites. However, the researcher of this study found a lack of theoretical framework to connect the specification and validation process for the four procedures provided in the Manual. Therefore, amendments to the linking model are suggested and other validation procedures in order to verify the whole linking process.

Khalifa, French and Salamoura (2010) also conducted a study which aims to sustain the linkage already established between FCE (First Certificate in English) with

the CEFR. In their study, they mainly focused on the first two procedures (Familiarization and Specification) outlined in the Manual. This study also aimed to explore how to familiarize expert panels with the CEFR by using the activities provided in the Manual. A total of 14 participants who had extensive experience in developing and validating tests took part in the project. The Familiarization stage involved three workshops in which pre-, face-to-face and post-activities were carried out respectively. Preparation tasks (a background reading of the CEFR global scale, a selection of B1 to C1 descriptors of listening, reading, speaking and writing) were carried out. The face-to-face workshop started with an introduction focusing on the origins, aims and nature of the CEFR, its relevance for language assessment and its implications for participants as professional language testers working with Cambridge ESOL. A descriptor sorting-activity then proceeded to apply language use descriptors to the CEFR levels. For post-activities, many tasks were conducted to evaluate the effectiveness of the Familiarization and training activities and materials. In the Specification stage, an external consultant who had prior knowledge of the FCE exam and the CEFR was consulted to complete this stage. The activities carried out at the specification phase were: 1) obtaining introductory and background information about the Manual by reading Chapters 1, 2 and 4 of the Manual, 2) consulting all the CEFR scales outlined in the Specification forms and related FCE documents (FCE test specifications & task specifications), 3) completing related forms in the Manual, and 4) providing a written report on the process. The study revealed that FCE should be classified as B2 level across all four skills and the researchers also found that the procedures provided in the Manual and the forms were useful for their linking project. This study provided users with a number of ways of using and complementing the Manual activities to build up a

link as well as maintain a linkage between an examination and the CEFR.

2.5.2 Linking projects in Asia

Apart from numerous linking projects conducted in European contexts, some examinations or tests were also related or linked to the CEFR levels in Asia. For example, Wu and Wu (2007) conducted a study to map the GEPT onto the CEFR in Taiwan. This study adopted three procedures including Familiarization, Specification and Standardization (only judgment session) proposed in the Pilot Manual. The participants comprised 20 experienced professionals in English teaching or testing. In the Familiarization stage, participants carried out a series of activities such as a presentation of the CEFR to the participants, self-assessment of their English proficiency by using the self-assessment grid in the CEFR, sorting individual proficiency statements from a lower level to a higher level and then comparing their results to the CEFR level descriptors. In the Specification stage, the Dutch CEFR Grid was applied to analyze different levels (excepting the Superior Level) of the GEPT Reading Comprehension Tests to reflect and differentiate its difficulty levels in terms of the CEFR levels. Text dimensions including text sources, text types, communication themes and domains were presented in the Dutch CEFR Grid. Three judges participated in the process of mapping for each level of the GEPT Reading Comprehension Tests; a different pair of judges analyzed each level and another person double-checked the results of the analyses for all four levels. A total number of 49 texts and 200 items in the GEPT Reading Comprehension Tests were analyzed using the Dutch CEFR Grid. The experience of this mapping project revealed that the Reading Comprehension Tests correspond in general to the CEFR levels. The results of the content analysis showed that the GEPT reading texts become more and more complex as the GEPT level ascends.

The study also found that expeditious reading and non-verbal task input are not addressed in the CEFR. The researchers commented on a lack of precision and clear guidelines to make comparisons of constructs and difficulty levels between different testing systems. Nevertheless, researchers of this study also pointed out that the Specification outlined in the Manual has facilitated detailed reports on the quality of tests.

Although linking projects are relatively scarce in Mainland China, some attempts have also made by researchers. For instance, following the procedures provided in the Manual, Liu (2012) firstly aligned TEM-4 to the CEFR through a method of comparative analysis in order to find out the particular CEFR level that TEM-4 is related to. Materials in this study included the Teaching Syllabus for English Majors, the Syllabus for TEM-4 and five administered papers from 2007 to 2011. All of these documents were analyzed and compared using all the illustrative scales in the CEFR. This present study started with the researcher herself following the Familiarization procedure by reading the CEFR, the Manual and the some linking projects. Later, Familiarization activities suggested in the Manual were undertaken by the researcher. In the Specification stage, forms A1-A8 in the Manual were used to describe and analyze the quality of TEM-4 based on the Teaching Syllabus for English Majors and the Syllabus for TEM-4 and to establish the TEM-4 level in relation to the CEFR. Then, forms A9-22 were utilized to analyze and compare the Teaching Syllabus for English Majors, the Syllabus for TEM-4 and the test papers from 2007 to 2011 with the help of the illustrated scales in the CEFR. Through this procedure, the researcher found the particular CEFR level that TEM-4 is related to. In order to calculate the cut-off score of TEM-4, 60 sophomores majoring in English were organized to take TEM-4 and TOEFL,

and each paper was marked by two English teachers. The Standardization procedure was carried out using by comparing 60 sophomores' TEM-4 scores with their usual performances and the cut-off scores were calculated. The students were divided into two groups in terms of their usual performances, and then the frequency distribution of their TEM-4 scores was graphically displayed. The cut-off score was then determined out by calculating the average score of each group and then calculating the value midway between these two averages. Finally, in the validation stage, the reliability and validity of the tests and the correlation of the TEM-4 and TOEFL scores were calculated, and then the distribution of the two tests were compared in order to confirm the results of the linking with the help of the Common Reference Level. The research revealed that the language proficiency level of TEM-4 is B1+ in the CEFR, which means, according to Liu (2012), that the test-takers who pass TEM-4 have an English listening, reading and writing competence of level B1+ in the CEFR. The researcher also spotted two problems in TEM-4. One of the problems reported by the researcher is that TEM-4 lacks transparency in the aspects of marking, reporting results, data analysis, and grading and test development. The other problem is that the "requirements of the proficiency level for each language skill in TEM-4 are uneven" (Liu, 2012, p.59).

In 2016, Zou conducted a study aiming at aligning the NMET with the CEFR by using the 2015 test paper in Jiangxi province, China, namely the National English Test Paper 1. The reading and listening sections from the test paper were chosen to establish a link between the NMET and the CEFR. The participants of the study consisted of 10 panelists (1 leader and 9 members) and 5 external experts. The panelists of 10 members included 3 college English teachers, 3 high school English teachers and 4 M.A students majoring in English, and the external judges were experienced

professionals in English language teaching and assessment. Similarly, the study followed the four procedures provided in the Manual. It started with the first Familiarization procedure outlined in the Manual, and data was collected and analyzed through SPSS in terms of participants' intra-reliability and inter-reliability as well as internal consistency on a common understanding of the CEFR descriptors. In the specification stage, the participants analyzed the content and quality of the 2015 National English Test Paper I with the help of the question forms provided in the Manual to establish a first link between NMET and CEFR. In the Standardization stage, a re-familiarization and training workshop and standard-setting workshop were organized to corroborate the claims made in the Specification phase. The standard-setting method used in this study was a Basket Method to calculate the cut-off scores. In order to verify the validity of this study at the Empirical validation phase, 127 grade three high school students from a key senior high school in Mainland China were selected to take part in NMET and TOEFL-iBT tests which were previously aligned with the CEFR levels. The test results of the listening skills were compared between NMET and TOEFL-iBT test. The results in this study showed that the listening tasks of receptive skills in the NMET can be related to the A2 level of the CEFR and the reading tasks to the B1 level of the CEFR, which means, as the researcher stated, that "the test-takers who have passed the NMET have an English listening competence of level A2 and a reading competence at B1 level of the CEFR" (Zou, 2016, p.41). The researcher also believes that the Manual needs to be further improved in terms of the Familiarization phase. She suggests that it would be much better to turn the Familiarization stage in the Manual into a cyclical process through a linking project since the more familiarization activities the participants carry out, the more significant

and reliable the results obtained.

2.5.3 Linking project of CET with the CEFR

As described in chapter one, the only linking study was conducted by Huang and Jia in 2012 to investigate the feasibility of linking CET to the CEFR in Mainland China. In their study on College English Curriculum Requirements in the year 2007, texts and test items in both the listening and reading sections of CET-4 and CET-6 from 2006 to 2010 (20 sets in total) with 140 texts in the listening sections, 30 in the reading sections, and 250 items in total were analyzed based on the CEFR descriptors and the Dutch CEFR Grid. The research procedures included two steps: 1) the linkage of College English Curriculum Requirements with the CEFR; and 2) the linkage of CET-4 and CET-6 with the CEFR. In Huang and Jia's study, the two researchers, also as the judges, analyzed the texts and items of CET-4 and CET-6 based on the CEFR descriptors and the Dutch CEFR Grid. They did not include the first procedure of Familiarization, but their study revealed that the content of both CET-4 and CET-6 is basically consistent with the CEFR and that CET can be linked to the CEFR. The gaps found in Huang and Jia's study have already been detailed in chapter one. Thus, in order to bridge the gaps found in this previous study, the present study aims to link CET-4 Reading Comprehension Tests to the CEFR levels by following the first two procedures provided in the Manual. The Dutch CEFR Grid was used to analyze the content of CET-4 Reading Comprehension Tests in the Specification procedure. The CEFR descriptor scales, test specification of CET-4 and copies of the last three administered CET-4 Reading Comprehension Tests from December 2016 to December 2017 will serve as materials in this study. Texts and items in the reading section of CET-4 will be analyzed by using the CEFR Grid since some users found the Grid useful in helping them to analyze the content and coverage

of some examinations and tests. A total of 3 participants took part in this study. Firstly, the Familiarization activities were carried out step by step to train all the participants to have an adequate understanding of the CEFR scales and descriptors. Then, the Specification procedure was carried out. The participants were asked to re-familiarize themselves with the CEFR by carrying out the Familiarization activities again. Next, a presentation concerning the use of the Dutch CEFR Grid and a pre-analyzed test sample were given to the participants. Later, participants were required to analyze the content of the CET-4 reading comprehension texts by using the Dutch CEFR Grid. Lastly, a discussion concerning the outcome of the linking activity was undertaken by the participants in order to reach a consensus.

2.6 Summary of the chapter

The origin, the content and the impact of the CEFR and the Manual were introduced at the beginning of this Chapter. Then some previous studies on linking an examination or test to the CEFR both outside and inside Asia were discussed. Lastly, related studies concerning the validity, reliability of CET and a linking project between the CET and the CEFR in China were also described. The next chapter discusses how the present study was conducted it focuses on the design of the research methodology.

CHAPTER 3

METHODOLOGY

The aim of this chapter is to describe the main research procedures of the study. After outlining the research design, the chapter moves on to detailing the research questions, procedures, materials, research context and participants as well as instruments of the present study. Finally, the data collection and data analysis are discussed.

3.1 Research design

The present study employed a mixed design of quantitative and qualitative methods. The selection of research methods was mainly guided and determined by the research questions which aimed at analyzing CET-4 Reading Comprehension Tests in order to find the levels of the texts and test items in relation to the CEFR. A content analysis technique and descriptive statistics were employed to analyze the data and finally the overall picture of the findings is presented. In addition, since the Familiarization training was crucial in order for the participants to generate the most accurate level of CET-4 Reading Comprehension Tests, their perceptions on all of the activities were gathered through an in-depth interview and an open-ended questionnaire. The collected data will provide some evidence concerning the effectiveness of the training which might affect the final results of the linking activities.

One of the merits of using a mixed method is that it provides a range of techniques

to collect and analyze the data. Though they were obtained from different techniques, they were complementary. The inclusion of quantitative data could make up for any deficiencies in the qualitative data, and the use of the qualitative data in turn enabled the investigators to explore the underlying processes which emerged from the quantitative data (Onwuegbuzie & Leech, 2004). Therefore, it can be concluded that the convergence and corroboration of findings both from the quantitative and qualitative data provides strong evidence for the participants' feedback on the Familiarization and Specification workshops as well as the final results of the linking process.

3.2 Research questions

As explained in Chapter 1, the aim of this study was to explore the CEFR levels for the CET-4 Reading Comprehension Tests and text items and also the major characteristics of the texts and test items. In order to realize these aims, two precise research questions were formulated. They are as follows:

1. Which levels of CEFR were CET-4 reading comprehension texts situated in and what were the reading texts' major characteristics?
2. Which levels of CEFR were CET-4 reading comprehension test items situated in and what were the reading items' major characteristics?

3.3 Research Procedures

As addressed in 3.1, this study aimed at linking CET-4 Reading Comprehension Tests to the CEFR by using the first two steps provided in the CEFR Manual, namely, the Familiarization and Specification procedures. In order to show the overall picture

of the whole linking project, the procedures of the study are presented in Table 3.1 with details.

Table 3.1 Linking procedures

The Whole Linking Process	
Phase 1 Preparatory work	Sending a preparatory package to the participants for self-study
Phase 2 Familiarization Workshop	1) Introductory Activities a) Presentation of the CEFR and the workshop b) Descriptor-sorting activity 1 (CEFR levels for interaction & production) c) Self-assessment of participants English level with self-assessment grid 2) Qualitative analysis of the CEFR scales d) Descriptor-sorting activity 2)scales of reading(e (Reconstructing self-assessment grid 3) Preparing for rating f) Descriptor-sorting activity 3 (CEFR level for Reception) 4) Conclusion
Phase 3 Specification Workshop	1) Re-familiarization 2) Introduction to the Dutch CEFR Grid 3) Analyzing the content of CET-4 Reading Comprehension Tests with the Dutch CEFR Grid 4) Discussion

3.3.1 Familiarization

The importance of familiarization with the CEFR was emphasized in numerous studies (Wu & Wu, 2007; Kecher & Eckes, 2010; Khalifa, French & Salamoura, 2010; O'Sullivan, 2010; Liu, 2012; Wang, 2015; Zou, 2016) and in several places such as on pages 7, 9, 10, and 17 in the CEFR Manual (CoE, 2009). The Familiarization procedure served as a fundamental and logical pre-requisite stage for effective linking. Without an in-depth understanding of the CEFR, it would be exceptionally difficult to make reliable decisions on the CEFR levels of a test during the Specification phase. This study thus adopted the activities of the Familiarization procedure from the Manual to train the participants to familiarize themselves with the

CEFR. The Familiarization phase included two steps: preparatory work and the Familiarization workshop. Details for each step are described below.

Preparatory work before the Familiarization workshop

As outlined in Table 3.1, the linking project started with preparatory work by sending a preparatory package to participants for self-study by e-mail on the first week. This was to ensure that all the participants had enough time to understand the CEFR levels so as to make the Familiarization workshop to be held later as successful as possible (CoE, 2009). The preparatory package included 1) a simplified version of the salient features of the CEFR levels for interaction and production (Appendix B), 2) a simplified version of the salient features of the CEFR levels for reception (Appendix C), 3) a self-assessment grid for reading (Appendix D), and 4) the CEFR scales for reading (e.g. overall reading comprehension, reading for information and argument, and reading for orientation), which are listed in Appendix E. Lastly, a written information sheet (Appendix F) detailing the nature of the linking project, the estimated length of time and participants' involvement were also included in the "pre-task package". Table 3.2 lists all the documents and materials in the preparatory package sent to the participants.

Table 3.2 Preparatory package for self-study

Preparatory Package for Self-study
1) A simplified version of the salient features of the CEFR levels for interaction & production 2) A simplified version of the salient features of the CEFR levels for reception 3) Self-assessment grid for reading 4) Scales for reading <ol style="list-style-type: none"> a) Overall reading comprehension b) Reading for information and argument c) Reading for orientation 5) Information sheet

Familiarization workshop

The Familiarization workshop was designed to train the participants in a sufficient understanding of the CEFR by proceeding with a series of activities proposed in the Manual. As a result of the workshop, the participants would be competent in analyzing CET-4 Tests by using the Dutch CEFR Grid at the Specification stage. The estimated time spent on each activity at the Familiarization workshop is shown in Table 3.3.

Table 3.3 Estimated time for Familiarization activities

Activities	Time
1) Introductory activity and discussion	(30 mins)
a) A presentation of CEFR and the Familiarization workshop by the trainer assistant b) Descriptor-sorting activity 1 (the CEFR levels for interaction and production) & c) Self-assessment of participants' English reading proficiency by using self-assessment grid	(45 mins)
2) Qualitative activity including group work	(45 mins)
d) Descriptor-sorting activity 2 (scales related to reading) e) Reconstructing self-assessment grid	(45 mins)
3) Preparation for rating	(45 mins)
f) Descriptor-sorting activity 3 (the CEFR levels for reception)	(45 mins)
4) Conclusion	(15 mins)

1) Introductory activities

The introductory activities mainly included a presentation of the CEFR and the workshop, descriptor-sorting activity 1 and activity for self-assessment of the participants' English reading proficiency.

a) Presentation of the CEFR and the workshop

At the first-stage of the Familiarization workshop, a presentation of the CEFR and the workshop was given by the trainer assistant which aimed at equipping the participants with general information of the nature, scope, content and the background of the CEFR as well as the objectives and contents of the Familiarization workshop. This phase took around 30 minutes.

b) Descriptor-sorting activity 1

After the presentation was completed, activities and discussions were organized. During this period of time, the participants were provided with a prepared sheet which contains 6 randomly scrambled descriptors coming from the simplified version of the salient features of the CEFR levels for interaction and production by the trainer assistant. Then, the participants were asked to assign the descriptors to their corresponding CEFR levels (A1-C2) without consulting the original version. While preparing these scrambled descriptors, level references were eliminated to ensure that no clue could be found among the descriptors for the sorting tasks by the trainer assistant. After the participants completed their sorting activity, the full version was distributed to them to compare and to check their results. Then, they were required to write down the number of descriptors assigned correctly to their corresponding levels on a judgment sheet (Appendix G). Finally, the participants shared their views on the salient features of each CEFR levels based on their self-study and the sorting activity.

c) Self-assessment of participants' reading ability

After completing descriptor-sorting activity 1, the participants were asked to assess their own English reading proficiency by using the CEFR self-assessment grid for reading and then discussing it with one another. This activity aimed at familiarizing the participants with the CEFR levels for reading.

2) Qualitative analysis of the CEFR

At this stage, in-depth work and discussion of the CEFR levels with respect to the scales and descriptors for reading comprehension were conducted. The activities forced the participants to consider the descriptors in isolation as independent criteria.

d) Descriptor-sorting activity 2

First of all, each participant was provided with envelopes containing some scales cut into constituent descriptors by the trainer assistant. These scales included overall reading comprehension, reading for information and argument, and reading for orientation. Each envelope contained one scale. Each participant thus got 3 envelopes. Next, they were required to sort the descriptors into levels one by one. After they completed sorting all of scales, they checked their answers by reference to the original version. Then, they were supposed to immediately note down the number of descriptors assigned correctly to their corresponding levels on the same judgment sheet again one by one until the last scale was completed. After they finished sorting and reporting for each sorting activity, the trainer guided them in a discussion to reach a consensus.

e) Reconstructing the reading part of the CEFR self-assessment grid

The last step of this phase was reconstructing the reading part of the

CEFR self-assessment grid. The participants were individually provided with a piece of paper which had the CEFR levels labelled on the left side but with empty cells on the right (Appendix H). Then, they were expected to reconstruct and place the descriptors into the correct cells. After finishing this step, the original version of the reading part of the Self-assessment grid was provided for them to check and revise their work.

3) Preparation for rating

The activity for this phrase used the simplified version of the salient characteristics of the CEFR levels for reception to train the participants to prepare themselves for analyzing the texts and items in CET-4 Tests. Details of the CEFR levels for reception are described in 2) in 3.4.

f) Descriptor-sorting activity 3

The activity for this phase was called descriptor-sorting activity (3). Each participant was provided with a sheet containing 6 scrambled descriptors from the simplified version of the salient characteristics of CEFR levels for reception, and then they were required to sort these descriptors into the CEFR levels. The activity was conducted in exactly the same way as described in the descriptor sorting activity 1. After finishing their sorting, the participants set down the number of descriptors they had assigned correctly to the corresponding levels on the judgment sheet.

4) Conclusion

After finishing all the activities above, the last step for this procedure was making a conclusion to the whole Familiarization workshop. This phase offered all the participants an opportunity to assess the work that they had done. It was hoped that the knowledge and understanding of the CEFR they had acquired would facilitate their work of the linking project later.

3.3.2 Specification

As described in chapter 2, the Specification procedure dealt with the analysis of the content of an examination or test in order to connect them to the CEFR levels from the point of view of coverage (CoE, 2009). This was a pivotal procedure for this study since the end product was concerned with accurate linking to the CEFR scale based on the specification. In this phase, the contents of the CET-4 Tests were analyzed by using the Dutch CEFR Grid (see more about the Grid in 3.6.2). This part consisted of four steps: 1) assuring adequate familiarization with the CEFR (re-familiarization); 2) an introduction and examples of the Dutch CEFR Grid; 3) analyzing the contents of the CET-4 Tests with reference to the Dutch CEFR Grid; and 4) discussing the analysis results.

Specification workshop

The Specification workshop and discussion was on weeks 4-5. After finishing the activities of re-familiarization which were the same as the Familiarization for the first procedure, an introduction to the Dutch CEFR Grid and analysis of a sample text were conducted in the morning. In the afternoon, participants analyzed the contents of CET-4 Tests (one set of paper) with the Grid. The estimated times for the Specification workshop are shown in Table 3.4.

Table 3.4 Estimated time for Specification workshop

Date	Day	Activities	Time
Sep 23 rd , 2018	morning	re-familiarization & introduction of the Grid (including analyzing one sample text)	3 hours
	afternoon	analyzing one test paper	3 hours
Sep 24 th , 2018	morning	analyzing one test paper	3 hours
	afternoon	analyzing one test paper	3 hours
Sep 30 th , 2018	morning	discussing the results	1 hours

1) Assuring adequate familiarization with the CEFR (re-familiarization)

To ensure that all the participants were thoroughly familiar with the CEFR, it was necessary to ask them to redo the Familiarization activity as described in 3.3.1, which was crucially important for the subsequent linking process. This activity was conducted in week 4 on the first day and it only took 30 minutes.

2) An introduction to the Dutch CEFR Grid

After completing the familiarization work, the project moved on to present the introduction of the Dutch CEFR Grid and link a sample test of reading extracted from CET-4 Tests to the CEFR guided by the trainer.

Firstly, the paper-based Dutch CEFR Grid (Appendix I) was distributed to each participant and then the trainer assistant presented the Grid covering its content, purposes, functions, and usage by means of a PowerPoint presentation to the participants. Then, the trainer guided the participants to scrutinize and discuss each key point on the Dutch CEFR Grid. Later, the participants were provided with a sample text and a judgment sheet to link the sample test with the CEFR with the guidance and help of the trainer. After all the members had completed their analysis of the sample test, a discussion session was held until a final consensus was reached. All these activities took around 2.5 hours.

3) Analyzing the content of CET-4 Tests by using the Dutch CEFR Grid

This phase consisted of an analysis of the content of CET-4 Tests and how to determine the appropriate CEFR level. The participants were provided with the judgment sheets (Appendix J) based on the Dutch CEFR Grid, the test specification of CET-4 (Appendix K), and the CET-4 Reading Comprehension Tests, respectively (Appendix L). Then, they analyzed one set test paper of CET-4 Reading Comprehension

Tests (4 texts and 30 test items) with the analysis results reported on the judgment sheet. The analysis procedure took place in the afternoon of the Specification workshop. On the second day, another two sets of test papers were analyzed by the participants (one set in the morning and one in the afternoon). The analysis of each set of the test papers lasted about 3 hours.

It should be mentioned that the analysis work was conducted individually. One paper-based Dutch CEFR Grid, three judgment sheets (one judgment sheet for one set of test paper, test specification of CET-4, and CET-4 Reading Comprehension Tests (three sets altogether) were given to each participant. After they finished the analysis, the trainer assistant collected all the data.

4) Discussing the analysis results

The last step in the Specification procedure was to discuss the results. This activity was conducted on week 5. After the results of the analysis were computerized, the trainer assistant emailed all the participants to inform them of the results. The discussion session mainly focused on those results which could not be decided because the answers given by each participant were totally different.

3.4 Research Context and Participants

This study was conducted in a university located in the Northeast of Thailand. A total of three teachers participated who were selected by using purposive sampling methods. In addition, an expert in language testing who had trained the three teachers on the CEFR also took part in the study. An assistant trainer also helped to ensure the successful and effective implementation of all the activities and discussions in the workshops.

(1) The 3 participants

Since the study dealt with the analysis of the CET test, it was considered that the participants who possessed relatively extensive teaching experience in universities and an adequate understanding of the CET-4 tests, would be able to do the linking project better than those who did not have any knowledge or experience of CET-4. Though they were not familiar with CEFR, the training activities of the Familiarization and Specification workshops were designed to help them. Therefore, the participants of this study were chosen based on two criteria :

- a) The participants must have at least 5-years' teaching experience at university level in Mainland China,
- b) The participants must be familiar with the CET-4 tests.

Details of the participants of this study are given below:

1. Assoc. Prof. Dr. Lu Huashan, a lecturer at Qingdao Agricultural University. He has been teaching English for 15 years.
2. Dr. Yimin Zhang, a lecturer at Sichuan Agricultural University. She has been teaching English for 7 years.
3. Dr. Daping Wu, a lecturer at Kunming University of Science and Technology. She has been teaching English for 15 years.

These volunteer participants attended the CEFR Familiarization workshop for 3 days and completed all of the recommended activities successfully.

(2) The trainer

For this study, the role of the trainer was extremely critical for the linking project since she would have to conduct the implementation of each procedure, train the participants to be familiar with the CEFR and guide the linking project by using the

sample test. The following qualifications were required:

- a) S/he should understand the CEFR thoroughly.
- b) S/he must have had some experience related to linking tests to the CEFR or organizing workshops about developing a CEFR-based English proficiency test.

These experiences should allow him/her to have a clear idea of exactly what should be carried out in each phase.

- c) S/he must be a professional in the field of testing since the linking project dealt with the test.

For this study, it was a great opportunity to have Asst. Prof. Dr. Kornwipa Poonpol, who is well known nationally and internationally in the field of testing and CEFR linking studies to serve as the trainer for the CEFR Familiarization and linking phase. She was responsible for

- a) training the participants to be familiar with the CEFR
- b) controlling the exact time allocation for each activity.
- c) organizing and guiding the linking project
- d) stimulating discussion among the participants while conducting the activities.

(3) The trainer assistant

In this study, the researcher took part as a training assistant who helped the trainer prepare and organize the linking project. The trainer assistant was responsible for:

- a) preparing all the related materials and documents which were used for the linking project.
- b) sending the preparatory package as shown in Figure 3.2 to the participants for

self-study in the first week.

c) preparing the PowerPoint presentation to introduce the CEFR, the Familiarization workshop, and the Dutch CEFR Grid.

d) checking and making sure that the room as well as the facilities and equipment were available for group work.

e) distributing the necessary materials and documents used in each stage.

f) collecting all the data from the participants both at the Familiarization and Specification Workshops.

g) preparing food and beverage for all participants.

The participants and the timetable for the whole process are shown in Table 3.5.

Table 3.5 Timetable and participants in the whole process

Stages	Time	Participants
Preparatory work	Sep 2 nd -15 th , 2018	participants (3)
Familiarization workshop	Sep 16 th , 2018	trainer (1) trainer assistant (1) participants (3)
Specification workshop	Sep 22 nd -30 th , 2018	trainer (1) trainer assistant (1) participants (3)

3.5 Materials

As described in 3.2, this study aimed at linking CET-4 Reading Comprehension Tests to the CEFR through a Familiarization workshop (training) and a Specification workshop (linking) by using the Dutch CEFR Grid. Thus, the materials used throughout the training and linking procedures in this study included 1) the CEFR scales, 2) the CEFR levels for interaction and production, and the CEFR levels for reception, 3) the test specifications of CET-4, and 4) copies of three sets of CET-4 Tests from December

2016 to December 2017.

1) The CEFR descriptor scales

The CEFR descriptor scales included a self-assessment grid for reading and reading scales (overall reading comprehension, reading for information and argument, and reading for orientation).

A self-assessment grid for reading helped the participants understand the English proficiency in terms of reading at certain levels by asking them to do a self-assessment. Scales for reading enabled users to further understand reading proficiency from different aspects rather than just having an overall picture of the proficiency at each level.

2) The CEFR levels for production and interaction, and the CEFR levels for reception

The CEFR levels for production and interaction and for reception presented the salient features of the CEFR levels of each type. The CEFR levels for reception mainly dealt with reception skills (listening and reading). Information at each level such as source, setting, and restrictions was presented. For example, the setting for A1 level was the most common situation, and the source for this level should be very short texts with visual support, a single phrase at a time (CoE, 2009).

3) Test specification of CET-4

The test specifications of CET-4 described in detail the test purpose, test domain, test takers, test method, test format and details for each sub-test including reading, writing, listening and translation. The sub-test covered many aspects such as test domain, operations, test type, topics, time duration, test length and item type, which helped the participants to analyze the CET-4 Reading Comprehension texts using the

Dutch CEFR Grid.

4) Copies of three sets of CET-4 Tests

The reasons for selecting the last three administered CET-4 tests were that they were the latest form of CET-4 implemented in 2016, which contributed to some changes of the test format of the Reading Comprehension Tests. Thus, It was reasonable to use the newly reformed CET-4 since it presents an up-to-date version of the CET-4 that college students take at present. On the other hand, the CEFR Manual also recommended the use of the last three administered tests to conduct a linking project. Therefore, it is evident that the results would be able to exemplify the test constructs, and at the same time reflect and differentiate the level of the CET-4 Reading Comprehension Tests in relation to the CEFR.

The CET-4 administered each time contained three different sets of papers. One set of papers was randomly selected out of three as sample material to be analyzed. The sample test papers of CET-4 included sets from December 2016, June 2017 and December 2017.

In each CET-4 Reading Comprehension Test set, there were 4 texts and 30 items. Thus, each participant analyzed 12 texts and 90 items as listed in Table 3.6.

Table 3.6 Texts and items of CET-4 Reading Comprehension Tests

CET-4 Reading Comprehension Tests				
Time for administration	Number of test sets and reading texts	Number of test items	Number of selected text sets	Number of items
Dec. 2016	3(4)	90	1	30
June 2017	3(4)	90	1	30
Dec.2017	3(4)	90	1	30
Total number of texts/items	12	90	3	90

3.6 Instruments

In order to achieve the objectives of the study, four instruments were used to collect the data: 1) descriptor-sorting activities, 2) workshop feedback questionnaires, 3) semi-structured interviews, and 4) the Dutch CEFR Grid with judgment sheets.

3.6.1 Descriptor-sorting activities

The descriptor-sorting activities both at the Familiarization workshop and the re-familiarization phase of the Specification workshop were utilized to train all the participants involved in the linking project to be familiar with the CEFR. These activities were also used to measure participants' understanding of the CEFR levels. There were three descriptor-sorting activities that are described in 3.3.1 b), d), and f).

3.6.2 The Dutch CEFR Grid

As mentioned above, the Dutch CEFR Grid for reading/listening is an on-line tool. However, this study used a paper-based version for two reasons. First, the online version of Dutch CEFR Grid required an efficient network and accidents like network jam or breakdown were not acceptable since it would affect the progress of the analysis. Second, compared with analyzing online, the paper-based approach was more convenient for the participants since they did not need a computer. Furthermore, it was easy to keep the data after analyzing the texts and test items of the CET-4 Tests for future reference. The paper-based analysis, moreover, offered participants the opportunity to discuss any disputes about the texts and test items.

While analyzing a test for reading, information about each task, texts and test items were entered into the Grid by specifying their characteristics from a set of options derived directly or indirectly from the CEFR. The data collected for texts included two aspects: input texts and items (question setting). The characteristics of the input texts

contained different dimensions such as text source, authenticity, discourse type, domain, topic, nature of content, text length, vocabulary, grammar, and the estimated text levels. The characteristics of the items referred to item type, operations, and the estimated level of the items.

It should be noted that three of the characteristics in the Dutch CEFR Grid were modified for this research study in order to better analyze the CET-4 Tests and to avoid subjectivity on the part of the participants. The first amendment concerned vocabulary. The original version of the Dutch CEFR Grid covered four categories: only frequently vocabulary, most frequently vocabulary, rather extended and extended as shown in Table 3.7.

Table 3.7 Vocabulary of the original Dutch CEFR Grid

8. Vocabulary	<input type="checkbox"/> Only frequently vocabulary <input type="checkbox"/> Mostly frequently vocabulary <input type="checkbox"/> Rather extended <input type="checkbox"/> extended
----------------------	---

Nevertheless, there were not any documents or criteria to classify vocabulary into these four categories. Analyzing the vocabulary of the CET-4 reading comprehension texts therefore became rather subjective with reference to the Dutch CEFR Grid. Owing to the need to analyze the CEFR level of texts and test items, finding the CEFR levels of vocabulary was necessary to provide more information for the participants to judge the level of the CET-4 reading comprehension texts and test items. The tool for classifying vocabulary into the CEFR's six levels was available online (see more details about the classification of vocabulary on <http://www.englishprofile.org/wordlists/evp>).

The innovative tools of English Vocabulary Profile Online together with English Grammar Profile Online were employed to analyze the grammar part in this study. They provided searchable databases that gave free access to the research findings on what English vocabulary and grammar were suitable for teaching at each CEFR level. This work was carried out as part of a ground-breaking collaborative project supported by the Council of Europe. It collected data from learners all over the world to inform the research. The research was led by two departments of the University of Cambridge, UK: Cambridge University Press and Cambridge English Language Assessment.

The grammar part was analyzed by using the English Grammar Profile Online which was the second characteristic amended in the present study. There were four categories: only simple structures, mainly simple structures, limited range of complex structures, and wide range of complex structures as provided on the Dutch CEFR Grid. The grammar structures are shown in Table 3.8.

Table 3.8 Grammar of the original Dutch CEFR Grid

9. Grammar	<input type="checkbox"/> Only simple structures <input type="checkbox"/> Mainly simple structures <input type="checkbox"/> Limited range of complex structures <input type="checkbox"/> Wide range of complex structures
-------------------	---

Similarly to vocabulary, due to the lack of documents or criterion to classify grammar structures into these four categories, grammar aspects of CET-4 reading comprehension texts were analyzed according to the CEFR level to avoid subjectivity on the part of the participants. CEFR's classifications of grammar are provided on

https://www.examenglish.com/CEFR/cefr_grammar.htm

The last part to be modified in the Dutch CEFR Grid was the operations of

items. The original version of the Dutch CEFR Grid in terms of operations is shown in Table 3.9.

Table 3.9 Operations of the original Dutch CEFR Grid

12. Operations	<input type="checkbox"/> Recognize	<input type="checkbox"/> Main idea/gist	<input type="checkbox"/> from explicit information
	<input type="checkbox"/> Make inferences	<input type="checkbox"/> Detail	<input type="checkbox"/> from implicit information
	<input type="checkbox"/> Evaluate	<input type="checkbox"/> Opinion	
		<input type="checkbox"/> Speaker's/Writer's attitude/mood	
		<input type="checkbox"/> Conclusion	
		<input type="checkbox"/> Communicative purpose	
		<input type="checkbox"/> Text structure/connections between parts	

However, based on the test specification of the CET-4 Reading Comprehension test, the researcher modified the content of the operations to make it fit with the context of the CET-4 reading comprehension items as well as to assist the participants in their judgements. The modified version of the part of operations is shown in Table 3.10.

Table 3.10 Operations of revised Dutch CEFR Grid

12. Operations	<input type="checkbox"/> Understand the detail
	<input type="checkbox"/> Infer speaker's/writer's attitude/mood/opinion
	<input type="checkbox"/> Generalize and draw the conclusion
	<input checked="" type="checkbox"/> Detect the topic and main idea/gist
	<input type="checkbox"/> Deduce the meaning of an unknown word from context
	<input type="checkbox"/> Identify the needed word from given options according to context
	<input type="checkbox"/> Distinguish the different forms of the same idea

3.6.3 Questionnaire

In a review of the related literature about the instruments used to investigate the effectiveness of the Familiarization procedure in linking projects, Salamoura (2011) used the Workshop Feedback Questionnaire in his project. Thus, the present study takes the questionnaire written by Salamoura as a reference and then the researcher revised some parts. It is an open-ended questionnaire with 10 questions. It consists of two sections:

section one for the Familiarization workshop and section two for the Specification workshop (Appendix M). Each participant responded by writing down their comments or notes in an exploratory investigation designed to elicit descriptive information with the aim of obtaining the participants' perceptions on both workshops and to detect the problems as well as the difficulties the participants encountered.

The reason for choosing open-ended questions was that it allowed the respondents to include more information such as feelings, attitudes and understanding of the subject, which allowed the researchers to better access the respondents' true feelings on an issue.

Since the questionnaire was developed by the researcher herself, it required validation. Norland (1990) noted that validity was the amount of systematic or inherent error in measurement. Validity is established by using a board of specialists and a field test. There are different types of validity (content, construct, criterion, and face etc.). The type of validity selected mainly depends on the objectives of the study. As for the present study, construct validity was conducted by using the index of item-objective congruence (IOC) developed by Rovinelli and Hambleton (1977).

Three experts in English language teaching were invited to validate the language and content in the questionnaire. They rated each item and checked the evaluation form by using IOC as a validation method for the relevance of the content and its objectives. Finally, the questionnaire was revised and improved on the basis of these experts' comments and suggestions.

The evaluation form used a 3-point scale (1 = relevant, 0 = uncertain, and 1 = irrelevant). The questionnaire items were refined and improved until the results of the IOC analysis showed that they were valid.

3.6.4 Semi-structured interview

The semi-structured interview refers to a simple conversation between the interviewer and interviewees using guiding questions. According to Numan (1992), interviews can be categorized into a structured interview, a semi-structured interview, and an unstructured interview. The reason for using a semi-structured interview in this study was that it is suitable for exploring attitudes, values, beliefs, and views so that a deeper and more detailed understanding of research questions could be explored (Fylan, 2005; Van Teijlingen, 2014). By using this tool, the problems and difficulties the participants encountered while carrying out the familiarization activities and analyzing the content of CET-4 Tests were investigated.

The questions for the semi-structured interview covered the effectiveness of the Familiarization and Specification workshops. The questionnaire was validated by three participants during the pilot study, and the final version of the questionnaire is shown in Appendix N.

In order to make the purposes of the instruments clearer and more understandable, the following table presents the information in detail:

Table 3.11 Purposes of instruments

Instruments	Procedures	Purposes
Descriptor-sorting activities	Familiarization workshop	- to familiarize all the participants involved in the linking project with the CEFR - to explore whether the participants have reached an adequate understanding of the proficiency of the CEFR descriptors
The Dutch CEFR Grid	Specification workshop	- to analyze the content of the CET-4 Reading Comprehension Tests and to figure out a particular level of the CET-4 Reading Comprehension Tests in relation to the CEFR
Workshop feedback questionnaire	Familiarization workshop	- to investigate the participants' perceptions to the workshops
Semi-structured interview	Post-workshop	- to detect the problems the participants encounter at workshops

3.7 Data collection

In this study, both qualitative and quantitative data were collected. Qualitative data were gathered by means of a semi-structured interview and a questionnaire. Quantitative data referred to the data from the descriptor-sorting activities and the Dutch CEFR Grid. The details of the data collection are given below.

3.7.1 Qualitative data

Semi-structured interview

The semi-structured interview was a group interview conducted after completing the Familiarization and Specification workshops. Three participants were interviewed one by one and each interview lasted around 10 minutes.

To begin with, the interviewer (the researcher) explained the purpose, time, and format of the interview to the interviewees. Moreover, permission to make use of a digital recording was obtained from the interviewees before the start of the interview. Then, the researcher organized the interview with a series of questions. During the interviewing process, the interviewer recorded the conversations and later transcribed them for analysis.

Questionnaire

A questionnaire was used to investigate the participants' perceptions of the Familiarization and Specification workshops. The questionnaire was given to the participants before the workshops started and collected from them afterwards.

3.7.2 Quantitative data

Descriptor-sorting activities

In order to explore whether the participants had reached an adequate understanding of the CEFR descriptors, the participants were required to perform three

descriptor-sorting activities at both workshops. The detailed procedures for carrying out the activities are described in 3.3.1 b), d), and f). As soon as the participants finished the activities, the trainer assistant collected the judgment sheets from the participants and then calculated the scores to analyze the inter- and intra-rater reliability of each activity using the SPSS Programme.

The Dutch CEFR Grid

The linking procedure was used to analyze the CET-4 Reading Comprehension Tests with regard to the Dutch CER Grid and the analysis results were reported on the judgment sheets based on the Grid and three sets of test papers. As mentioned above, the analyses of the CET-4 Tests using the Dutch CER Grid were conducted in three separate sessions which lasted for two days. After the participants finished the analysis of the CET-4 Tests each time, all the judgment sheets were returned to the trainer assistant. Thus, the data was collected in the afternoon of the first day. There was one judgment sheet for one test paper and two judgment sheets for two sets of test papers on the second day.

3.8 Data analysis

3.8.1 Qualitative data analysis

Questionnaire

Content analysis is a procedure for the categorization of verbal or behavioral data, for purposes of classification, summarization and tabulation (Taylor, Bogdan & DeVault, 2015). A content analysis was conducted with the data obtained from the questionnaire. It involved coding and classifying data to make logical sense of the data collected and to highlight the important messages, features and findings.

Semi-structured interview

The data collected from the semi-structured interview was also analyzed by using a content analysis method. Specifically, the qualitative data was analyzed through the following steps:

Transcribing interview recordings verbatim;

- 1) Making brief notes in the margin of the transcripts when interesting or relevant information is found;
- 2) Reading the notes made in the margins and listing the different types of information found;
- 3) Checking the list and categorizing items into the same topics and themes;
- 4) Reviewing all of the categories and ascertaining whether some categories can be merged or if some need to be sub-categorized;
- 5) Returning to the original transcripts and ensuring that all the information has been categorized.

3.8.2 Quantitative data analysis

Descriptor-sorting activities

After collecting the data from the descriptor-sorting activities and calculating the scores of the participants, all the data was processed by SPSS to analyze the inter- and intra-rater reliability.

The Dutch CEFR Grid

A descriptive statistics report was used to analyze the data from the Dutch CEFR Grid. As explained above, the data (judgment sheets) regarding the analysis of the content of the CET-4 Reading Comprehension Tests with reference to the Dutch CEFR Grid were collected by the trainer assistant after the participants completed their

analysis of each set of test papers. Then, the data was analyzed by using descriptive statistics. Finally, the characteristics and the CEFR levels of the CET-4 Reading Comprehension Tests and test items were revealed.

The following table illustrates the data collection and analysis methods used to answer the research questions.

Table 3.12 Data collection and analysis methods

Research questions	Data collection methods	Data analysis methods
1. Which level are CET-4 reading comprehension texts situated in the CEFR and what are the major characteristics of the reading texts?	The Dutch CEFR Grid	Descriptive statistics (median)
2. Which level are the CET-4 reading comprehension test items situated in the CEFR and what are the major characteristics of the test items?		

3.9 Summary of the chapter

This chapter mainly describes the research methodology and explains the research design, research procedure, research context and participants, materials, and instruments used to collect its data. A total of three participants took part in the linking process. They were firstly required to self-study the materials and then attend Familiarization and Specification workshops. Familiarization workshop was to train the participants to be familiar with the CEFR by conducting descriptor-sorting activities. Specification workshop analyzed the content of CET-4 Reading Comprehension Tests in relation to the CEFR by using the Dutch CEFR Grid. Lastly, the results from the data analysis were analyzed. The findings will be presented in the next chapter.

CHAPTER 4

RESULTS

This chapter reports the results of the Familiarization and Specification workshops where the participants were trained to be acquainted with the CEFR for the first step and then to analyze the CET-4 Reading Comprehension Tests by using the Dutch CEFR Grid for the second step. It begins with the results of intra-rater and inter-rater reliability of the descriptor-sorting activities during the Familiarization workshop and it is followed by the results of the linking analysis of the CET-4 Reading Comprehension Test contents.

4.1 Results of Familiarization Workshop

As described in the previous chapters, the Familiarization workshop was held to enhance the participants' understanding regarding CEFR by conducting a series of descriptor-sorting activities. The participants were expected to analyze the reading texts and test items and situate them in relation to the CEFR levels at the Specification stage. All the activities at the Familiarization workshop have been described in 3.3.1 b), d), and f) in Chapter 3. To examine the effectiveness of the activities, intra-rater reliability and interrater reliability were investigated by considering that coefficients above 0.7 were acceptable (Kaftanjieva & Takala, 2002).

Intra-rater reliability in this study refers to the degree of agreement between the two-round descriptor-sorting activities (one at the Familiarization workshop and one

with the re-familiarization procedure at the Specification workshop) made by a single participant. As shown in Table 4.1, the correlation coefficient of intra-rater was 0.930, which was higher than the criterion of 0.7. In other words, the result indicates that the judgments made by the participants in the two activities were quite consistent.

Table 4.1 Intra-rater reliability of descriptor-sorting activities during the training

Scale Reliability Statistics		
mean	sd	Cronbach's α
2.111	0.259	0.930

Inter-rater reliability in this study was the degree of agreement among the three participants in terms of the two descriptor-sorting activities. It explored whether the participants agreed on the ranking of the descriptors of the scales. Table 4.2 shows that the results of the Spearman coefficient were higher than the accepted criterion of 0.7. Therefore, it can be concluded that the judgments made by the three participants in relation to the descriptor-sorting activities were highly consistent.

Table 4.2 Inter-rater reliability of descriptor-sorting activities in two rounds

mean	sd	Cronbach's α
2.197	0.253	0.865

Note: Of the observations, 39 were used, 0 were excluded, and 39 were provided.

The results of the intra- and inter-rater reliability illustrate that as a result of the workshop, each participant understood the CEFR scale very well and was able to rate the test items accurately with can-do statements listed in the CEFR. In other words, the judgments made by each participant in the two activities were reliable.

4.2 Results of Specification workshop

The results were presented by using the two research questions as a framework.

4.2.1 Findings to research question 1: which levels were the CET-4 reading comprehension texts situated in the CEFR and what were the major characteristics of the reading texts?

As stated in previous chapters, three sets of the CET-4 Reading Comprehension Tests with 12 texts and 90 questions in total were analyzed in the Specification workshop by using the Dutch CEFR Grid to identify the major characteristics and the CEFR levels of texts and their questions. In the Dutch CEFR Grid, the major characteristics of the texts were mainly composed of ten parts: text source, authenticity, discourse type, domain, topic, nature of content, text length, vocabulary, grammar, and estimated levels of the texts. These characteristics are reported individually as follows.

1) Text source

The first analysis of the CET-4 Reading Comprehension Tests was concerned with the text source. There were 40 text sources in total provided in the Dutch CEFR Grid. However, only a total of 4 categories were found from the 12 reading texts as shown in Table 4.3.

Table 4.3 Text sources of CET-4 reading texts

Text source	Frequency & Percentage (%)			
	Test 1	Test 2	Test 3	Total
Journal articles	2	2	1	41.67%
Magazines	0	1	3	33.33%
Newspapers	1	1	0	16.67%
Reports	1	0	0	8.33%
Total	4	4	4	100%

In the 4 categories of text sources, the highest frequency was for journal articles (41.67%), followed by magazines (33.33%) and then newspapers (16.67%). The lowest frequency was for the reports (8.33%). The figures in the table represent the exact number of each type of text source. For example, there were 2 journal articles in Test 1 and Test 2 respectively but only 1 in Test 3. There were 12 texts in three sets of the CET-4 Reading Comprehension Tests in total. The number of journal articles was 5, which means that the percentage of journal articles was 41.67%.

2) Authenticity

Authenticity was the second characteristic of the CET-4 Reading Comprehension Tests in the Dutch CEFR Grid. There were three kinds of input texts including 1) genuine, 2) adapted and 3) pedagogic. In this study, the genuine texts were in the original form without any change. Adapted texts refer to texts with changes by the researcher, such as adding or deleting the content or changing some of the words in the original texts. The pedagogic texts type was simplified for instructional purposes.

Of the 12 texts in the CET-4 Reading Comprehension Tests, 9 were classified as adapted and 3 as genuine, but none of them were pedagogic (See Table 4.4. below).

Table 4.4 Authenticity of CET-4 reading texts

Authenticity	Test 1	Test 1	Test 3	Total
Adapted	3	3	3	75.00%
Genuine	1	1	1	25.00%
Pedagogic	0	0	0	0%
Total	4	4	4	100%

3) Discourse type

The next characteristic of the CET-4 reading texts was the discourse type which referred to 1) mainly argumentative, 2) mainly narrative, 3) mainly expository, and 4) mainly instructive. In this study, these 4 discourse types referred to the text types which were argumentative-oriented, narrative-oriented, expository-oriented and instructive-oriented. Some texts combined two or more types of discourse. However, it was the major type of the text that determined the discourse type.

In the texts of the CET-4 Reading Comprehension Tests, only three types were found including mainly narrative, mainly expository and mainly argumentative as shown in Table 4.5.

Table 4.5 Discourse type of CET-4 reading texts

Discourse type	Frequency & Percentage (%)			
	Test 1	Test 2	Test 3	Total
Mainly narrative	2	2	2	50.00%
Mainly expository	1	2	1	33.33%
Mainly argumentative	1	0	1	16.67%
Mainly instructive	0	0	0	0%
Total	4	4	4	100%

Of the three types, 50% of the texts were classified into mainly narrative type, 33.33% were mainly expository and 16.66% were mainly argumentative. Interestingly, none of the texts were of the mainly instructive discourse type.

4) Domain

The fourth characteristic of the CET-4 reading texts was related to text domains. The Dutch CEFR Grid provided four categories for this and all of them were

found in the CET-4 Reading Comprehension texts as shown in Table 4.6.

Table 4.6 Domain of CET-4 reading texts

Domain	Frequency & Percentage (%)			
	Test 1	Test 2	Test 3	Total
Public	2	2	2	50.00%
Educational	1	0	2	25.00%
Personal	0	2	0	16.67%
Occupational	1	0	0	8.33%
Total	4	4	4	100%

As shown in Table 4.6, of all the texts in the CET-4 tests, half of them came from the public domain. A less frequent category was educational. Of all the texts, 3 texts belonged to the educational category with the proportion of 25% and 2 texts were categorized as 'personal' with the proportion of 16.67%. Compared with the previous three categories, the number of occupational texts was the least, with only a proportion of 8.33%.

5) Topic

The next characteristic was the topic. In the Dutch CEFR Grid, there were 14 categories of topics which were personal and identification, house and home, environment, daily life, free time, entertainment, travel, regulations with other people, health and body care, education, shop, food and drink, services, language, and weather. According to the participants' analysis, 7 kinds of topics were used in these 3 sets of CET-4 tests as shown in Table 4.7.

Table 4.7 Topics of CET-4 reading texts

Topics	Frequency & Percentage (%)			
	Test 1	Test 2	Test 3	Total
Education	1	1	3	41.67%
Education and body care	1	0	1	16.67%
Health	1	0	0	8.33%
Environment	1	0	0	8.33%
Daily life	0	1	0	8.33%
House and home	0	1	0	8.33%
Services	0	1	0	8.33%
Total	4	4	4	100%

Generally, the topic of education was the highest with 12 texts accounting for 41.67%. However, the proportion occupied by education and body care was less than half of that of education with only 16.67%. Interestingly, the topics of health, environment, daily life, house and home, and services received the same proportions with 8.33%.

6) Nature of content

The sixth characteristic was the content. The Dutch CEFR Grid divides this characteristic into four categories: only concrete content, mostly concrete content, fairly abstract content and mainly abstract content. However, only two categories were found in the CET-4 reading comprehension texts as shown in Table 4.8.

Table 4.8 Nature of content of CET-4 reading texts

Nature of content	Frequency & Percentage (%)			
	Test 1	Test 2	Test 3	Total
Mostly concrete content	4	3	4	91.67%
Only concrete content	0	1	0	8.33%
Fairly abstract content	0	0	0	0%
Mainly abstract content	0	0	0	0%
Total	4	4	4	100%

In the CET-4 reading comprehension tests, the content of most of the texts was ‘mostly concrete’ and they accounted for 91.67%. On the contrary, only 1 text belonged to ‘only concrete content with 8.33%. In other words, there existed a significant difference between these two categories with regard to the frequency proportions.

7) Text length

In the Dutch CEFR Grid, text length depends on the number of words in each text. In general, no significant difference was found among these three texts. The specific text lengths are shown in Table 4.9.

Table 4.9 Text lengths of CET-4 reading texts

Text length				
Test	Text 1	Text 2	Text 3	Text 4
Test 1	216	1013	350	331
Test 2	253	1039	355	354
Test 3	239	1034	341	354

In each of the CET-4 reading tests, the words of the first text were around 210 to 250. The third and fourth text were about the same length, i.e., 331 to 355. Note that the length of the second text was much longer than the other three with the number of words ranging from 1031 to 1039. The number of words consistently illustrate a well-controlled text length in terms of the number of words and the sequence of texts in each test.

8) Vocabulary

Vocabulary was also one of the characteristics analyzed in the Dutch CEFR Grid by using an online tool supported by the Council of Europe called "English Vocabulary Profile Online" in order to avoid any subjective bias on the part of the participants. The results of the analysis are shown in Table 4.10.

Table 4.10. CEFR level of vocabulary in CET-4 reading comprehension texts

CEFR level of Vocabulary	Frequency & percentage (%)			
	Test 1	Test 2	Test 3	Total
C2	24	24	24	2.86%
C1	43	43	38	4.92%
B2	192	150	140	19.13%
B1	204	176	150	21.04%
A2	126	130	127	15.20%
A1	311	196	228	29.18%
Unlisted	57	53	83	7.67%

In Table 4.10, the vocabulary at A1 level occupied the highest proportion with 29.18% in total in the three sets followed by B1 level with 21.04%. The number of B2 vocabulary items was only 2% less than B1 with 19.13%. The vocabulary at A2 level was only 15.20%. As expected, the vocabulary at the C1 and C2 levels were the lowest with 4.92% and 2.86%, respectively. The proportion of vocabulary that could not be found using the tool of English Vocabulary Profile Online was 7.67%, for example, proper nouns (people's names), compound words and words that are not commonly seen in texts.

9) Grammar

Grammar had to be analyzed with reference to the Dutch CEFR Grid. Similarly, to vocabulary, English Grammar Profile Online was used to find the CEFR

level of each grammar structure. Table 4.11 shows the frequency and percentage of the grammar items for each level in all the texts.

Table 4.11 CEFR level of grammar in CET-4 reading texts

CEFR level of grammar	Frequency & Percentage (%)			
	Test 1	Test 2	Test 3	Total
C2	0	0	0	0%
C1	1	1	1	1.05%
B2	35	28	36	34.49%
B1	27	30	23	27.87%
A2	18	35	31	29.27%
A1	6	11	4	7.32%
Total	87	105	95	100%

As shown in Table 4.11, the percentage of grammar structures found in three sets of CET-4 Tests were 7.32%, 29.27%, 27.87%, 34.49% and 1.05% from A1 to C1 level, respectively. Among the five CEFR levels, C1 was the lowest with only 3 grammar structures. In contrast, B2 level had the highest proportion with 34.49%. The number of A2 grammar structures was similar to B1 with 84 and 80, respectively. The number at A1 level was 21. No grammar structures were found at C2 level.

10) Level of texts estimated

The last element analyzed in the Dutch CEFR Grid to estimate the text levels. As stated in the previous chapter, the six CEFR levels present the steps which ascend with A1 the lowest and C2 the highest. According to the participants' analysis, they finally came to the conclusion that the texts in CET-4 consisted of three levels. These three levels are shown in Table 4.12.

Table 4.12 Levels of texts estimated for CET-4 reading comprehension texts

Texts level estimated	Frequency & Percentage (%)			
	Test 1	Test 2	Test 3	Total
C2	0	0	0	0%
C1	0	0	0	0%
B2	3	1	2	50%
B1	1	2	2	41.67%
A2	0	1	0	8.33%
A1	0	0	0	0%
Total	4	4	4	100%

It was found that 50% of the texts in CET-4 were graded at B2 level. These texts consisted of 3 from the first set (December, 2016), 1 from the second set (June, 2017), and 2 from the third set (December, 2017). Five texts were found to be at B1 level. However, only one out of 12 texts was at A2 level while none of the of the CET-4 texts was found to be at levels A1, C1, or C2.

4.2.2 Findings for research question 2: Which levels were CET-4 reading comprehension test items situated in the CEFR and what were the reading items' major characteristics?

In the previous section, the analysis was mainly concerned with the characteristics of the input texts of the CET-4 Reading Comprehension Tests. The following section focuses on the characteristics of the test items. In the Dutch CEFR Grid these are composed of three aspects including item type, operations and the estimated level of the item.

1) Item type

The first part of the characteristics of the items concerns the item type which is divided into three categories: selected responses, short constructed responses

and extended constructed responses, and there were corresponding types under each category. For example, multiple choice, banked multiple choice, and true/false belonged to the category of selected responses. Likewise, gap filling, short answers, and clozes, pertained to the category of short constructed responses.

According to the participants' analysis of the CET-4 Tests, three kinds of item types were found: cloze, matching information and multiple choice, and the number of each item type was 10 as shown in Table 4.13. Therefore, the proportion of each type was 33.33%.

Table 4.13 Item types of CET-4 reading comprehension tests

Item type	Frequency & Percentage (%)			
	Test 1	Test 2	Test 3	Total
Cloze	10	10	10	33.33%
Matching information	10	10	10	33.33%
Multiple choice	10	10	10	33.33%
Total	30	30	30	100%

2) Operations

The second part was the operations of the items. As mentioned in chapter 3, changes were made by the researcher in terms of the operations in the Dutch CEFR Grid to make them more applicable to the CET-4 Reading Comprehension Tests (see the revised version of operations in Table 3.9 in chapter 3).

As shown in Table 4.14, the operations of identifying the correct word from given options according to the context was the same as distinguishing the different forms of the same idea with 10 items in each test paper. Next, understanding the details also occurs at a high frequency with a percentage of 24.44% while inferring

speaker's/writer's attitude / mood / opinion had a lower frequency than understanding the details with 6 out of 90. However, there were only 2 items for identifying the topic and main idea/gist and 1 for generalizing and drawing a conclusion. Interestingly, the operation of deducing the meaning of an unknown word from context, which was commonly seen in the items of other kinds of tests such as NMET and TEM-4, did not occur in the three sets of CET-4 Reading Comprehension tests.

Table 4.14 Operations of CET-4 Reading Comprehension Test Items

Operations	Frequency & Percentage (%)			
	Test 1	Test 2	Test 3	Total
Identify the correct word from given options according to context	10	10	10	33.33%
Distinguish the different forms of the same idea	10	10	10	33.33%
Understand the detail	8	8	6	24.44%
Infer speaker's/writer's attitude/mood/opinion	2	1	3	6.67%
Detect the topic and main idea/gist	0	1	1	2.22%
Generalize and draw the conclusion	0	0	1	1.11%
Deduce the meaning of an unknown word from context	0	0	0	0%
Total	30	30	30	100%

3) Item level estimated

The last characteristic of the items was estimated level of the item. As mentioned in chapter 3, the CET-4 Reading Comprehension Tests for this study included 12 texts and 90 items. The CEFR levels of the items were the same as the number of texts. However, there existed a significant difference between texts and items in terms of the percentage at each level. As for the test items, the highest proportion was at B1 level (76.67%), followed by B2 (18.89%). The lowest frequency was A2 with

just 4 items (see Table 4.15).

Table 4.15 Estimated level of Items

Item level estimated	Frequency & Percentage (%)			
	Test 1	Test 2	Test 3	Total
C2	0	0	0	0%
C1	0	0	0	0%
B2	4	3	10	18.89%
B1	25	24	20	76.67%
A2	1	3	0	4.44%
A1	0	0	0	0%
Total	30	30	30	100%

4.3 Summary of this chapter

This chapter presents the results of an analysis of three sets of CET-4 Reading Comprehension Tests by using the Dutch CEFR Grid in relation to the CEFR levels. For the input texts of CET-4 tests, text source, authenticity, discourse type, domain, topic, nature of content, vocabulary, grammar, and estimated levels of texts were analyzed. In terms of test items (question setting), item type, operations and the estimated levels of the items were examined.

CHAPTER 5

DISCUSSION, IMPLICATIONS AND CONCLUSION

This chapter is composed of five sections. They are, first, the summary of findings, second, the discussions of two workshops and findings based on former studies and interpretations, third, the implications of the study fourth, the limitations and recommendations for future research and last, the conclusion of the study.

5.1 Summary of findings

This study aimed at linking CET-4 Reading Comprehension Tests to the CEFR. The findings are summarized in Tables 5.1 and 5.2.

Table 5.1 Findings of the first research question

Research question 1	Findings	
Which levels were CET-4 reading comprehension texts situated in the CEFR and what were the reading texts' major characteristics?	Text source	Journal article
	Authenticity	Adapted
	Discourse type	Mainly narrative
	Domain	Public
	Topic	Education
	Nature of content	Mostly concrete content
	Vocabulary	A1
	Grammar	B2
	Estimated text level	B1

Table 5.1 displays the final results of the characteristics of the texts and the linking of the CET-4 texts to the CEFR. In three sets of the CET-4 texts, adapted texts, journal articles, and mainly narrative accounted for the highest frequency. Of all the texts, half

of them were from the public domain and the major topic was education. The content of most of the texts was concrete rather than abstract. For the CEFR levels of vocabulary, grammar and texts, the results were completely different with A1 for vocabulary, B2 for grammar and B1 for texts.

Table 5.2 Findings of the second research question

Research question 2	Findings	
Which levels were CET-4 reading comprehension test items situated in the CEFR and what were the reading items' major characteristics?	Item type	Cloze/ Matching information/ Multiple choice
	Operations	Identify text structure/ connections of parts from explicit information/ Recognize the main idea or theme in the paragraphs or sections of a text
	Item level estimated	B1

The second part involved the characteristics of the test items which were divided into three aspects: item type, operations and the CEFR. Cloze, matching information and multiple choice were the three main item types found and the proportion of them were the same. In terms of operation, there were mainly two types including identifying text structure/connections of parts from explicit information and recognizing the main idea or theme in the paragraphs or sections of a text. Lastly, the CEFR level of items was the same as for the texts at B1 level.

5.2 Discussion

5.2.1 The Effects of Familiarization Workshop on Linking Consistency

Even though the Familiarization workshop did not deal with the research questions in this study, it is still worth discussing specifically because of the crucial role

of the training session at this workshop for the later analysis of the test papers. As explained in previous chapters, the Familiarization workshop was held to train the participants to have an in-depth knowledge of the CEFR and its descriptors so that they were capable of accurately linking the CET-4 Tests to the CEFR in the Specification workshop. In short, the training session was highly crucial for the success of determining the CEFR levels. The effectiveness of the Familiarization workshop was demonstrated through the inter-rater reliability which was shown in chapter 4.

With regard to the participants' opinions concerning the effectiveness of the workshop, an open-ended questionnaire was used. The results revealed that most of the participants were quite satisfied with the activities. In their opinions, the pace as well as the connection of each activity was appropriate to gradually strengthen their understanding of the CEFR. Furthermore, the content of each activity was useful, adequate, and self-contained which helped to consolidate their background knowledge. Not only did they understand the purpose of the workshop, but they also mastered use of the different descriptors at each CEFR level after two training sessions. All the participants stated that they were confident that they were capable of linking the CET-4 Tests to the CEFR though it was the first time they had learnt about the CEFR in detail.

The main objective of this activity was to provide the participants with a fuller understanding of the descriptors in the CEFR scales. Since English language teaching and learning in China has its own national reference and criteria which were designated by the government and had no connection with the global CEFR levels, it was difficult for the research participants to determine their own language proficiency levels. For example, in reality they might have some language competence at C2 level but they failed to perform some of the B2 tasks. This kind of incongruence between

their reading ability and the CEFR, on the however, served to develop their interest and deepen their understanding of the descriptors. In other words, according to the participants this activity was very useful although it was also complicated and time-consuming. Therefore, the trainer played a key role in leading the Familiarization workshop.

Although the arrangement of activities during the Familiarization workshop was found to be reasonable in terms of the pace and content, the time for the descriptor-sorting activity of the reading scales in the first round was insufficient in terms of the time originally stipulated. The estimated time for this activity was 45 minutes in the CEFR Manual, but additional 25 minutes was necessary during the workshop. The two reasons for extending the time were, firstly, the time individuals spent in completing each familiarization activity depended mainly upon the extent of their familiarity with the CEFR (CoE, 2001). Secondly, the effectiveness of the descriptor-sorting activity of reading scales directly affected the participants' understanding of the CEFR and their ability to determine the appropriate levels. Thus, it was suggested that the time of 45 minutes for this activity was insufficient so it was extended appropriately in the first session. After completing the first training session, the participants had acquired sufficient knowledge of the CEFR scales and also, they had a clearer idea of the salient features of the learners' language proficiency in reading skills at different levels after the the second session was completed. The time for the descriptor-sorting activity in terms of reading was shortened to 30 minutes in the second session as a result of to the effectiveness of the workshop and their understanding of the CEFR levels.

5.2.2 Discussion of Specification workshop

The activities at the Familiarization workshop were widely praised and

except for the time allocated for the descriptor-sorting activities of the reading scales in the first training session. In contrast, some problems occurred while linking the CET-4 Reading Comprehension Tests to the CEFR at the Specification workshop.

The most serious problem for the participants was that the boundaries between the different levels were hard to determine due to the ambiguity of expressions such as long, short, simple, and complex, etc. in some of the descriptors in the linking process at the Specification workshop. Differentiating these factors thus became extremely difficult on account of the lack of clear criteria for linking.

Another difficulty encountered by the participants was that the identification of some texts fell somewhere between two CEFR levels. In other words, the span between two language levels was relatively large. For example, the following two levels in the overall reading scales are explained as:

B1: Can read straightforward factual texts on subjects related to his/her field and interest with a satisfactory level of comprehension.

B2: Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using appropriate reference sources selectively.

The descriptor of B1 level is concerned with straightforward factual texts related to the learners' field. However, B2 level expects the reader to adapt reading speed and style according to different texts independently. As mentioned in chapter 2, the plus levels were added to some scales to provide the salient characteristics of the CEFR levels for interaction and production and the salient characteristics of CEFR levels for reception. Unfortunately, not every language scale has plus levels such as B1 and B1+, B2 and B2+. Because of the large gap between the two levels, the researcher

suggests that the plus levels should be provided for all the language levels. If this is done, the linking work could be conducted more easily.

It should also be clearly clarified that, as a result of the participant's problems with the levels of texts and items mentioned above, a discussion session was held in order to reach a consensus on the final results.

5.2.3 Discussion of final results

In 2012, Huang and Jia conducted a study to investigate the feasibility of linking CET-4 to the CEFR by using the Dutch CEFR Grid in Mainland China. The results of their study claimed the feasibility of linking and the high level of consistency between the characteristics of the CET-4 texts and the CEFR with regard to textual origin, types, dimension, and topics. In a review of Huang and Jia's study on this linking project, the differences and similarities which were found area shown in Table 5.3 and each characteristic is discussed in detail as follows.

Table 5.3 A comparison of the findings in this study and with those of Huang and Jia's study (2012)

Characteristics of texts	Huang and Jia (2012)	The present study
Texts source	a variety of text sources (14 types) were found	only 4 types were found
Discourse type	mainly communicative	mainly narrative
Domain	all four types were found with the highest frequency for personal	all four types were found, but the public domain had the highest frequency
Topic	multiple types were found but education had the highest frequency	multiple types were found but education had the highest frequency

5.2.3.1 Texts source: Journal article

It is important to shed light on the completely different results found between these two studies, in which a variety of text sources (14 types) were found in Huang and Jia's study. However, only 4 types were found in this study where journal articles covered the largest proportion. As mentioned in chapter two, CET-4 was reformed several times in terms of its format and contents. According to the test specification for reading in 2016 (the latest reform of CET-4), the description explicitly revealed that the test takers should be able to understand relatively straightforward journal articles and other kinds of English materials. However, in the 2016 test specification, the description for reading comprehension showed that students should be able to understand general articles. This might be the reason why the differences in the main text sources as well as the number of text sources exist in the two studies.

5.2.3.2 Domain: personal and public

The findings of this study with regard to text domain resemble those of Huang and Jia (2012), which means that all four types of domain were found in both studies. However, the priority given to the types of text domain was quite different. In Huang and Jia's study, the texts in the personal domain had the highest frequency with 122 out of 170 including listening and reading texts together, but in this study the most common was the public domain with half out of all the texts analyzed. The possible reason for this difference might be that Huang and Jia analyzed not only reading with 30 texts but also listening with 140 texts. Furthermore, the reform of the CET-4 tests in 2016 could also be one of the factors in causing these differences.

It is worth mentioning that the CEFR emphasizes the personal domain for language teaching, learning and testing. Nonetheless, from the results found in this

study, the text in the personal domain only accounted for 16.67%. Therefore, it is necessary for the test developers to consider the need to lay stress on the number of texts in the personal domain in CET-4 reading comprehension tests in order to be in line with international standards. Another reason for adding the texts in the personal domain is the motivational effects of domains on language learners. For example, students may be chiefly interested in developing personal relations. Adding more reading texts from the personal domain in CET-4 thus could possibly interest the test takers more than other domains.

5.2.3.3 Topic

The results of the text topics of the CET-4 reading comprehension texts were similar to those of Huang and Jia's study in two aspects. The first aspect was concerned with the many different types of topics found in the two studies. A total of 7 topics was found from 12 reading texts. In both this study and Huang and Jia's study nearly all the types listed in Dutch CEFR Grid occurred. The findings of the two studies are in line with the requirements of the CEFR for the diversity of topics. The second aspect is related to the proportion of a certain type of topic in both studies. Among the various topics found in the two studies, emphasis was placed on the topic of education.

5.2.3.4 Text length, vocabulary and grammar

It was found that the length of the CET-4 reading comprehension texts followed certain patterns with words 250/1000/350/350 in each set of test papers and the vocabulary was mostly at A1 level, the most basic level in the three tests. On the other hand, most of the grammar structures were quite challenging at the level of B2. While analyzing the texts, an interesting finding concerning the second text in each set of papers was reflected by the participants as shown in Table 5.4.

Table 5.4 Text length, vocabulary and grammar of the second texts in three sets of test papers

Test 1			Test 2		Test 3	
Levels	Vocabulary	Grammar	Vocabulary	Grammar	Vocabulary	Grammar
C2	3	0	15	0	12	0
C1	11	1	20	0	21	3
B2	107	25	67	11	75	21
B1	111	9	88	14	75	7
A2	59	9	61	25	63	19
A1	226	1	89	7	113	0

As shown in Table 5.4, most of the vocabulary in the second text is at A1 level, which is the same as the results found in the other three texts. However, the level of grammar was not lower except for the text in the second set and the length of the second text was longer than the other texts. However, the CEFR level of most items in the second texts was B1 on the basis of the participants' analyses. More specifically, the length of the second texts was nearly four times longer than that of the other texts in each test paper while the level of difficulty for the grammar and other items remained the same, which means that there is a considerable disparity between the second texts and the others in the same test paper.

It was very difficult for the researcher to find a reasonable explanation for this disparity. In order to make the level of difficulty relatively consistent in terms of CEFR level or in other words, to reduce the disparity between the second text and other three in one set of test papers, a practical alternative would be to shorten the second text to match the other texts. It is difficult to give a specific number of words in a text since studies need to be conducted to find the ideal by the test developers. Alternatively, the CEFR level of grammar or test items was lower to a certain extent,

similar to the difference in length of the second text in the second set of test papers. Either of these two possibilities could be used to make further improvements to the CET-4 Reading Comprehension Tests.

5.2.3.5 CEFR level of texts

Since Huang and Jia (2012) did not link the CET-4 texts and test items to a certain CEFR level, no comparison could be made between the two studies in this area. The reading comprehension texts in the present study were at B1 level according to the participants' analyses. In order to shed light on the research findings more clearly, the descriptors of B1 level for reading in the CEFR are listed below.

Table 5.5 B1 level of reading part in CEFR

READING FOR ORIENTATION	
B1	Can scan longer texts in order to locate desired information, and gather information from different parts of a text, or from different texts in order to fulfil a specific task.
READING FOR INFORMATION AND ARGUMENT	
B1	Can identify the main conclusions in clearly signaled argumentative texts. Can recognize the line of argument in the treatment of the issue presented, though not necessarily in detail. Can recognize significant points in straightforward newspaper articles on familiar subjects.
OVERALL READING COMPREHENSION	
B1	Can read straightforward factual texts on subjects related to his/her field and interest with a satisfactory level of comprehension.

As shown in Table 5.5, learners at B1 level have the ability to locate and gather the necessary information from different parts of texts. Furthermore, learners at this level are able to discern the main conclusion from texts and generally identify the line of argument. More importantly, they are also able to identify important ideas in newspaper articles on familiar topics.

As for overall reading comprehension at B1 level, the descriptors revealed that the subjects of the texts were related to the learners' fields and the texts were relatively longer than those at A1 and A2 levels. Since the test takers of CET-4 Reading Comprehension Tests were all college students and the topic of most of the texts in the three sets of papers was about education, they were related to the test takers' field. This explains why the participants considered that the texts analyzed could be linked to B1 level. Although the participants were sure about the results of their analyses, they also encountered some problems and difficulties. As mentioned in 5.2.1, the most serious problem was the descriptors used such indefinable expressions as short, longer and long texts without providing examples which could help to classify the CEFR input texts. Deciding whether a text was short or long thus became somewhat subjective merely depending on the participants' own experience and language competence. With respect to the foregoing, the researcher would suggest that a document or standard can be provided by the CEFR to clearly demonstrate the requirements of the level and length of texts so that the linking work can be conducted more on an evidence-based rather than experience-based procedure. In addition, it would be easy for the test developers to develop CEFR-based tests whose CEFR levels were completely suitable for the test takers in terms text length.

5.3 Implications

5.3.1 Implications for CET-4 test developers

As described previously, the objective of the study was to link the CET-4 Reading Comprehension Tests to the CEFR. This research offered an opportunity to review the quality of the newly reformed CET-4 Reading Comprehension Tests and the

findings could possibly provide some suggestions for the further improvement of the test papers. Even though no single test paper will be able to meet the requirements or standards of different language contexts completely, the development of any large-scale standardized test must undergo a process of continuous improvement (Jin, 2005). This explains why the CET-4 tests have been reformed many times. In order to make the CET-4 test more sophisticated and internationally acceptable to measure college students' language proficiency, it is strongly recommended that the test developers make further improvements to the test.

As revealed in this study, the CET-4 test papers provide fewer texts in the personal domain, which in contrast, are highly valued in the CEFR document. Therefore, more emphasis should be placed on the personal domain by the test developers. In addition to this, text length is another factor to be taken into consideration as well. To improve this factor, shortening text lengths or lowering the levels of difficulty should be investigated.

5.3.2 Implications for college English teachers

It cannot be denied that teachers play a crucial role in spreading knowledge, stimulating and raising students' interest in learning as well as training them in the skills of problem-solving. Through the analysis of the CET-4 Reading Comprehension Tests, the present study has tried to provide some useful suggestions for college English teachers.

First of all, it is necessary to increase the teaching and training of various kinds of reading texts which are from different sources, domains and topics in normal teaching practice. In this way, students will be more confident when they encounter familiar texts related to the three aspects during a test.

Secondly, since the second text in each set of reading test paper is relatively long it is strongly recommended that teachers should train students to improve their reading speeds and strategies for finding the correct answers. According to the test specification of the CET-4 Reading Comprehension tests, students should not spend more than 40 minutes on the reading part, which means that the average time for reading and answering questions for each text is 10 minutes. In this respect, improving reading speed and skills in answering questions will be very useful for students and will enable them to perform better in the tests.

Last but not least, the CEFR level of grammar of most of the CET-4 reading texts is B2 and the most frequent grammar structures which belong to this level in reading texts are clauses such as objective clauses and attributive clauses as counted by the researcher. Therefore, it is extremely important for college English teachers to consider the need to train students in these two types of grammar structures until the students can fully master them. If this can be done, students will encounter less difficulties with the grammar structures and understand the texts better.

5.4 Limitations of the study and recommendations for future research

This study set out to link the CET-4 Reading Comprehension Tests to the CEFR and the results obtained from this study illustrate the feasibility and reliability of linking the CET-4 Reading Comprehension Tests to the CEFR by using the Dutch CEFR Grid. Inevitably, the design of the current study may have some inherent limitations even though it provided some penetrating insights into the improvement of the CET-4 Reading Comprehension Tests.

First of all, the number and composition of the analysts who participated in the

linking process is not sufficient as suggested in the Manual. According to the Manual, members and experts engaged in the linking process should include item writers, test administrators, language assessors, and so on. However, due to the limited number of qualified analysts available, this study was carried out on a relatively small scale. Thus, any future studies should work on a larger scale and include a set of more experienced analysts with in-depth training to participate in the linking project. Moreover, as Martyniuk (2010) recommended, it is impossible that a single one-off study will provide sufficient evidence for complete alignment. To this end, studies with more professionals in this topic are strongly needed to verify the results obtained from the current study and the research findings generated from future projects will then be more reliable and significant.

Secondly, this study merely aimed at linking the CET-4 Reading Comprehension Tests to the CEFR due to the limitation of time. Therefore, it is strongly suggested that future researchers should also link both the listening and reading sections and even all four skills in the CET-4 tests to the CEFR on a larger scale. On the one hand, the results would be more comprehensive and thorough. On the other hand, analysts who are involved in the linking process would obtain a better understanding of the CEFR.

Thirdly, it must be acknowledged that the research results reported in this study are based on just three sets of tests as recommended in the Manual. This sample might be too small. Therefore, future studies should consider using more sets of tests to conduct the linking project. At the same time, this study also recommends that the CEFR Manual provide more sets of test papers to enable researchers to conduct further linking projects since the number of texts in different types test papers varies. For example, the CET-4 reading part includes 4 texts, but there are 5 in NMET and 3 in

IELTS and TOEFL. Due to the different number of texts in each test paper, there will be discrepancies in terms of their results. As a consequence, using more representative texts to conduct a linking project will result in more comprehensive findings.

Lastly, this study only focuses on linking the CET-4 Reading Comprehension Tests to the CEFR and examining the characteristics of texts and test items instead of finding the cut-off scores of CET-4. Future studies could investigate the cut-off scores closely which would provide another angle for a linking project of CET-4 and CEFR. If this can be achieved, test takers will clearly understand what a particular CEFR level means in relation to learners' reading proficiency and how a CET-4 test-taker will be able to perform in terms of the criteria specified in the CEFR levels.

5.5 Conclusion

These final remarks bring the present study to a conclusion. The study aimed at linking the CET-4 Reading Comprehension Tests to the CEFR by following the procedures provided in the Manual. To conduct the linking project, the Dutch CEFR Grid was used to examine the content of the tests in terms of the specification procedure. It was found that both the texts and test items of the CET-4 reading comprehension tests at B1 level of the CEFR and improvements to the CET-4 regarding text domains and length of texts should be taken into consideration. The researcher hopes that more studies on linking CET-4 Tests to the CEFR can be conducted so as to validate the research findings of the current study and to improve the quality of the CET-4 Test so that it will become internationally acceptable.

REFERENCES

- Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal*, 91(4), 659-663.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Cao, Y. Y., & Wang, J. Y. (2015). A study on the content validity and parallel-forms reliability of “multiple-volume test” of CET-4 reading comprehension subtests. *Journal of Heilongjiang College of Education*, 34(2), 160-161.
- Cheng, L. (2005). *Changing language teaching through language testing*. Cambridge: Cambridge University Press.
- Cheng, H. H. (2016). Research on Validity Analysis of Careful Reading in CET 4. *Advanced in Social Science, Education and Humanities Research*, 92, 385-389.
- Cheng, L., & Curtis, A. (Eds.). (2010). *English language assessment and the Chinese learner*. Routledge.
- Council of Europe (2001) *Common European Framework of Reference for Languages: learning, teaching and assessment*. Cambridge: Cambridge University Press.
- Council of Europe (2009) *Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching and assessment. A manual*. Strasbourg: Council of Europe.
- Eckstein, M. A., & Noah, H. J. (1993). *Secondary school examinations: International perspectives on policies and practice*. Yale University Press.

- Fan, J., Ji, P., & Song, X. (2014). Washback of university-based English language tests on students' learning: A case study. *The Asian Journal of Applied Linguistics*, 1(2), 178-91.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. *Washback in language testing: Research contexts and methods*, 191-210.
- Figueras, N., North, B., Takala, S., Verhelst, N., & Van Avermaet, P. (2005). Relating examinations to the Common European Framework: A manual. *Language Testing*, 22(3), 261-279.
- Figueras, N. (2012). The impact of the CEFR. *ELT journal*, 66(4), 477-485.
- Fleckenstein, J., Keller, S., Krüger, M., Tannenbaum, R. J., & Köller, O. (2020). Linking TOEFL iBT® writing rubrics to CEFR levels: Cut scores and validity evidence from a standard setting study. *Assessing Writing*, 43, 100420.
- Franz, J., & Teo, A. (2018). 'A2 is Normal'—Thai Secondary School English Teachers' Encounters with the CEFR. *RELC Journal*, 49(3), 322-338.
- Fu, G. F. (2011). An introduction to the Common European Framework of Reference and European Language portfolio. *Higher Education Development and Evaluation*, 27(1), 110-115.
- Fylan, F. (2005). Semi structured interviewing. In Miles, J., & Gilbert, P (Eds.), *A handbook of research methods for clinical and health psychology* (pp. 65-78). Oxford University Press.
- Gao, C. (2011). A comparative study on the listening subtests of CET and IELTS listening. *Science Survey*, (17), 26-26.

- Green, A. (2018). Linking tests of English for academic purposes to the CEFR: The score user's perspective. *Language Assessment Quarterly*, 15(1), 59-74.
- Gronlund, N. E. (1998). *Assessment of student achievement*. Allyn & Bacon Publishing, Longwood Division, 160 Gould Street, Needham Heights, MA 02194-2310; tele.
- Han, B. (2002). The need for the reform of English proficiency test for college students. *National Research Center for Foreign Language Education: Foreign Languages and Literature*, 34(6), 410-411.
- Han, B., Dai, M., & Yang, L. (2004). Problems with College English Test as emerged from a survey. *Foreign languages and their teaching*, 179(2), 17-23.
- Hao, W. L., Feng, J., Li, N., & Yang, J. H. (2015). A content validity analysis of CET-4 reading comprehension subtests. *Examination Research*, 4, 012.
- Hou, X, N. (2010). Problems of speaking proficiency among university the students in China. *The Border Economy and Culture*, 84(12), 160-162.
- Huang, B., & Wu, Y. (2004). Problems exist in Chinese university the students' speaking and solutions. *Journal of Shenyang Medical College*, 4(6), 251-252.
- Huang, T., & Jia, G. D. (2012). The feasibility of aligning language tests with the Common European Framework- taking CET as an example. *Foreign Language Testing and Teaching*, (1), 38-49.
- Hughes, A. (2007). *Testing for language teachers*. Ernst Klett Sprachen.
- Hulstijn, J., C. Alderson, and R. Schoonen (2010) Developmental stages in second language acquisition and levels of second language proficiency: are there links between them? In I. Bartning, M. Martin, and I. Vedder (eds.), *Communicative proficiency and linguistic development: intersections between SLA and language testing research*. (pp. 11–20). Rome: Eurosla Monographs. 11–20.

- Hung, N. N. (2015). Vietnam's National Foreign Language 2020 Project: Challenges, Opportunities, and Solutions. *English for ASEAN integration: Policies and practices in the region*, 62.
- Jin, Y. (2005). The national college English test of China. *International Association of Applied Linguistics (AILA), Madison, WI*.
- Jin, Y. (2010). College English education in China: From testing to assessment. *Language assessment in Asia: Local, regional, or global*, 1-28.
- Jin, Y., Wu, Z., Alderson, C., & Song, W. (2017). Developing the China Standards of English: challenges at macropolitical and micropolitical levels. *Language Testing in Asia*, 7(1), 1.
- Kecker, G., & Eckes, T. (2010). Putting the Manual to the test: The TestDaF–CEFR linking project. *Aligning tests with the CEFR: Reflections on using the Council of Europe's draft Manual*, 50-79.
- Khalifa, H., & French, A. (2009). Aligning Cambridge ESOL Examinations to the CEFR: Issues & Practice. *Cambridge ESOL Research Notes*, 37, 10-14.
- Khalifa, H., French, A., & Salamoura, A. (2010). Maintaining alignment to the CEFR: the FCE case study. *Aligning tests with the CEFR*, 80-101.
- Kunnan, A. J. (2014). *The Companion to Language Assessment: Abilities, Contexts, and Learners*. Wiley Blackwell.
- Li T. A. (2008). *A comparative study on the reading comprehension subtests of the new CET4, old CET4 and the IELTS* (Master's thesis, Shanghai International Studies University).
- Little, D. (2006). The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact. *Language Teaching*, 39(3), 167-190.

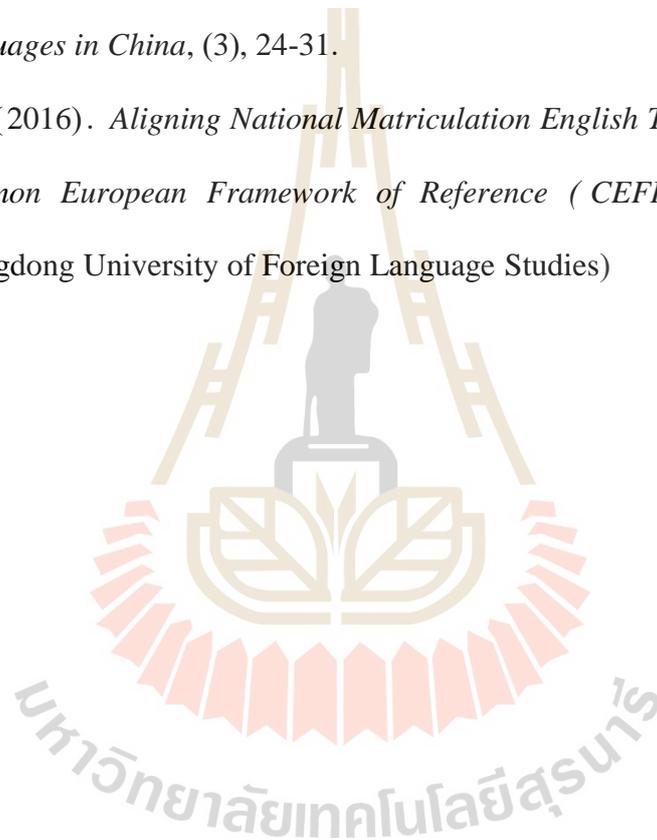
- Liu, J. G. (2012). *Aligning TEM-4 with the CEFR* (Master's dissertation, Henan Normal University).
- Ma, F. (2014). College English Test: To Be Abolished or To Be Polished. *Journal of Language Teaching & Research*, 5(5).
- Martyniuk, W. (Ed.) (2010). *Aligning tests with the CEFR: Reflections on using the Council of Europe's draft Manual (Vol.33)*. Cambridge: Cambridge University Press.
- National College English Testing Committee. (2016). *National College English Test Syllabus*.
- Negishi, M. (2012). The Development of the CEFR-J: Where we are, where we are going. *New perspectives for foreign language teaching in higher education: Exploring the possibilities of application of CEFR*, 105-116.
- Negishi, M., & Tono, Y. (2014). An update on the CEFR-J project and its impact on English language education in Japan. *ALTE Paris Retrieved March, 14, 2016*.
- Norland, E. V. T. (1990). Controlling error in evaluation instruments. *Journal of extension*, 28(2), 56.
- North, B. (2000). *The development of a common framework scale of language proficiency* (Vol. 8). Peter Lang Pub Incorporated.
- North, B. (2007). The CEFR: Development, theoretical and practical issues. *Babylonia*, 1, 22-29.
- North, B. (2014). *The CEFR in practice* (Vol. 4). Cambridge University Press.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge University Press.

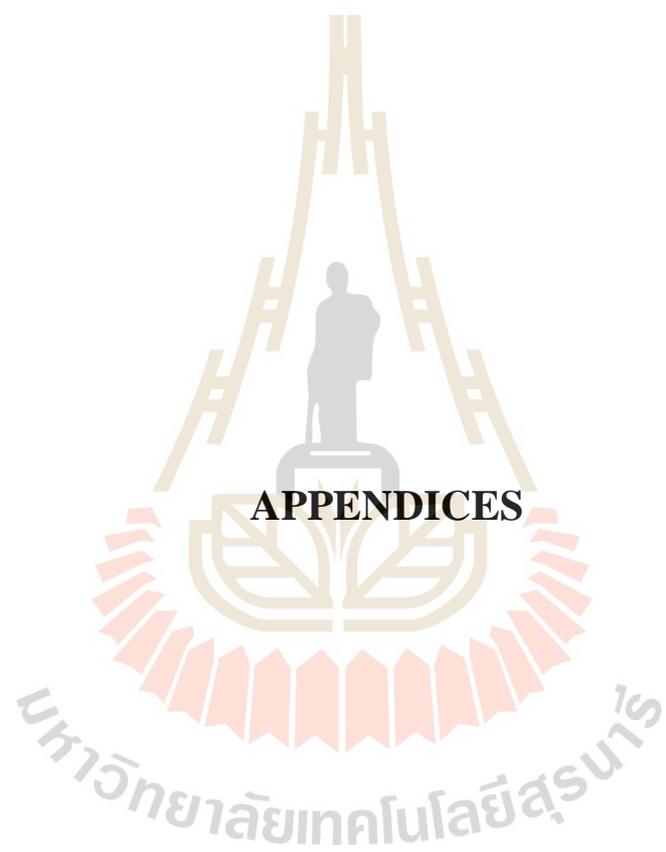
- Office of the Basic Education Commission. (2014). Guidelines for practices by Ministry of Education: English education policy reform. Office of the Basic Education Commission, Ministry of Education. Available at http://old.drs.ac.th/ext/tch_data/tch_02.pdf. Office of the Higher Education Commission. (2015). Policy of upgrading English education standards of higher education institutions. Office of the Higher Education Commission, Ministry of Education. Available at http://www.mua.go.th/users/bhes/front_home/Data%20Bhes_2559/04052559.pdf.
- Onwuegbuzie A. J., & Leech N. L. (2004). Enhancing the interpretation of “significant” findings: The role of mixed methods research. *The Qualitative Report*, 9(4), 770-792.
- O’Sullivan, B. (2010). The City & Guilds Communicator Examination Linking Project: A brief overview with reflections on the process. In W. Martyniuk (Ed.), *Relating language examinations to the Common European Framework of Reference for Languages: Case studies and reflections on the use of the Council of Europe’s Draft Manual*, (pp. 33-49). Cambridge; Cambridge University Press.
- Qiu, Y. Q. (2006). A comparative study on the listening subtests of CET-4, TOEFL and IELTS. *Shanghai International Studies University*, 2006f2.
- Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal for Educational Research*.
- Salamoura, A. (2011). Maintaining alignment to the CEFR: the FCE case study. *Aligning Tests with the CEFR*, 80.

- Shao, H. (2006). An empirical study of washback from CET-4 on college English teaching and learning. *CELEA Journal*, 29(1), 54-59.
- Socialist Republic of Vietnam. 2008. "National Foreign Language 2020 Project." Article 1.1, Decision 1400. Available online at: <http://www.scribd.com/doc/237474972/National-Foreign-Language-2020>.
- Subong, P. E. (2005). *Statistics for research*. Sampaloc, Manila: Rex Book Store, Inc.
- Sun, Y. (2016). *Context, construct, and consequences: Washback of the College English Test in China* (Doctoral dissertation).
- Sun, Q., & Xie, N. (2015). A longitudinal study on the content validity of CET-4 Careful Reading. *Journal of Anshun College*, 17(6), 68-69.
- Van Teijlingen, E. (2014, December). Semi-structured interviews. In *PGR Workshop December*.
- Vietnamese Government. (2008). *Teaching and learning foreign languages in the national education system from 2008 to 2020*. Retrieved from http://www.chinhphu.vn/portal/page/portal/chinhphu/hethongvanban?class_id=1&page=18&mode=detail&document_id=78437
- Wang, B. T. (2015). *The Feasibility of Relating TEM to CEFR. --Taking Reading Comprehension Test for Example* (Doctoral dissertation, Yangzhou University).
- Wang, J. P. (2012) European Language Testing Standards and Foreign Language Testing reform in China. *Foreign Language Testing and Teaching*, (1) 50-56
- Weir, C. J. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing*, 22(3), 281-300.

- Westhoff, G. (2007). Challenges and opportunities of the CEFR for reimagining foreign language pedagogy. *The Modern Language Journal*, 91(4), 676-679.
- Wu, J. R., & Wu, R. Y. (2007). Using the CEFR in Taiwan: The perspective of a local examination board. *The Language Training and Testing Center Annual Report*, 56.
- Wu, J. (2012). 18 Policy Perspectives from Taiwan. *The Common European Framework of Reference: The globalisation of language education policy*, 23, 213.
- Wu, Z. (2014). *Washback effects of the reformed CET-4 on college English teaching and learning in China: Students' perspectives*. Indiana University of Pennsylvania.
- Wudthayagorn, J. (2018). Mapping the CU-TEP to the Common European Framework of Reference (CEFR). *LEARN Journal: Language Education and Acquisition Research Network*, 11(2), 163-180.
- Xu, L. M. (2016). A Study of Content Validity of Reading Comprehension in CET-4. *Campus English*, (29), 14-15.
- Yan, J. L. (2008). Perceptions on the development of China Standards of English based on Common European Framework of Reference for languages. *Journal of Southwest University of Science and Technology*. 25(4). 70-73.
- Yang, H., & Gui, S. (2007). On developing a unified Asian framework of English rating scales. *Foreign Languages in China*, 4(2), 34–37. 67.
- Yang, H., & Weir, C. (2001). Validation study of the national college English test *Foreign Language Education Press, Shanghai, China. successful*.
- Yao, J. (2015). Research on the content validity analysis of CET-4 reading comprehension. *Journal of Inner Mongolia Normal University: Philosophy & Social Science*. 44(3), 151-153.

- Zheng, Y., & Cheng, L. (2008). Test review: college English test (CET) in China. *Language Testing*, 25(3), 408-417.
- Zhong, Y. (2007). *A study on the validity of new CET-4* (Master's thesis, Xiangtan University).
- Zou, S., Zhang, W. X., & Kong, J. F. (2015). The Common European Framework of Reference in China: Research status and prospects of application. *Foreign Languages in China*, (3), 24-31.
- Zou, Y. Y. (2016). *Aligning National Matriculation English Test (NMET) with the Common European Framework of Reference (CEFR)* (Master's thesis, Guangdong University of Foreign Language Studies)





APPENDICES

APPENDIX A

GLOBAL SCALE OF THE CEFR

Proficient user	C2	Can understand with ease virtually everything heard or read. Can summaries information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.
Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics, which are familiar, or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

APPENDIX B

SALIENT CHARACTERISTICS OF CEFR LEVELS FOR INTERACTION AND PRODUCTION

Level	Salient Characteristics: (CEFR Section 3.6, simplified)	
Proficient user	C2	<p>It cannot be overemphasized that Level C2 is not intended to imply native speaker competence or even near native speaker competence. Both the original research and a project using CEFR descriptors to rate mother-tongue as well as foreign language competence (North 2002: CEFR Case Studies volume) showed the existence of ambilingual speakers well above the highest defined level (C2). Wilkins had identified a seventh level of “Ambilingual Proficiency” in his 1978 proposal for a European scale for unit-credit schemes.</p> <p>Level C2 is intended to characterize the degree of precision, appropriateness and ease with the language which typifies the speech of those who have been highly successful learners. Descriptors calibrated here include: convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices; has a good command of idiomatic expressions and colloquialisms with awareness of connotative level of meaning; backtrack and restructure around a difficulty so smoothly the interlocutor is hardly aware of it.</p>
	C1	<p>Level C1 is characterized by a broad range of language, which allows fluent, spontaneous communication, as illustrated by the following examples: Can express him/herself fluently and spontaneously, almost effortlessly. Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. There is little obvious searching for expressions or avoidance strategies; only a conceptually difficult subject can hinder a natural, smooth flow of language. The discourse skills appearing at B2+ are more evident at C1, with an emphasis on more fluency, for example: select a suitable phrase from a fluent repertoire of discourse functions to preface his remarks in order to get the floor, or to gain time and keep it whilst thinking; produce clear, smoothly flowing, well-structured speech, showing controlled use of organizational patterns, connectors and cohesive devices.</p>
Independent User	B2	<p>Level B2 represents a break with the content so far. Firstly there is a focus on effective argument: account for and sustain his opinions in discussion by providing relevant explanations, arguments and comments; explain a viewpoint on a topical issue giving the advantages and disadvantages of various options; develop an argument giving reasons in support of or against a particular point of view; take an active part in informal discussion in familiar contexts, commenting, putting point of view clearly, evaluating alternative proposals and making and responding to hypotheses. Secondly, at this level one can hold your own in social discourse: e.g. understand in detail what is said to him/her in the standard spoken language even in a noisy environment; initiate discourse, take his/her turn when appropriate and end conversation when he/she needs to, though he/she may not always do this elegantly; interact with a degree of fluency and spontaneity that makes regular interaction with native speakers</p>

		quite possible without imposing strain on either party. Finally, there is a new degree of language awareness: correct mistakes if they have led to misunderstandings; make a note of “favorite mistakes” and consciously monitor speech for it/them; generally correct slips and errors if he/she becomes conscious of them.
	B1	Level B1 reflects the Threshold Level specification and is perhaps most categorized by two features. The first feature is the ability to maintain interaction and get across what you want to, for example: generally follow the main points of extended discussion around him/her, provided speech is clearly articulated in standard dialect; express the main point he/she wants to make comprehensibly; keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production. The second feature is the ability to cope flexibly with problems in everyday life, for example cope with less routine situations on public transport; deal with most situations likely to arise when making travel arrangements through an agent or when actually travelling; enter unprepared into conversations on familiar topics.
Basic User	A2	Level A2 has the majority of descriptors stating social functions like use simple everyday polite forms of greeting and address; greet people, ask how they are and react to news; handle very short social exchanges; ask and answer questions about what they do at work and in free time; make and respond to invitations; discuss what to do, where to go and make arrangements to meet; make and accept offers. Here too are to be found descriptors on getting out and about: make simple transactions in shops, post offices or banks; get simple information about travel; use public transport: buses, trains, and taxis, ask for basic information, ask and give directions, and buy tickets; ask for and provide everyday goods and services.
	A1	Level A1 is the lowest level of generative language use-the point at which the learner can interact in a simple way, ask and answer simple questions about themselves, where they live, people they know, and things they have, initiate and respond to simple statements in areas of immediate need or on very familiar topics, rather than relying purely on a very finite rehearsed, lexically organized repertoire of situation-specific phrases.

APPENDIX C

SALIENT CHARACTERISTICS OF CEFR LEVELS FOR RECEPTION

Table A2. Salient Characteristics: Reception					
	Setting	Action	What is understood	Score	Restriction
C1	•Abstract and complex topics encountered in social, academic and professional life, whether or not they relate to own field/speciality	•Follow, maybe with a little difficulty		Films with a considerable degree of slang and idiomatic usage Poor quality, audially distorted public announcements	May occasionally need to: confirm details (with dictionary, from speaker) if outside field re-read difficult sections
		Understand	•Finer points of detail •Implied as well as stated opinions •A wide range of idiomatic expressions and colloquialisms •Register shifts •Implied attitudes and relationships	Lengthy, complex texts of various kinds Extended speech – lectures, discussions, debates even when not clearly structured Complex interactions between third parties in interaction and debate A wide range of recorded and broadcast texts, including some nonstandard Any correspondence	
B2+	•A wide range of familiar and unfamiliar topics encountered in social, academic And professional life	•Follow, maybe with a little difficulty		Animated conversation between native speakers	Standard, non-idiomatic: Adequate discourse structure Low background noise May occasionally need to confirm details (with dictionary, from speaker) if outside field if above conditions not met
		•Understand		Spoken language, live broadcast Specialised texts (highly specialised if within field)	
B2	•Reasonably familiar concrete and abstract topics related to field of interest/speciality	•Follow, maybe with a little difficulty	Much of what is said	Discussion around him/her by native speakers	Standard Clearly signposted/signalled with explicit markers If native speakers talking together modify language If can re-read difficult sections
		•Scan quickly	Relevance Whether closer study is worthwhile Specific details	Long and complex texts News items, articles and reports	
		•Understand (with a large degree of independence)	Main ideas Essentials/essential meaning Complex lines of argument Speaker/writer mood, tone etc.	Extended speech: lectures, talks, presentations, reports, discussions Propositionally and linguistically complex text Technical discussions; lengthy, complex instructions; details on conditions or warnings Most TV and current affairs programmes TV documentaries, interviews, talk shows, highly specialised sources Announcements and messages Most radio documentaries, recorded audio materials Correspondence	
B1	•Common everyday or job-related topics •Topics in his/her field of (personal) interest	•Follow, though not necessarily in detail	Line of argument in treatment of the issue	Argumentative text	Standard – (Familiar accent) Straightforward Clearly signposted/signalled with explicit markers
		•Scan	Desired information	Longer texts Different texts, different parts of a text	
		•Understand	Straightforward factual information content General message Main conclusions Specific details	Argumentative text Lectures and talks within own field Large part of many TV programmes: interviews, short lectures, news reports Majority of recorded and broadcast audio material	

Table A2. Salient Characteristics: Reception (continued)

	Table A2. Salient Characteristics: Reception (continued)				
	Setting	Action	What is understood	Score	Restriction
B1	<ul style="list-style-type: none"> Familiar topics regularly encountered in a school, work or leisure context Topics in his/her field of 	<ul style="list-style-type: none"> Follow, though not necessarily in detail 	<ul style="list-style-type: none"> Significant points 	<ul style="list-style-type: none"> Extended discussion around him/her Many films in which visuals and action carry much of the story line TV programmes: interviews, short lectures, news reports Straightforward newspaper articles 	<ul style="list-style-type: none"> Clear Standard Straightforward Relatively slow
		<ul style="list-style-type: none"> Understand with satisfactory comprehension 	<ul style="list-style-type: none"> Main points Relevant information 	<ul style="list-style-type: none"> Straightforward factual texts Short narratives Descriptions of events, feelings, wishes Detailed directions Short talks Radio news bulletins and simpler recorded materials Everyday written materials: letters, brochures, short official documents Simple technical information e.g. operating instructions 	
A2+	<ul style="list-style-type: none"> Familiar topics of a concrete type 	<ul style="list-style-type: none"> Identify 	<ul style="list-style-type: none"> Main points 	<ul style="list-style-type: none"> TV news items reporting events, accidents etc. in which visuals support the commentary 	<ul style="list-style-type: none"> Clearly and slowly articulated
		<ul style="list-style-type: none"> Understand enough to meet needs 		<ul style="list-style-type: none"> Basic types of standard letters, faxes (enquiries, orders, confirmations) Short texts with simpler, high frequency everyday and job-related language Regulations, e.g. safety 	<ul style="list-style-type: none"> Expressed in simple language
A2	<ul style="list-style-type: none"> Predictable everyday matters Areas of most immediate priority: basic personal, family, shopping, local area, employment 	<ul style="list-style-type: none"> Identify 	<ul style="list-style-type: none"> Specific, predictable information Topic of discussion Changes of topic An idea of the content 	<ul style="list-style-type: none"> Simpler everyday material: advertisements, menus, reference lists, timetables, brochures, letters Discussion around him/her Short newspaper articles describing events Factual TV news items 	<ul style="list-style-type: none"> Clearly and slowly articulated
		<ul style="list-style-type: none"> Understand 	<ul style="list-style-type: none"> Main point Essential information 	<ul style="list-style-type: none"> Simpler everyday material: advertisements, menus, reference lists, timetables, brochures, letters Discussion around him/her Short newspaper articles describing events Factual TV news items 	
A1	<ul style="list-style-type: none"> The most common everyday situations 	<ul style="list-style-type: none"> Identify 	<ul style="list-style-type: none"> Familiar words, phrases, names An idea of the content 	<ul style="list-style-type: none"> Simple notices Simpler informational material 	<ul style="list-style-type: none"> Very slow, carefully articulated, with long pauses to allow assimilation of meaning Familiar names, words and basic phrases A chance to re-read/get repetition
		<ul style="list-style-type: none"> Understand 	<ul style="list-style-type: none"> (Main point) 	<ul style="list-style-type: none"> Very short simple texts with visual support, a single phrase at a time: <ul style="list-style-type: none"> messages on postcards directions descriptions 	

APPENDIX D

SELF-ASSESSMENT FOR READING

C2	I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialized articles and literary works.
C1	I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialized articles and longer technical instructions, even when they do not relate to my field.
B2	I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.
B1	I can understand texts that consist mainly of high frequency every day or job- related language. I can understand the description of events, feelings and wishes in personal letters.
A2	I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.
A1	I can understand familiar names, words and very simple job sentences, for example on notices and posters or in catalogues.

APPENDIX E

CEFR SCALES FOR READING

OVERALL READING COMPREHENSION

OVERALL READING COMPREHENSION	
C2	Can understand and interpret critically virtually all forms of the written language including abstract, structurally complex, or highly colloquial literary and non-literary writings. Can understand a wide range of long and complex texts, appreciating subtle distinctions of style and implicit as well as explicit meaning.
C1	Can understand in detail lengthy, complex texts, whether or not they relate to his/her own area of specialty, provided he/she can reread difficult sections.
B2	Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using appropriate reference sources selectively. Has a broad active reading vocabulary, but may experience some difficulty with low frequency idioms.
B1	Can read straightforward factual texts on subjects related to his/her field and interest with a satisfactory level of comprehension.
A2	Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency every day or job-related language. Can understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared international vocabulary items.
A1	Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required.

READING FOR ORIENTATION

READING FOR ORIENTATION	
C2	As B2
C1	As B2
B2	Can scan quickly through long and complex texts, locating relevant details. Can quickly identify the content and relevance of news items, articles and reports on a wide range of professional topics, deciding whether closer study is worthwhile.
B1	Can scan longer texts in order to locate desired information, and gather information from different parts of a text, or from different texts in order to fulfil a specific task.
A2	Can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus, reference lists and timetables. Can locate specific information in lists and isolate the information required (e.g. use the 'Yellow Pages' to find a service or tradesman). Can understand everyday signs and notices: in public places, such as streets, restaurants, railway stations; in workplaces, such as directions, instructions, hazard warnings.
A1	Can recognize familiar names, words and very basic phrases on simple notices in the most common everyday situations.

READING FOR INFORMATION AND ARGUMENT

READING FOR INFORMATION AND ARGUMENT	
C2	As C1
C1	Can understand in detail a wide range of lengthy, complex texts likely to be encountered in social, professional or academic life, identifying finer points of detail including attitudes and implied as well as stated opinions.
B2	Can obtain information, ideas and opinions from highly specialised sources within his/her field. Can understand specialised articles outside his/her field, provided he/she can use a dictionary occasionally to confirm his/her interpretation of terminology. Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.
B1	Can identify the main conclusions in clearly signaled argumentative texts. Can recognize the line of argument in the treatment of the issue presented, though not necessarily in detail. Can recognize significant points in straightforward newspaper articles on familiar subjects.
A2	Can identify specific information in simpler written material he/she encounters such as letters, brochures and short newspaper articles describing events.
A1	Can get an idea of the content of simpler informational material and short simple descriptions, especially if there is visual support.

APPENDIX F

INFORMATION SHEET OF THE LINKING PROJECT

Dear teachers,

You are coordinately invited to take part in a linking project which aims to link CET-4 Reading Comprehension Tests to the CEFR. In order to achieve the research objective, you will be coordinately trained to familiarize the CEFR and then asked to analyze the content of CET-4 Reading Comprehension Test with the help of a tool (the Dutch CEFR Grid). The duration of this project is about six weeks in total. In the first two weeks, you will do self-study using related materials and documents sent by the researcher. Then, Familiarization workshop (around 3-5 hours) will be held on week 3 and week 4 to ensure that all of you are fully trained. On week 5, the same workshop will be held again with the same training activities again in the morning and with some introduction to the usage of the Grid. In the afternoon, you will be asked to analyze the content of the CET-4 Reading Comprehension Test (one test set). On the second day, you will be asked to do the analysis with one test set in the morning and one test set in the afternoon. During this period of time, some of your statements will be recorded. Your response to the questionnaire will be kept confidential. Your name will not appear in this research project. A brief report on the results of this study will be given to you once the study is complete. If you have any questions about the present project, please feel free to contact me. Estimated time for the linking project is shown in the following table.

Estimated time for the linking project

Stages	Dates	Length of time
Self-study	1-2 weeks	Two weeks
Familiarization (Training)	Week 3	3 hours
Specification (analyzing & linking)	Week 4	9 hours
Discussion	Week 5	3 hours

APPENDIX G

JUDGMENT SHEET FOR DESCRIPTOR-SORTING ACTIVITIES

Location: _____

Participant Name (optional): _____

Date: _____

Name of activities	No. of descriptors	No. of accuracy
Sorting activity 1 The CEFR level for Production & Interaction	6	
Sorting activity 2 Overall reading comprehension	6	
Reading for information and argument	6	
Reading for orientation	4	
Sorting activity 3 The CEFR level for Reception	6	



APPENDIX H

RECONSTRUCTING SELF-ASSESSMENT GRID FOR READING

C2	
C1	
B2	
B1	
A2	
A1	

APPENDIX I

PAPER-BASED Dutch CEFR GRID

Characteristics of Input Text		
Dimension	Description	
1. Text source <i>(taken from CEF Table 5 pages 48/9)</i>	Reading <input type="checkbox"/> Abstracts <input type="checkbox"/> Advertising materials <input type="checkbox"/> Advertising materials <input type="checkbox"/> Blackboard text <input type="checkbox"/> Broadcast & recorded spoken text <input type="checkbox"/> Brochures <input type="checkbox"/> Business letter <input type="checkbox"/> Computer screen text <input type="checkbox"/> Contracts <input type="checkbox"/> Dictionaries <input type="checkbox"/> Exercise materials <input type="checkbox"/> Guarantees <input type="checkbox"/> Instructional manuals <input type="checkbox"/> Instructional material <input type="checkbox"/> Job description <input type="checkbox"/> Journal articles <input type="checkbox"/> Junk mail <input type="checkbox"/> Labeling and packaging <input type="checkbox"/> Leaflets, graffiti <input type="checkbox"/> Life safety notices <input type="checkbox"/> Magazines <input type="checkbox"/> Menus <input type="checkbox"/> Newspapers <input type="checkbox"/> Notices, regulations <input type="checkbox"/> Novels <input type="checkbox"/> OP text <input type="checkbox"/> Personal letters <input type="checkbox"/> Programmes <input type="checkbox"/> Public announcements & notices <input type="checkbox"/> Recipes <input type="checkbox"/> Reference books <input type="checkbox"/> Regulations <input type="checkbox"/> Report, memorandum <input type="checkbox"/> Sacred texts, sermons, hymns <input type="checkbox"/> Sign posting <input type="checkbox"/> Teletext <input type="checkbox"/> Textbooks, readers <input type="checkbox"/> Tickets, timetables <input type="checkbox"/> Videotext <input type="checkbox"/> Visiting cards	
2. Authenticity	Input text appears to be <input type="checkbox"/> Genuine <input type="checkbox"/> Adapted/simplified <input type="checkbox"/> Pedagogic	
3. Discourse type	Discourse types	Examples (discourse types)
	<input type="checkbox"/> Mainly Descriptive impressionistic descriptions	e.g. sports commentaries, physical appearance, layout of room, house, landscape, places
	technical descriptions	e.g. presentation of a product
	<input type="checkbox"/> Mainly Narrative stories, jokes, anecdotes	
	reports	e.g. news reports, features, documentaries

	<input type="checkbox"/> Mainly Expository	definitions	brief definitions																
		explications	broader accounts of (especially) abstract phenomena e.g. lectures, talks																
		outlines	e.g. program listings on the radio, time-tables																
		summarize	e.g. an oral account of the plot of a book, summarizing minutes of a meeting																
		interpretations	e.g. describing a book, an article etc.																
	<input type="checkbox"/> Mainly Instructive	personal instructions	e.g. announcements, ads, propaganda, routine, commands																
	<input type="checkbox"/> Mainly Argumentative	comments, pros and cons of an issue, opinions	by any individual in any situation																
		formal argumentation	e.g. formal debate																
<p>4. Domain (Source: CEFR, page 45) NB. In many situations, more than one domain may be involved</p>	<p><input type="checkbox"/> Personal: Domain in which the person concerned lives as a private individual, centers on home life with family and friends and engages in individual practices such as reading for pleasure, keeping a personal diary, pursuing a special interest or hobby, etc.</p> <p><input type="checkbox"/> Public: Domain in which the person concerned acts as a member of the general public or of some organization and is engaged in transactions of various kinds for a variety of purposes.</p> <p><input type="checkbox"/> Occupational: Domain in which the person concerned is engaged in his or her job or profession.</p> <p><input type="checkbox"/> Educational: Domain in which the person concerned is engaged in organized learning, especially but not necessarily within an educational institution.</p>																		
<p>5. Topic (Source = CEFR page 52)</p>	<table style="width: 100%; border: none;"> <tbody> <tr> <td><input type="checkbox"/> Personal identification</td> <td><input type="checkbox"/> Shopping</td> </tr> <tr> <td><input type="checkbox"/> House and home, environment</td> <td><input type="checkbox"/> Food and drink</td> </tr> <tr> <td><input type="checkbox"/> Daily life</td> <td><input type="checkbox"/> Services</td> </tr> <tr> <td><input type="checkbox"/> Free time, entertainment</td> <td><input type="checkbox"/> Places</td> </tr> <tr> <td><input type="checkbox"/> Travel</td> <td><input type="checkbox"/> Language</td> </tr> <tr> <td><input type="checkbox"/> Regulations with other people</td> <td><input type="checkbox"/> Weather</td> </tr> <tr> <td><input type="checkbox"/> Health and body care</td> <td><input type="checkbox"/> Other (please specify_____)</td> </tr> <tr> <td><input type="checkbox"/> Education</td> <td></td> </tr> </tbody> </table>			<input type="checkbox"/> Personal identification	<input type="checkbox"/> Shopping	<input type="checkbox"/> House and home, environment	<input type="checkbox"/> Food and drink	<input type="checkbox"/> Daily life	<input type="checkbox"/> Services	<input type="checkbox"/> Free time, entertainment	<input type="checkbox"/> Places	<input type="checkbox"/> Travel	<input type="checkbox"/> Language	<input type="checkbox"/> Regulations with other people	<input type="checkbox"/> Weather	<input type="checkbox"/> Health and body care	<input type="checkbox"/> Other (please specify_____)	<input type="checkbox"/> Education	
<input type="checkbox"/> Personal identification	<input type="checkbox"/> Shopping																		
<input type="checkbox"/> House and home, environment	<input type="checkbox"/> Food and drink																		
<input type="checkbox"/> Daily life	<input type="checkbox"/> Services																		
<input type="checkbox"/> Free time, entertainment	<input type="checkbox"/> Places																		
<input type="checkbox"/> Travel	<input type="checkbox"/> Language																		
<input type="checkbox"/> Regulations with other people	<input type="checkbox"/> Weather																		
<input type="checkbox"/> Health and body care	<input type="checkbox"/> Other (please specify_____)																		
<input type="checkbox"/> Education																			

6. Nature of Content	<input type="checkbox"/> Only concrete content <input type="checkbox"/> Mostly concrete content <input type="checkbox"/> Fairly abstract content <input type="checkbox"/> Mainly abstract content
7. Text Length	_____
8. Vocabulary	<input type="checkbox"/> A1 <input type="checkbox"/> A2 <input type="checkbox"/> B1 <input type="checkbox"/> B2 <input type="checkbox"/> C1 <input type="checkbox"/> C2
9. Grammar	<input type="checkbox"/> A1 <input type="checkbox"/> A2 <input type="checkbox"/> B1 <input type="checkbox"/> B2 <input type="checkbox"/> C1 <input type="checkbox"/> C2
10. Text level estimated	<input type="checkbox"/> A1 <input type="checkbox"/> A2 <input type="checkbox"/> B1 <input type="checkbox"/> B2 <input type="checkbox"/> C1 <input type="checkbox"/> C2

Characteristics of Item:	
11. Item type <i>(from DIALANG Assessment Specifications)</i>	<p>Selected response</p> <input type="checkbox"/> 1. Multiple choice <input type="checkbox"/> 2. Banked multiple choice <input type="checkbox"/> 3. True False <input type="checkbox"/> 4. Multiple matching <input type="checkbox"/> 5. Sequencing / ordering jumbled text <input type="checkbox"/> 6. Citing
	<p>Short constructed response</p> <input type="checkbox"/> 7. Short answer <input type="checkbox"/> 8. Cloze (every nth) <input type="checkbox"/> 9. Gap filling (one word) <input type="checkbox"/> 10. C-Test <input type="checkbox"/> 11. Summary completion <input type="checkbox"/> 12. Information transfer <input type="checkbox"/> 13. Sentential response <input type="checkbox"/> 14. Justify by citing
	<p>Extended constructed response (creative, etc)</p> <input type="checkbox"/> 15. Essay <input type="checkbox"/> 16. Summary <input type="checkbox"/> 17. Report in own words <input type="checkbox"/> 18. Justify in own words <input type="checkbox"/> 19. Other Please specify: _____

12. Operations	<input type="checkbox"/> Understand the detail <input type="checkbox"/> Infer speaker's/ writer's attitude/mood/opinion <input type="checkbox"/> Generalize and draw the conclusion <input type="checkbox"/> Detect the topic and main idea/gist <input type="checkbox"/> Deduce the meaning of an unknown word <input type="checkbox"/> Decide the needed word from given options according to context <input type="checkbox"/> Distinguish the different forms of the same idea
13. Item level estimated	<input type="checkbox"/> A1 <input type="checkbox"/> A2 <input type="checkbox"/> B1 <input type="checkbox"/> B2 <input type="checkbox"/> C1 <input type="checkbox"/> C2



APPENDIX J

JUDGMENT SHEET FOR TEXTS AND ITEMS

Test name _____ Analyst _____

Time for Administration _____ Date _____

JUDGMENT SHEET FOR TEXTS			
Characteristics of texts	Judgment		
	Text 1	Text 2	Text 3
1. text source			
2. Authenticity			
3. Discourse type			
4. Domain			
5. Topic			
6. Nature of Content			
7. Text Length			
8. Vocabulary			
9. Grammar			
10. Texts level estimated			

JUDGMENT SHEET FOR ITEMS			
	Judgment		
Number of Items	14. Item type	15. Operations	16. Item level estimated
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			

APPENDIX K

TEST SPECIFICATION FOR CET-4

Test Purpose

College English Test Band 4(CET4) is the norm-referenced and criterion-referenced test which is designed to assess college students' English proficiency accurately and objectively. It focuses the test of overall ability including writing, listening, reading and translating skills and so on. CET4 aims to promote the implement of syllabus and level of English teaching.

Test Domain

Theoretical Construct:

To test students' ability to listen verbal information, process information and solve problems in reading text, translate Chinese into English accurately and frequently and express one's idea correctly and clearly in English.

Test Takers

Test takers are the undergraduates and graduates who have finished the college English Band 4 curriculum required by syllabus.

Test Method

Discrete-point test and integrative test are used in combination with emphasis on the latter.

Test organization and weighting

Structure	Content	Question types	Question number	Score percentage	Test time
Writing	writing	a short essay	1	15%	30 mins
Listening comprehension	short conversations	multiple choice	7	7%	25 mins
	long conversations	multiple choice	8	8%	
	passage comprehension	multiple choice	10	20%	
Reading comprehension	banked cloze	cloze	10	5%	40 mins
	long passage comprehension	matching	10	10%	
	reading in depth	multiple choice	10	20%	
Translation	from Chinese to English	passage translation	1	15%	30 mins
Total			57	100%	125 mins

Test Administration

Test Administration

The students should be informed of the procedures of test administration beforehand. Test instructions should be read to student before the test. When the test begins, the students should first finish the writing section on Answer Sheet 1 within 30 minutes. Later, when Listening Comprehension begins, students should write their answer on Answer Sheet 1. 30 minutes later, supervisors should collect the Answer Sheet 1 within 5minutes and students should finish the rest of the test on Answer Sheet 2.

Supervisors are expected to administer the test strictly in accordance with the stated procedures.

Measures should be taken to maintain a supportive testing environment throughout the test.

Specifications for Writing

Test purpose

The writing test is a proficiency test designed to assess students' English Writing ability.

Test domain

Theoretical Construct:

To test students' ability to produce different types of English writings in limited time as required in the syllabus.

Operations:

- a) Be able to produce writing at an efficient rate of speed and required number of words to suit the purpose
- b) Be able to produce correctly spelling words and use appropriate words and phrases
- c) Be able to use acceptable grammatical system, patterns and rules
- d) Be able to express a particular meaning in different grammatical forms
- e) Be able to use cohesive devices in written discourse
- f) Be able to use the rhetorical forms and conventions of written discourse
- g) Be able to develop the writing logically
- h) Be able to appropriately accomplish the communicative functions of written texts according to form.

Test content**Text types:**

Picture description, form completion tasks, letters, reports, advertisements, notes, argumentation, etc.

Topics:

History, customs, culture and society, school life, environment, technology, etc.

Focus:

Test task should focus on checking students' ability to express particular meaning with grammatical system and develop the writing logically and cohesively and coherently, with more emphasizes on the communicative functions of written texts.

Test task and item type

Type	Expected Response	Mode
writing	constructed	written

Time duration

The test should be completed within 30 minutes.

Scoring criteria

13-15 points. The passage is relevant to the given topic and expresses the idea clearly with coherent writing, and without linguistic errors, or only with few errors.

10-12 points. The passage is relevant to the given topic and expresses the idea clearly with relatively coherent writing. But the passage has a small number of errors.

7-9 points. The passage is basically relevant to the given topic. Some expressions are not clear enough and the discourse is barely coherent with a large number of linguistic errors, especially some of which are fatal.

4-6 points. The passage is basically relevant to the given topic with unclear expressions of ideas and a large amount of fatal errors. The discourse is poorly coherent

1-3 points. The passage is improperly arranged. The language in writing is disordered and most of sentences have errors, some of which are fatal.

0 point. Unanswered or only a few isolated words. Or the passage has no relevance to the given topic.

Specifications for Listening Comprehension

Test purpose

The listening test is a proficiency test designed to assess students' listening skills and ability.

Test domain

Theoretical Construct:

To test students' ability to get information, comprehend the given listening material and solve the problems with the help of the basic listening skills, such as scanning and skimming, etc.

Operations

- a) Be able to detect the specific information, such as time, space, participants, etc.
- b) Be able to grasp the key words
- c) Be able to infer the links and connections a of events and deduce causes and effects
- d) Be able to distinguish the literal information from implied information
- e) From the events and ideas described, predicts the outcomes and the behavior or attitude of people
- f) Be able to detect the main idea, supporting idea
- g) Be able to generalize
- h) Be able to dictate the specific words, phrases correctly

Test content

Text types:

Conversations, reports, dictation

Topics:

Daily conversation, news, reports, stories, advertisements, radio program, etc.

Focus:

Test task or items should focus on checking students' knowledge, comprehension, application of specified skills, with more emphasis on comprehension and application.

Text length

The whole test consists of three sections, including 35 questions. Section A has 8 short conversations and 2 long conversations. Section B includes 3 passages. Section C is a passage with 10 items.

Test task and item type

Type	Expected Response	Mode
Multiple-choice questions	Selected	written
dictations	constructed	written

Time duration

The test should be completed within 30 minutes.

Scoring criteria

A standard key is provided for the selected-response type task or items and the constructed-response type task or items.

Test development chart

Component	Skill (see operations)	Number	Test focus	Number of Items for Each Focus
conversation listening (15 items)	a	2	Knowledge	3
	b	1	Comprehension	12
	c	5		
	d	3		
	e	3		
	f	1		
passage listening (10 items)	a	6	Knowledge	6
	b	2	Comprehension	4
	c			
	d	2		
	e			
	f			
	g			
h	10			

Specifications for Reading Comprehension

Test purpose

The reading test is a proficiency test designed to assess students' reading ability and skills.

Test domain

Theoretical Construct:

To test students' ability to process information in reading texts, comprehend the target text and use different reading strategies for different purposes as required.

Operations

- a) Be able to scan text to locate specific information.
- b) Be able to search through text to establish which part is relevant to a specified need.
- c) Be able to deduce the meaning of an unknown word from context.
- d) Be able to distinguish the different forms of the same idea.
- e) Be able to understand the relations within sentences.
- f) Be able to search through text to evaluate the content in terms of previously received information.
- g) Be able to distinguish the literal information from implied information.
- h) Be able to comprehend the text globally and detect the topic and main idea.
- i) Be able to generalize and draw the conclusion.

Test content

Text types:

Cloze, essay, arguments, newspaper articles, reports.etc

Topics:

Education, economy, agriculture, medical, technology, etc

Focus:

Test task or items should focus on checking students' knowledge, comprehension, application of specified skills, with more emphasis on comprehension and application.

Text length

The test consists of three sections. Section A has a cloze with 10 items. Section B includes 10 short paragraphs, each has around 50 words. There are two texts of no more than 350 words.

Test task and item type

Type	Expected Response	Mode
Multiple-choice questions	Selected	written

Time duration

The test should be completed within 40 minutes.

Scoring criteria

A standard key is provided for this selected-response type test.

Test development chart

Component	Skill (see operations)	Number	Test Focus	Number of Items for Each Focus
Section A: Cloze (10 items)	g	10	Application	10
Section B: Matching (10 items)	a b d e h i	2 3 5	Comprehension	10
C: Multiple-choice questions (10 items)	a b	2 3	Knowledge	5
	c d e f g h i	1 2 1 1	Comprehension	5

Specifications for Translation

Test purpose

The translating test is a proficiency test designed to assess students' translating ability and skills.

Test domain

Theoretical Construct:

To test students' ability to process translation between Chinese and English according to certain contexts as required in the translation syllabus.

Operations

- a) Be able to use appropriate words and phrases.
- b) Be able to analyze the sentence structures.
- c) Be able to comprehend the text globally.
- d) Be able to grasp language transformation strategies.
- e) Be able to understand Bilingual culture.
- f) Be able to translate the source text into target text faithfully and smoothly.

Test content

Text types:

Short essays, etc

Topics:

History, stories, customs, education, social development, etc

Focus:

Test task or items should focus on checking students' knowledge, comprehension and application of translation skills, with more emphasis on application and translation.

Text length

There is a Chinese text of 140-160 words in the test.

Test task and item type

Type	Expected Response	Mode
translation	constructed	written

Time duration

The test should be completed within 30 minutes.

Scoring criteria

13-15 points. The translation is smooth and faithful to the original with appropriate wording. The whole translation does not have language errors, or just have a few errors.

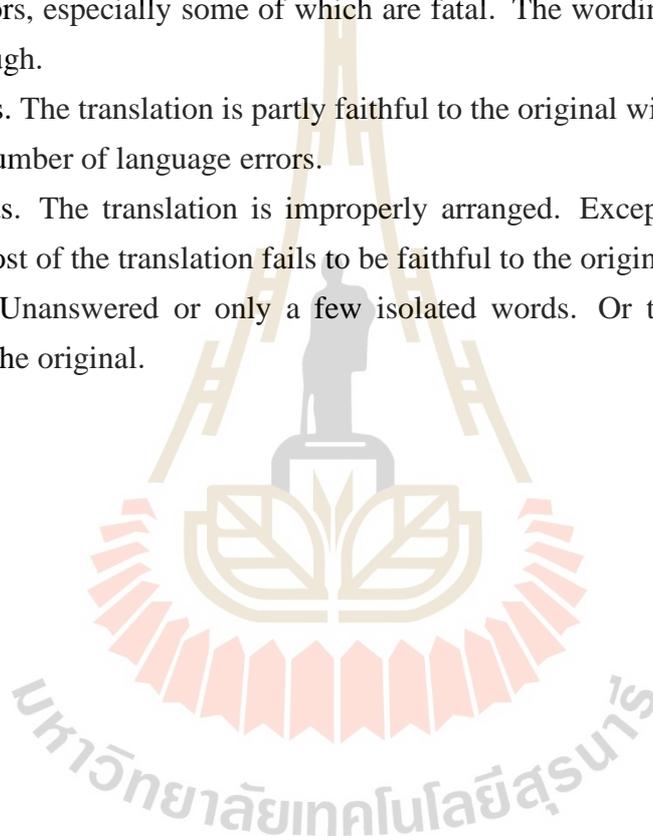
10-12 points. The translation is basically smooth and faithful to the original without fatal language errors.

7-9 points. The translation is barely faithful to the original with a number of language errors, especially some of which are fatal. The wording in translation is not accurate enough.

4-6 points. The translation is partly faithful to the original with inaccurate wording and a large number of language errors.

1-3 points. The translation is improperly arranged. Except for a few words or sentences, most of the translation fails to be faithful to the original.

0 point. Unanswered or only a few isolated words. Or the translation has no relevance to the original.



APPENDIX L

CET-4 Reading Comprehension Tests (2016-2017)

(December, 2016 (The first set))

Part III Reading Comprehension (40 minutes)

Part III Reading Comprehension (40 minutes)

Section A

Directions: *In this section, there is a passage with ten blanks. You are required to select one word for each blank from a list of choices given in a word bank following the passage. Read the passage through carefully before making your choices. Each choice in the bank is identified by a letter. Please mark the corresponding letter on **Answer Sheet 2** with a single line through the centre. You may not use any of the words in the bank more than once.*

The ocean is heating up. That's the conclusion of a new study that finds that Earth's oceans now (27) heat at twice the rate they did 18 years ago. Around half of ocean heat intake since 1865 has taken place since 1997, researchers report online in Nature Climate Change.

Warming waters are known to (28) coral bleaching (珊瑚白化) and they take up more space than cooler waters, raising sea (29). While the top of the ocean is studied, its depths are more difficult to (30) The researchers gathered 150 years of ocean temperature data in order to get better (31) of heat absorption from surface to seabed. They gathered together temperature readings collected by everything from a 19th century (32) of British naval ships to modern automated ocean probes. The extensive data sources, (33) with computer simulations (计算机模拟), created a timeline of ocean temperature changes, including cooling from volcanic outbreaks and warming from fossil fuel (34).

About 35 percent of the heat taken in by the oceans during the industrial era now resides at a (35) of more than 700 meters, the researchers found. They say they're unsure (36) whether the deep-sea warming canceled out warming at the sea's surface.

A. absorb	I. heights
B. combined	J. indifferent
C. contribute	K. level
D. depth	L. mixed
E. emission	M. picture
F. explore	N. unsure
G. explore	O. voyage
H. floor	

Section B

Directions: *In this section, you are going to read a passage with ten statements attached to it. Each statement contains information given in one of the paragraphs. Identify the paragraph from which the information is derived. You may choose a paragraph from more than once. Each paragraph is marked with a letter. Answer the questions by marking the corresponding letter on Answer Sheet 2.*

The Secret to Raising Smart Kids

- A) I first began to investigate the basis of human motivation-and how people persevere after setbacks-as a psychology graduate student at Yale University in the 1960s. Animal experiments by psychologists at the University of Pennsylvania had shown that after repeated failures, most animals conclude that a situation is hopeless and beyond their control. After such an experience an animal often remains passive even when it can effect change-a state they called learned helplessness.
- B) People can learn to be helpless, too. Why do some students give up when they encounter difficulty, whereas others who are no more skilled continue to strive and learn? One answer, I soon discovered, lay in people's beliefs about why they had failed.
- C) In particular, attributing poor performance to a lack of ability depresses motivation more than does the belief that lack of effort is to blame. When I told a group of school children who displayed helpless behavior that a lack of effort led to their mistakes in math, they learned to keep trying when the problems got tough. Another group of helpless children who were simply rewarded for their success on easier problems did not improve their ability to solve hard math problems. These

experiments indicated that a focus on effort can help resolve helplessness and generate success.

- D) Later, I developed a broader theory of what separates the two general classes of learners- helpless versus mastery- oriented. I realized these different types of students not only explain their failures differently, but they also hold different “theories” of intelligence. The helpless ones believe intelligence is a fixed characteristic: you have only a certain amount, and that's that. I call this a "fixed mind-set (思维模式)." Mistakes crack their self-confidence because they attribute errors to a lack of ability, which they feel powerless to change. They avoid challenges because challenges make mistakes more likely. The mastery- orient children, on the other hand, think intelligence is not fixed and can be developed through education and hard work. Such children believe challenges are energizing rather than intimidating (令人生畏);
- E) We validated these expectations in a study in which two other psychologists and I monitored 373 student for two years during the transition to junior high school, when the work gets more difficult and the grading more strict, to determine how their mind- sets might affect their math grades. At the beginning of seventh grade, we assessed the students' mind- sets by asking them to agree or disagree with statements such as "Your intelligence is something very basic about you that you can't really change." We then assessed their beliefs about other aspects of learning and looked to see what happened to their grades.
- F) As predicted, the students with a growth mind-set felt that learning was more important goal than getting good grades. In addition, they held hard work in high regard. They understood that even geniuses have to work hard. Confronted by a setback such as a disappointing test grade, students with a growth mind-set said they would study harder or try a different strategy. The students who held a fixed mind- set, however, were concerned about looking smart with less regard for learning. They had negative views of effort, believing that having to work hard was a sign of low ability. They thought that a person with talent or intelligence did not need to work hard to do well. Attributing a bad grade to their own lack of ability, those with a fixed mind-set said that would study less in the future, try never to take that subject again and consider cheating on future tests.
- G) Such different outlook had a dramatic impact on performance. At the start of junior high, the math achievement test scores of the students with a growth mind-set were comparable to those of students who displayed a fixed mind-set. But as the work

became more difficult, the students with a growth mind-set showed greater persistence. As a result, their math grades overtook those of the other students by the end of the first semester-and the gap between the two groups continued to widen during the two years we followed them.

- H) A fixed mind-set can also hinder communication and progress in the workplace and discourage or ignore constructive criticism and advice. Research shows that managers who have a fixed mind-set are less likely to seek or welcome feedback from their employees than are managers with a growth mind-set.
- I) How do we transmit a growth mind-set to our children? One way is by telling stories about achievements that result from hard work. For instance, talking about mathematical geniuses who were more or less born that way puts students in a fixed mind-set, but mathematicians who fell in love with math and developed amazing skills produce a growth mind-set.
- J) In addition, parents and teachers can help children by providing explicit instruction regarding the mind as a learning machine, I designed an eight-session workshop for 91 students whose math grades were declining in their first year of junior high. Forty-eight of the students received instruction in study skills only, whereas the others attended a combination of study skills sessions and classes in which they learned about the growth mind-set and how to apply it to schoolwork. In the growth mind-set classes, students read and discussed an article entitled “You Can Grow Your Brain.” They were taught that the brain is like a muscle that gets stronger with use and that learning prompts the brain to grow new connections. From such instruction, many students began to see themselves as agents of their own brain development. Despite being unaware that there were two types of instruction, teachers reported significant motivational changes in 27% of the children in the growth mind-set workshop as compared with only 9% of students in the control group.
- K) Research is converging (汇聚) on the conclusion that great accomplishment and even genius is typically the result of years of passion and dedication and not something that flows naturally from a gift.
36. The author's experiment shows that students with a fixed mind-set believe having to work hard is an indication of low ability.
37. Focusing on effort is effective in helping children overcome frustration and achieve success.

38. We can cultivate a growth mind-set in children by telling success stories that emphasize hard work love of learning.
39. Students' belief about the cause of their failure explains their attitude toward setbacks.
40. In the author's experiment, student with a growth mind-set showed greater perseverance in solving difficult math problems.
41. The author conducted an experiment to find out about the influence of students' mind-sets on math learning.
42. After failing again and again, most animals give up hope.
43. Informing students about the brain as a learning machine is a good strategy to enhance their motivation for learning.
44. People with a fixed mind-set believe that one's intelligence is unchangeable.
45. In the workplace, feedback may not be so welcome to managers with a fixed mind-set.
36. The author's experiment shows that students with a fixed mind-set believe having to work hard is an indication of low ability.

Section C

Directions: *There are 2 passages in this section. Each passage is followed by some questions or unfinished statements. For each of them there are four choices marked A),B),C) and D).You should decide on the best choice and mark the corresponding letter on Answer Sheet 2 with a single line through the center.*

Passage One

Questions 46 to 50 are based on the following passage.

"Sugar, alcohol and tobacco," economist Adam Smith once wrote, "are commodities which are nowhere necessities of life, which have become objects of almost universal consumption, and which are, therefore, extremely popular subjects of taxation."

Two and a half centuries on, most countries impose some sort of tax on alcohol and tobacco. With surging obesity levels putting increasing strain on public health systems, governments around the world have begun to toy with the idea of taxing sugar as well.

Whether such taxes work is a matter of debate. A preliminary review of Mexico's taxation found a fall in purchases of taxed drinks as well as a rise in sales if untaxed

and healthier drinks. By contrast, a Danish tax on foods high in fats was abandoned a year after its introduction, amid claims that consumers were avoiding it by crossing the border to Germany to satisfy their desire for cheaper, fattier fare.

The food industry has, in general, been firmly opposed to such direct government action. Nonetheless, the renewed focus on waistlines means that industry groups are under pressure to demonstrate their products are healthy as well as tasty.

Over the past three decades, the industry has made some efforts to improve the quality of its offerings. For example, some drink manufactures have cut the amount of sugar in their beverages.

Many of the reductions over the past 30 years have been achieved either by reducing the amount of sugar, salt or fat in a product, or by finding an alternative ingredient. More recently, however.

Some companies have been investing money in a more ambitious undertaking: learning how to adjust the fundamental make-up of the food they sell. For example, having salt on the outside, but none on the inside, reduces the salt content without changing the taste.

While reformulating recipes (配方) is one way to improve public health, it should be part of a multi-sided approach. The key is to remember that there is not just one solution. To deal with obesity, a mixture of approaches-including reformulation, taxation and adjusting portion sizes-will be needed. There is no silver bullet.

46. What did Adam Smith say about sugar, alcohol and tobacco?

- A.They were profitable to manufacture.
- B.They were in ever-increasing demand.
- C.They were subject to taxation almost everywhere.
- D.They were no longer considered necessities of life.

47. Why have many countries started to consider taxing sugar?

- A.They are under growing pressures to balance their national budgets.
- B.They find it ever harder to cope with sugar-induced health problems.
- C.They practice of taxing alcohol and tobacco has proved both popular and profitable.
- D.The sugar industry is overtaking alcohol and tobacco business in generating profits.

48. What do we learn about Danish taxation on fat-rich foods?
- A. It did not work out as well as was expected.
 - B. It gave rise to a lot of problems on the border.
 - C. It could not succeed without German cooperation.
 - D. It met with firm opposition from the food industry.
49. What is the more recent effort by food companies to make foods and drinks both healthy and tasty?
- A. Replacing sugar or salt with alternative ingredients.
 - B. Setting a limit on the amount of sugar or salt in their products.
 - C. Investing in research to find ways to adapt to consumers' needs.
 - D. Adjusting the physical composition of their products.
50. What does the author mean by saying, at the end of the passage, "There is no silver bullet"(Line 4, Para 7)?
- A. There is no single easy quick solution to the problem.
 - B. There is no hope of success without public cooperation.
 - C. There is on hurry in finding ways to solve the obesity problem.
 - D. There is no effective way to reduce people's sugar consumption.

Passage Two

Questions 51 to 55 are based on the following passage.

You may have heard some of the fashion industry horror stories: models eating tissues or cotton balls to hold off hunger, and models collapsing from hunger-induced heart attacks just seconds after they step off the runway.

Excessively skinny models have been a point of controversy for decades, and two researchers say a model's body mass should be a workplace health and safety issue. In an editorial released Monday in the American Journal of Public Health, Katherine Record and Bryn Austin made their case for government regulation of the fashion industry.

The average international runway model has a body mass index (BMI) under 16—low enough to indicate starvation by the World Health Organization's standard. And Record and Austin are worried not just about the models themselves, but about the vast number of girls and women their images influence.

"Especially girls and teens", says Record. "Seventy percent of girls aged 10 to 18 report that they define perfect body image based on what they see in magazines." That's especially worrying, she says, given that anorexia (厌食症) results in more

deaths than does any other mental illness, according to the National Institute of Mental Health.

It's commonly known that certain diseases are linked with occupations like lung disease in coal miners. Professional fashion models are particularly vulnerable to eating disorders resulting from occupational demands to maintain extreme thinness.

Record's suggestion is to prohibit agents from hiring models with a BMI below 18.

In April, France passed a law setting lower limits for a model's weight. Agents and fashion houses who hire models with BMI under 18 could pay \$82,000 in fines and spend up to 6 months in jail. Regulating the fashion industry in the United States won't be easy, Record says. But with the new rules in France, U.S. support could make a difference. "A designer can't survive without participating in Paris Fashion Week", she says, adding, "Our argument is that the same would be true of New York Fashion Week."

51. What do Record and Austin say about fashion models' body mass?
- A. It has caused needless controversy.
 - B. It is focus of the modeling business.
 - C. It is but a matter of personal taste.
 - D. It affects models' health and safety.
52. What are Record and Austin advocating in the Monday editorial?
- A. A change in the public's view of female beauty.
 - B. Government legislation about models' weight.
 - C. Elimination of forced weight loss by models.
 - D. Prohibition of models eating non-food stuff.
53. Why are Record and Austin worried about the low body mass index of models?
- A. It contributes to many mental illnesses.
 - B. It defines the future of the fashion industry.
 - C. It has great influence on numerous girls and women.
 - D. It keeps many otherwise qualified women off the runway.
54. What do we learn about France's fashion industry?
- A. It has difficulty hiring models.
 - B. It has now a new law to follow.
 - C. It allows girls under 18 on the runway.
 - D. It has overtaken that of the United States.

55. What does Record expect of New York Fashion Week?
- A.It will create a completely new set of rules.
 - B.It will do better than Paris Fashion Week.
 - C.It will differ from Paris Fashion Week.
 - D.It will have models with a higher BMI.



CET-4 Reading Comprehension Tests

(June, 2017 (The second set))

Part III Reading Comprehension (40 minutes)

Section A

Directions: *In this section, there is a passage with ten blanks. You are required to select one word for each blank from a list of choices given in a word bank following the passage. Read the passage through carefully before making your choices. Each choice in the bank is identified by a letter. Please mark the corresponding letter on **Answer Sheet 2** with a single line through the centre. You may not use any of the words in the bank more than once.*

America's Internet is faster than ever before, but people still complain about their Internet being too slow.

New York's Attorney General's office ___26___ an investigation in the fall into whether or not Verizon, Cablevision and Time Warner are delivering broadband that's as fast as the providers ___27___ it is. Earlier this month, the office asked for the public's help to measure their speed results, saying consumers ___28___ to get the speeds they were promised.

"Too many of us maybe paying for one thing, and getting another," the Attorney General said.

If the investigation uncovers anything, it wouldn't be the first time a telecom provider got into ___29___ over the broad speeds it promised and delivered customers.

Back in June, the Federal Communications Commission fined AT&T \$100 million over ___30___ that the carrier secretly reduced wireless speeds after customers consumed a certain amount of ___31___.

Even when they stay on the right side of the law, Internet providers arouse customers' anger over bandwidth speed and cost. Just this week, an investigation found that media and telecom giant Comcast is the most ___32___ provider. Over 10 months, Comcast received nearly 12,000 customer complaints, many ___33___ to its monthly data cap and overage 超过额度的 charges.

Some Americans are getting so ___34___ with Internet providers they're just giving up. A recent study found that the number of Americans with high-speed Internet at

home today __35__fell during the last two years, and 15% of people now consider themselves to be "cord-cutters."

A. accusations	F. data	K. relating
B. actually	G. deserved	L. times
C. claim	H. frustrated	M. trouble
D. communicating	I. hated	N. usually
E. complain	J. launched	O. worried

Section B

Directions: *In this section, you are going to read a passage with ten statements attached to it. Each statement contains information given in one of the paragraphs. Identify the paragraph from which the information is derived. You may choose a paragraph from more than once. Each paragraph is marked with a letter. Answer the questions by marking the corresponding letter on Answer Sheet 2.*

From Accountant to Yogi: Making a Radical Career Change

- A) At some point, almost all of us will experience a period of radical professional change. Some of us will seek it out; for others it will feel like an unwelcome intrusion into otherwise stable careers. Either way, we have choices about how we respond to it when it comes.
- B) We recently caught up with yoga entrepreneur Leah Zaccaria, who put herself through the fire of change to completely reinvent herself. In her search to live a life of purpose, Leah left her high-paying accounting job, her husband, and her home. In the process, she built a radically new life and career. Since then, she has founded two yoga studios, met a new life partner, and formed a new community of people. Even if your personal reinvention is less drastic, we think there are lessons from her experience that apply.
- C) Where do the seeds of change come from? The Native American Indians have a saying: "Pay attention to the whispers so you won't have to hear the screams." Often the best ideas for big changes come from unexpected places—it's just a matter of tuning in. Great leaders recognize the weak signals or slight signs that point to big changes to come. Leah reflects on a time she listened to the whispers: "About the time my daughter was five years old, I started having a sense that 'this isn't right.'" She then realized that her life no longer matched her vision for it.

- D) Up until that point, Leah had followed traditional measures of success. After graduating with a degree in business and accounting, she joined a public accounting firm, married, bought a house, put lots of stuff in it, and had a baby. "I did what everybody else thought looked successful," she says. Leah easily could have fallen into a trap of feeling content; instead, her energy sparked a period of experimentation and renewal.
- E) Feeling the need to change, Leah started playing with future possibilities by exploring her interests and developing new capabilities. First trying physical exercise and dieting, she lost some weight and discovered an inner strength. "I felt powerful because I broke through my own limitations," she recalls.
- F) However, it was another interest that led Leah to radically reinvent herself. "I remember sitting on a bench with my aunt at a yoga studio," she said, "and having a moment of clarity right then and there: Yoga is saving my life. Yoga is waking me up. I'm not happy and I want to change and I'm done with this." In that moment of clarity Leah made an important leap, conquering her inner resistance to change and making a firm commitment to take bigger steps.
- G) Creating the future you want is a lot easier if you are ready to exploit the opportunities that come your way. When Leah made the commitment to change, she primed herself to new opportunities she may otherwise have overlooked. She recalls:
- H) One day a man I worked with, Ryan, who had his office next to mine, said, "Leah, let's go look at this space on Queen Anne." He knew my love for yoga and had seen a space close to where he lived that he thought might be good to serve as a yoga studio. As soon as I saw the location, I knew this was it. Of course I was scared, yet I had this strong sense of "I have to do this." Only a few months later Leah opened her first yoga studio, but success was not instant.
- I) Creating the future takes time. That's why leaders continue to manage the present while building toward the big changes of the future. When it's time to make the leap, they take action and immediately drop what's no longer serving their purpose. Initially Leah stayed with her accounting job while starting up the yoga studio to make it all work.
- J) Soon after, she knew she had to make a bold move to fully commit to her new future. Within two years, Leah shed the safety of her accounting job and made the switch complete. Such drastic change is not easy.

- K) Steering through change and facing obstacles brings us face to face with our fears. Leah reflects on one incident that triggered her fears, when her investors threatened to shut her down: "I was probably up against the most fear I've ever had," she says. "I had spent two years cultivating this community, and it had become successful very fast, but within six months I was facing the prospect of losing it all. "
- L) She connected with her sense of purpose and dug deep, cultivating a tremendous sense of strength. "I was feeling so intentional and strong that I wasn't going to let fear just take over. I was thinking, 'OK, guys, if you want to try to shut me down, shut me down.' And I knew it was a negotiation scheme, so I was able to say to myself, 'This is not real.'" By naming her fears and facing them head- on, Leah gained confidence. For most of us, letting go of the safety and security of the past gives us great fear. Calling out our fears explicitly, as Leah did, can help us act decisively.
- M) The cycle of renewal never ends. Leah's growth spurred her to open her second studio—and it wasn't for the money.
- N) I have no desire to make millions of dollars. It's not about that; it's about growth for me. Honestly, I didn't need to open a second studio. I was making as much money as I was as an accountant. But I know if you don't grow, you stand still, and that doesn't work for me.
- O) Consider the current moment in your own life, your team or your organization. Where are you in the cycle of renewal: Are you actively preserving the present, or selectively forgetting the past, or boldly creating the future? What advice would Leah give you to move you ahead on your journey? Once we're on the path of growth, we can continually move through the seasons of transformation and renewal.
36. Readiness to take advantage of new opportunities will make it easier to create one's desired future.
37. By conventional standards, Leah was a typical successful woman before she changed her career.
38. Leah gained confidence by laying out her fears and confronting them directly.
39. In search of a meaningful life, Leah gave up what she had and set up her own yoga studios.
40. Leah's interest in yoga prompted her to make a firm decision to reshape her life.
41. Small signs may indicate great changes to come and therefore merit attention.

42. Leah's first yoga studio was by no means an immediate success.
43. Some people regard professional change as an unpleasant experience that disturbs their stable careers.
44. The worst fear Leah ever had was the prospect of losing her yoga business.
45. As she explored new interests and developed new potentials. Leah felt powerful internally.

Section C

Directions: *There are 2 passages in this section. Each passage is followed by some questions or unfinished statements. For each of them there are four choices marked A),B),C) and D).You should decide on the best choice and mark the corresponding letter on Answer Sheet 2 with a single line through the center.*

Passage One

Questions 46 to 50 are based on the following passage.

Urbanization-migration away from the suburbs to the city center-will be the biggest real-estate trend in 2015, according to a new report.

The report says America's urbanization will continue to be the most significant issue affecting the industry, as cities across the country imitate the walkability and trans-oriented development making cities like New York and San Francisco so successful.

As smaller cities copy the model of these "24-hour cities," more affordable versions of these places will be created. The report refers to this as the coming of the "18-hour city," and uses the term to refer to cities like Houston, Austin, Charlotte, and Nashville, which are "positioning themselves as highly competitive, in terms of livability, employment offerings, and recreational and cultural facilities."

Another trend that looks significant in 2015 is that America's largest population group, *Millennials* (千禧一代), will continue to put off buying a house. Apartments will retain their appeal for a while for Millennials, haunted by what happened to home-owning parents.

This trend will continue into the 2020s, the report projects. After that, survey respondents disagree over whether this generation will follow in their parents' footsteps, moving to the suburbs to raise families, or will choose to remain in the city center.

Another issue affecting real estate in the coming year will be America's failing infrastructure. Most roads, bridges, transit, water systems, the electric grid, and

communications networks were installed 50 to 100 years ago, and they are largely taken for granted until they fail.

The report's writers state that America's failure to invest in infrastructure impacts not only the health of the real-estate market, but also our ability to remain globally competitive.

Apart from the specific trends highlighted above, which cause some investors to worry, the report portrays an overall optimism borne by the recent healthy real-estate "upcycle" and improving economy. Seventy-four percent of the respondents surveyed report a "good to excellent" expectation of real-estate profitability in 2015. While excessive optimism can promote bad investment patterns, resulting in a real-estate "bubble," the report's writers downplay that potential outcome in that it has not yet occurred.

46. According to the new report, real estate development in 2015 will witness _____.
- an accelerating speed
 - a shift to city centers
 - a new focus on small cities
 - an ever-increasing demand
47. What characterizes "24-hour cities" like New York?
- People can live without private cars.
 - People are generally more competitive.
 - People can enjoy services around the clock.
 - People are in harmony with the environment.
48. Why are Millennials reluctant to buy a house?
- They can only afford small apartments.
 - The house prices are currently too high.
 - Their parents' bad experience still haunts them.
 - They feel attached to the suburban environment.
49. What might hinder real estate development in the U.S.?
- The continuing economic recession in the country.
 - The lack of confidence on the part of investors.
 - The fierce global competition.
 - The worsening infrastructure.
50. How do most of the respondents in the survey feel about the U.S. real-estate market in 2015?
- Pessimistic.

- B. Hopeful.
- C. Cautious.
- D. Uncertain.

Passage Two

Questions 51 to 55 are based on the following passage.

The brain is a seemingly endless library, whose shelves house our most precious memories as well as our lifetime's knowledge. But is there a point where it reaches capacity?

The answer is no, because brains are more sophisticated than that. Instead of just crowding in, old information is sometimes pushed out of the brain for new memories to form.

Previous behavioral studies have shown that learning new information can lead to forgetting. But in a new study, researchers demonstrated for the first time how this effect occurs in the brain.

In daily life, forgetting actually has clear advantages. Imagine, for instance, that you lost your bank card. The new card you receive will come with a new personal identification number (PIN). Each time you remember the new PIN, you gradually forget the old one. This process improves access to relevant information, without old memories interfering.

And most of us may sometimes feel the frustration of having old memories interfere with new, relevant memories. Consider trying to remember where you parked your car in the same car park you were at a week earlier. This type of memory (where you are trying to remember new, but similar information) is particularly vulnerable to interference.

When we acquire new information, the brain automatically tries to incorporate (合并) it within existing information by forming associations. And when we retrieve (检索) information, both the desired and associated but irrelevant information is recalled.

The majority of previous research has focused on how we learn and remember new information. But current studies are beginning to place greater emphasis on the conditions under which we forget, as its importance begins to be more appreciated.

A very small number of people are able to remember almost every detail of their life. While it may sound like an advantage to many, people with this rare condition often find their unusual ability burdensome.

In a sense, forgetting is our brain's way of sorting memories, so the most relevant memories are ready for retrieval. Normal forgetting may even be a safety mechanism to ensure our brain doesn't become too full.

51. What have past behavioral studies found about our brain?
- A. Its capacity actually knows no limits.
 - B. It grows sophisticated with practice.
 - C. It keeps our most precious memories until life's end.
 - D. New information learned pushes old information out.
52. What is the benefit of forgetting?
- A. It frees us from painful memories.
 - B. It helps slow down our aging process.
 - C. It facilitates our access to relevant information.
 - D. It prevents old information from forming associations.
53. What is the emphasis of current studies of memory?
- A. When people tend to forget.
 - B. What contributes to forgetting.
 - C. How new technology hinders memory capacity.
 - D. Why learning and forgetting are complementary.
54. What do people find about their rare ability to remember every detail of their life?
- A. It adds to the burden of their memory.
 - B. It makes their life more complicated.
 - C. It contributes to their success in life.
 - D. It constitutes a rare object of envy.
55. What does the passage say about forgetting?
- A. It can enlarge our brain capacity.
 - B. It helps get rid of negative memories.
 - C. It is a way of organizing our memories.
 - D. It should not cause any alarm in any way.

CET-4 Reading Comprehension Tests
(December, 2017 (The third set))

Part III Reading Comprehension (40 minutes)

Section A

Directions: *In this section, there is a passage with ten blanks. You are required to select one word for each blank from a list of choices given in a word bank following the passage. Read the passage through carefully before making your choices. Each choice in the bank is identified by a letter. Please mark the corresponding letter on **Answer Sheet 2** with a single line through the center. You may not use any of the words in the bank more than once.*

We all know there exists a great void (空白) in the public educational system when it comes to ___26___ to STEM (Science, Technology, Engineering and Mathematics) courses. One educator named Dori Roberts decided to do something to change this system. Dori taught high school engineering for 11 years. She noticed there was a real void in quality STEM education at all ___27___ of the public educational system. She said, "I started Engineering For Kids (EFK) after noticing a real lack of math, science and engineering programs to ___28___ my own kids in."

She decided to start an afterschool program where children ___29___ in STEM-based competitions. The club grew quickly and when it reached 180 members and the kids in the program won several state ___30___, she decided to devote all her time to cultivating and ___31___ it. The global business EFK was born.

Dori began operating EFK out of her Virginia home, which she then expanded to ___32___ recreation centers. Today, the EFK program ___33___ over 144 branches in 32 states within the United States and in 21 countries. Sales have doubled from \$5 million in 2014 to \$10 million in 2015, with 25 new branches planned for 2016. The EFK website states, "Our nation is not ___34___ enough engineers. Our philosophy is to inspire kids at a young age to understand that engineering is a great ___35___."

- | | |
|------------------|-----------------|
| A) attracted | I) feeding |
| B) career | J) graduating |
| C) championships | K) interest |
| D) degrees | L) levels |
| E) developing | M) local |
| F) enroll | N) operates |
| G) exposure | O) participated |
| H) feasible | |

Section B

Directions: *In this section, you are going to read a passage with ten statements attached to it. Each statement contains information given in one of the paragraphs. Identify the paragraph from which the information is derived. You may choose a paragraph more than once. Each paragraph is marked with a letter. Answer the questions by marking the corresponding letter on Answer Sheet 2.*

Why aren't you curious about what happened?

- A) "You suspended Ray Rice after our video," a reporter from TMZ challenged National Football League Commissioner Roger Goodell the other day. "Why didn't you have the curiosity to go to the casino (赌场) yourself?" The implication of the question is that a more curious commissioner would have found a way to get the tape.
- B) The accusation of incuriosity is one that we hear often, carrying the suggestion that there is something wrong with not wanting to search out the truth. "I have been bothered for a long time about the curious lack of curiosity," said a Democratic member of the New Jersey legislature back in July, referring to an insufficiently inquiring attitude on the part of an assistant to New Jersey Governor Chris Christie who chose not to ask hard questions about the George Washington Bridge traffic scandal. "Isn't the mainstream media the least bit curious about what happened?" wrote conservative writer Jennifer Rubin earlier this year, referring to the attack on Americans in Benghazi, Libya.

- C) The implication, in each case, is that curiosity is a good thing, and a lack of curiosity is a problem. Are such accusations simply efforts to score political points for one's party? Or is there something of particular value about curiosity in and of itself?
- D) The journalist Ian Leslie, in his new and enjoyable book *Curious: The Desire to Know and Why Your Future Depends on It*, insists that the answer to that last question is 'Yes'. Leslie argues that curiosity is a much-overlooked human virtue, crucial to our success, and that we are losing it.
- E) We are suffering, he writes, from a "serendipity deficit." The word "serendipity" was coined by Horace Walpole in an 1854 letter, from a tale of three princes who "were always making discoveries, by accident, of things they were not in search of." Leslie worries that the rise of the Internet, among other social and technological changes, has reduced our appetite for aimless adventures. No longer have we the inclination to let ourselves wander through fields of knowledges, ready to be surprised. Instead, we seek only the information we want.
- F) Why is this a problem? Because without curiosity we will lose the spirit of innovation and entrepreneurship. We will see unimaginative governments and dying corporations make disastrous decisions. We will lose a vital part of what has made humanity as a whole so successful as a species.
- G) Leslie presents considerable evidence for the proposition that the society as a whole is growing less curious. In the U.S. and Europe, for example, the rise of the Internet has led to a declining consumption of news from outside the reader's borders. But not everything is to be blamed on technology. The decline in interest in literary fiction is also one of the causes identified by Leslie. Reading literary fiction, he says, makes us more curious.
- H) Moreover, in order to be curious, "you have to be aware of a gap in your knowledge in the first place." Although Leslie perhaps paints a bit broadly in contending that most of us are unaware of how much we don't know, he's surely right to point out that the problem is growing: "Google can give us the powerful illusion that a question have definite answers."
- I) Indeed, Google, for which Leslie express admiration, is also his frequent whipping boy (替罪羊). He quotes Google co-founder Larry Page to the effect that the "perfect search engine" will "understand exactly what I mean and give me back exactly what

I want.” Elsewhere in the book, Leslie writes: “Google aims to save you from the thirst of curiosity altogether.”

- J) Somewhat nostalgically (怀旧地), he quotes John Maynard Keynes’s justly famous words of praise to the bookstore: “One should enter it vaguely, almost in a dream and allow what is there freely to attract and influence the eye. To walk the rounds of the bookshops, dipping in as curiosity dictates, should be an afternoon’s entertainment.” If only!
- K) Citing the work of psychologists and cognitive (认知的) scientists, Leslie criticizes the received wisdom that academic success is the result of a combination of intellectual talent and hard work. Curiosity, he argues, is the third key factor—and a difficult one to preserve. If not cultivated, it will not survive: “Childhood curiosity is a collaboration between child and adult. The surest way to kill it is to leave it alone.”
- L) School education, he warns, is often conducted in a way that makes children incurious. Children of educated an upper-middle-class parents turn out to be far more curious, even at early ages, than children of working class and lower class families. That lack of curiosity produces a relative lack of knowledge, and the lack of knowledge is difficult if not impossible to compensate for later on.
- M) Although Leslie’s book isn’t about politics, he doesn’t entirely shy away from the problem. Political leaders, like leader of other organizations, should be curious. They should ask questions at crucial moments. There are serious consequences, he warns, in not wanting to know.
- N) He presents as an example the failure of the George W. Bush administration to prepare properly for the after-effects of the invasion of Iraq. According to Leslie, those who ridiculed former Defense Secretary Donald Rumsfeld for his 2002 remark that we have to be wary of the “unknown unknowns” were mistaken. Rumsfeld’s idea, Leslie writes, “wasn’t absurd—it was smart.” He adds, “The tragedy is that he didn’t follow his own advice.”
- O) All of which brings us back to Goodell and the Christie case and Benghazi. Each critic in those examples is charging, in a different way, that someone in authority is intentionally being incurious. I leave it to the reader’s political preference to decide which, if any, charges should stick. But let’s be careful about demanding curiosity about the other side’s weaknesses and remaining determinedly incurious about our

own. We should be delighted to pursue knowledge for its own sake—even when what we find out is something we didn't particularly want to know.

36. To be curious, we need to realize first of all that there are many things we don't know.
37. According to Leslie, curiosity is essential to one's success.
38. We should feel happy when we pursue knowledge for knowledge's sake.
39. Political leaders' lack of curiosity will result in bad consequences.
40. There are often accusations about politicians' and the media's lack of curiosity to find out the truth.
41. The less curious a child is, the less knowledge the child may turn out to have.
42. It is widely accepted that academic accomplishment lies in both intelligence and diligence.
43. Visiting a bookshop as curiosity leads us can be a good way to entertain ourselves.
44. Both the rise of the Internet and reduced appetite for literary fiction contribute to people's declining curiosity.
45. Mankind wouldn't be so innovative without curiosity.

Section C

Directions: *There are 2 passages in this section. Each passage is followed by some questions or unfinished statements. For each of them there are four choices marked A), B), C) and D). You should decide on the best choice and mark the corresponding letter on Answer Sheet 2 with a single line through the center.*

Passage One

Questions 46 and 50 are based on the following passage.

Aging happens to all of us, and is generally thought of as a natural part of life. It would seem silly to call such a thing a “disease.” On the other hand, scientists are increasingly learning that aging and biological age are two different things, and that the former is a key risk factor for conditions such as heart disease, cancer and many more.

In that light, aging itself might be seen as something treatable, the way you would treat high blood pressure or a vitamin deficiency.

Biophysicist Alex Zhavoronkov believes that aging should be considered a disease. He said that describing aging as a disease creates incentives to develop treatments.

“It unties the hands of the pharmaceutical (制药的) industry so that they can begin treating the disease and not just the side effects,” he said.

“Right now, people think of aging as natural and something you can’t control,” he said. “IN academic circles, people take aging research as just an interest area where they can try to develop interventions. The medical community also takes aging for granted, and can do nothing about it except keep people within a certain health range.”

But if aging were recognized as a disease, he said, “It would attract funding and change the way we do health care. What matters is understanding that aging is curable.” “It was always known that the body accumulates damage,” he added. “The only way to cure aging is to find ways to repair that damage. I think of it as preventive medicine for age-related conditions.”

Leonard Hayflick, a professor at the University of California, San Francisco, said the idea that aging can be cured implies the human lifespan can be increased, which some researchers suggest is possible. Hayflick is not among them.

“There’re many people who recover from cancer, stroke, or heart disease. But they continue to age, because aging is separate from their disease,” Hayflick said. “Even if those causes of death were eliminated, life expectancy would still not go much beyond 92 years.”

46. What do people generally believe about aging?

- A) It should cause not alarm whatsoever.
- B) They just cannot do anything about it.
- C) It should be regarded as a kind of disease.
- D) They can delay it with advances in science.

47. How do many scientists view aging now?

- A) It might be prevented and treated.
- C) It results from a vitamin deficiency.
- B) It can be as risky as heart disease.
- D) It is an irreversible biological process.

48. What does Alex Zhavoronkov think of “describing aging as a disease”?
- A) It will prompt people to take aging more seriously.
 - B) It will greatly help reduce the side effects of aging.
 - C) It will free pharmacists from the conventional beliefs about aging.
 - D) It will motivate doctors and pharmacists to find ways to treat aging.
49. What do we learn about the medical community?
- A) They now have a strong interest in research on aging.
 - B) They differ from the academic circles in their view on aging.
 - C) They can contribute to people’s health only to a limited extent.
 - D) They have ways to intervene in people’s aging process.
50. What does Professor Leonard Hayflick believe?
- A) The human lifespan cannot be prolonged.
 - B) Aging is hardly separable from disease.
 - C) Few people can live up to the age of 92.
 - D) Heart disease is the major cause of aging.

Passage Two

Questions 51 to 55 are based on the following passage.

Female applicants to postdoctoral positions in geosciences were nearly half as likely to receive excellent letters of recommendation, compared with their male counterparts. Christopher Intagliata reports.

As in many other fields, gender bias is widespread in the sciences. Men score higher starting salaries, have more mentoring (指导), and have better odds of being hired. Studies show they’re also perceived as more competent than women in STEM (Science, Technology, Engineering, and Mathematics) fields. And new research reveals that men are more likely to receive excellent letters of recommendation, too.

“Say, you know, this is the best student I’ve ever had,” says Kuheli Dutt, a social scientist and diversity officer at Columbia University’s Lamont campus. “Compare those excellent letters with a merely good letter: ‘The candidate was productive, or intelligent, or a solid scientist or something that’s clearly solid praise,’ but nothing that singles out the candidate as exceptional or one of a kind.”

Dutt and her colleagues studied more than 1,200 letters of recommendation for postdoctoral positions in geoscience. They were all edited for gender and other

identifying information, so Dutt and her team could assign them a score without knowing the gender of the student. They found that female applicants were only half as likely to get outstanding letters, compared with their male counterparts. That includes letters of recommendation from all over the world, and written by, yes, men and women. The findings are in the journal *Nature Geoscience*.

Dutt says they were not able to evaluate the actual scientific qualifications of the applicants using the data in the files. But she says the results still suggest women in geoscience are at a potential disadvantage from the very beginning of their careers starting with those less than out-standing letters of recommendation.

“We're not trying to assign blame or criticize anyone or call anyone consciously sexist. Rather, the point is to use the results of this study to open up meaningful dialogues on implicit gender bias, be it at a departmental level or an institutional level or even a discipline level.” Which may lead to some recommendations for the letter writers themselves.

51. What do we learn about applicants to postdoctoral positions in geosciences?
 - A) There are many more men applying than women.
 - B) Chances for women to get the positions are scarce.
 - C) More males than females are likely to get outstanding letters of recommendation.
 - D) Male applicants have more interest in these positions than their female counterparts.
52. What do studies about men and women in scientific research show?
 - A) Women engaged in postdoctoral work are quickly catching up.
 - B) Fewer women are applying for postdoctoral positions due to gender bias.
 - C) Men are believed to be better able to excel in STEM disciplines.
 - D) Women who are keenly interested in STEM fields are often exceptional.
53. What do the studies find about the recommendation letters for women applicants?
 - A) They are hardly ever supported by concrete examples.
 - B) They contain nothing that distinguishes the applicants.
 - C) They provide objective information without exaggeration.
 - D) They are often filled with praise for exceptional applicants.
54. What did Dutt and her colleagues do with the more than 1,200 letters of recommendation?
 - A) They asked unbiased scholars to evaluate them.
 - B) They invited women professionals to edit them.

- C) They assigned them randomly to reviewers.
D) They deleted all information about gender.
55. What does Dutt aim to do with her study?
- A) Raise recommendation writers' awareness of gender bias in their letters.
B) Open up fresh avenues for women post-doctors to join in research work.
C) Alert women researchers to all types of gender bias in the STEM disciplines.
D) Start a public discussion on how to raise women's status in academic circles.



APPENDIX M

WORKSHOP FEEDBACK QUESTIONNAIRE

Dear teachers,

This questionnaire is to investigate your opinions of the Familiarization work shop. There is no right or wrong answer. Please feel free to respond to the statements below by writing some comments or notes that best suit to your feeling. The details of your statement or notes will be discussed during the interview. Your response to the questionnaire will be kept confidential.

Thank you for your cooperation.

Workshop Name: _____

Participant Name: _____

Date: _____

Job Title: _____

Years in present position? <1 1-3 3-5 5+ 10+ 15+

INSTRUCTIONS

Please write down your comments and notes under each question.

Your feedback is sincerely appreciated. Thank you.

Section 1. Participants' feedback towards Familiarization workshop

1. Do you clearly understand the objective of the workshop? If yes, what is it?

_____.

2. What is your opinion of the pace and content of the activities? Is the pace **appropriate**? Is the content **clear** and **useful** for you to understand the CEFR? Have you encountered any **difficulties** and **problems** while conducting these activities?

2.1 Presentation (30 minutes)

- the background of the CEFR
- the content of the CEFR
- the impact of the CEFR
- the CEFR and language tests

Pace: _____.

Content: _____.

Difficulties in understanding: _____.

Others, please specify: _____.

2.2 Descriptor-sorting activity & self-assessment of own English proficiency (45 minutes)

- providing the randomly unordered descriptors to the participants
- participants sort the descriptors into the CEFR levels
- checking the sorting result with the original version
- reporting the sorting results to the judgment sheet
- discussion

Self-assessment of own English proficiency

- making self-assessment of participants' own English reading proficiency
- discussing with other participants.

Pace: _____.

Content: _____.

Difficulties in understanding: _____.

Others, please specify: _____.

2.3 Descriptor-sorting activity 2 & reconstructing self-assessment grid (45 minutes)

- providing the randomly unordered descriptors to the participants
- participants sort the descriptors into the CEFR levels
- checking the sorting result with the original version
- reporting the sorting results to the judgment sheet
- discussion

Reconstructing the self-assessment grid

- providing a piece of paper to the participants with the CEFR levels on the left and the descriptors empty on the right
- participants reconstruct the descriptors in corresponding to each CEFR level
- checking answer with the original version of the self-assessment grid
- discussing with other participants

Pace: _____.

Content: _____.

Difficulties in understanding: _____.

Others, please specify: _____.

2.4 Descriptor-sorting activity 3 (45 minutes)

- providing the randomly unordered descriptors to the participants
- participants sort the descriptors into the CEFR levels
- checking the sorting result with the original version
- reporting the sorting results to the judgment sheet
- discussion

Pace: _____.

Content: _____.

Difficulties in understanding: _____.

Others, please specify: _____.

2.5 conclusion (15 minutes)

Pace: _____.

Content: _____.

Difficulties in understanding: _____.

Others, please specify: _____.

3. Do you think of you are capable of linking a test to the CEFR based on your current understanding.

Section 2: Participants' feedback towards Specification workshop

1. Do you clearly understand the objectives of the workshop? If yes, what is it?

2. What is your opinion of the pace and content of the workshop? Is the pace **appropriate**? Is the content **clear** and **useful**? Have you encountered any **difficulties** and **problems** at each stage? What are they?

2.1 Re-familiarization (120 minutes)

2.2 Introduction of the Dutch CEFR Grid (120 minutes)

-the content of the Dutch CEFR Grid

-the usage of the Dutch CEFR Grid

-pre-analyzed test example with the Dutch CEFR Grid

Pace: _____.

Content: _____.

Difficulties in understanding: _____.

Others, please specify: _____.

2.3 Analyzing the content of the CET-4 Reading Comprehension Tests (180 minutes)

Pace: _____.

Content: _____.

Difficulties in understanding: _____.

Others, please specify: _____.

APPENDIX N

A SEMI-STRUCTURED INTERVIEW ON THE FAMILIARIZATION WORKSHOP

Dear teachers,

This interview is to investigate your opinions both on the Familiarization workshop and Specification workshop. Section 1 is for Familiarization workshop, and section 2 is for Specification workshop. There is no specific answer. Please feel free to talk about your feeling and ideas about the workshop. Your statement to the interview will be kept confidential.

Thank you for your cooperation.

Workshop Name: _____

Training Location: _____

Participant Name: _____

Date: _____

Job Title: _____

Years in present position? <1 1-3 3-5 5+ 10+ 15+

Section 1: Interview questions for Familiarization workshop

1. Have you encountered any problems or difficulties while carrying out the following sorting activities? If yes, what are the problems or difficulties?

- a) descriptor-sorting activity 1
- b) self-assessment of own English proficiency
- c) descriptor-sorting activity 2
- d) reconstructing self-assessment grid
- e) descriptor-sorting activity 3

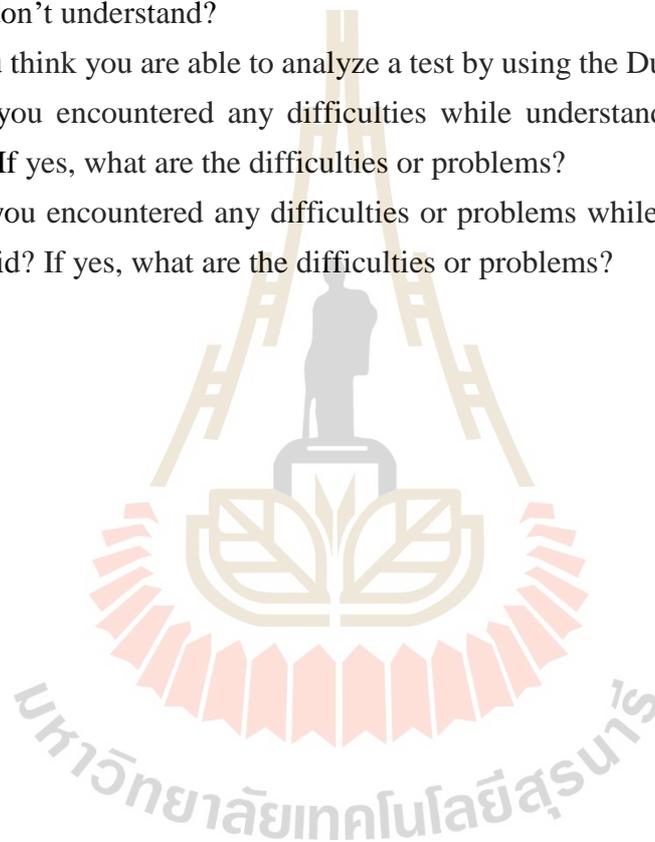
2. Do you think the following sorting activities helpful for you to understand the CEFR? why?

- a) descriptor-sorting activity 1
- b) self-assessment of own English proficiency
- c) descriptor-sorting activity 2
- d) reconstructing self-assessment grid
- e) descriptor-sorting activity 3

3. Do you think of you are capable of linking a test to the CEFR based on your current understanding.

Section 2: Interview questions for Specification workshop

1. Do you think the training of the Dutch Grid useful? Why?
 - the introduction of the usage of the Dutch Grid
 - the pre-analyzed examples with the Dutch Grid
2. Do you understand how to use the Dutch Grid to analyze a test? If no, what aspects you don't understand?
3. Do you think you are able to analyze a test by using the Dutch Grid? If no, why?
4. Have you encountered any difficulties while understanding the usage of the Dutch Grid? If yes, what are the difficulties or problems?
5. Have you encountered any difficulties or problems while analyzing a test with the Dutch Grid? If yes, what are the difficulties or problems?



APPENDIX O

THE INFORMED CONSENT FORM

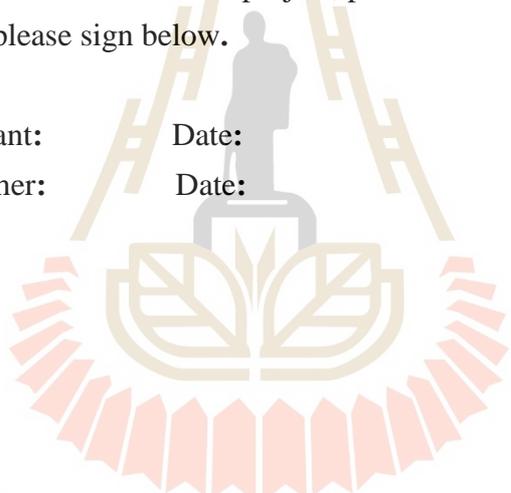
The research project in which you will participate is to linking CET-4 Reading Comprehension Tests to the CEFR. It will be conducted for a period of 3 weeks in July. Your role in this project is to analyze the content of CET-4 Reading Comprehension Tests by using a tool (the Dutch CEFR Grid). You will be fully trained to have an adequate understanding to the CEFR and to know how to use the tool for doing analysis. During the project, you will be interviewed by the researcher and talks will be recorded. All of the information that you provide will be kept completely confidential. When I write up my report, I promise not to use your real name. I shall give you the result of the project once the research is finished. If you have any questions about the research project, please feel free to ask me. If you agree to participate in it, please sign below.

Signed Participant:

Date:

Signed Researcher:

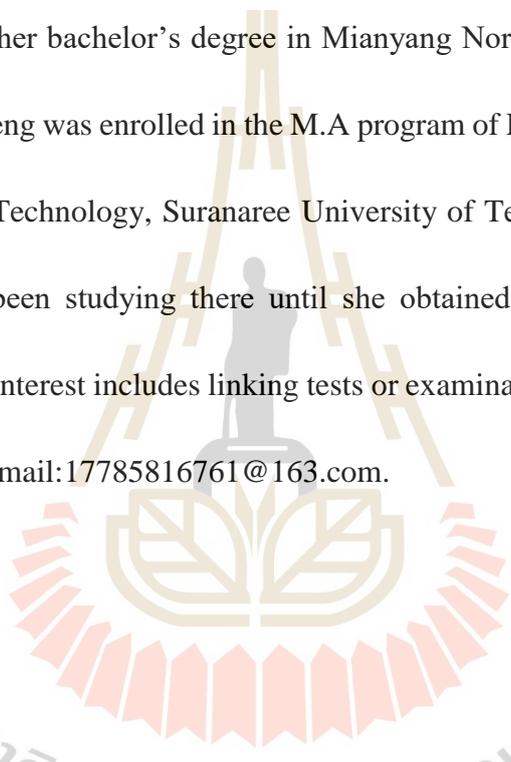
Date:



มหาวิทยาลัยเทคโนโลยีสุรนารี

CURRICULUM VITAE

Xinli Deng was born on September 21, 1990 in Guiyang, Guizhou province, China. She earned her bachelor's degree in Mianyang Normal University in 2014. In June 2015, Xinli Deng was enrolled in the M.A program of English Languages Studies, Institute of Social Technology, Suranaree University of Technology, Thailand. Since then on, she had been studying there until she obtained her master's degree. Her academic research interest includes linking tests or examinations to the CEFR. She can be reached at the e-mail: 17785816761@163.com.



มหาวิทยาลัยเทคโนโลยีสุรนารี