

Aligning Academic Reading Tests to the Common European Framework of Reference for Languages (CEFR)

SIVAKORN TANGSAKUL

Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand

KORNWIPA POONPON*

Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand

Corresponding author email: korpul@kku.ac.th

Article information	Abstract
Article history:	<i>Given the significant global influence of the Common European Framework of Reference for Languages: Teaching, Learning, and Assessment (CEFR) on English language education, this study deals with aligning a university's academic reading tests to the CEFR. It aimed at validating the test construct of the academic reading tests in relation to the proficiency levels defined by the CEFR. The study employs two standard setting procedures outlined in the CEFR Manual: the Familiarization procedure and the Specification procedure, to explore the CEFR level of the academic reading tests as well as the prominent characteristics of the reading texts and the test items in terms of their level and key features. Three academic reading tests were randomly selected. The CEFR Content Analysis Grid for Reading was employed to characterize the content of test items and test tasks. The results indicated that 9 out of 18 reading texts were estimated to correspond to the B2 CEFR level. Texts estimated as B1 and C1 levels were evenly distributed, and none of the reading texts were classified as A1 or C2 levels. Moreover, the findings demonstrated a significant prevalence of B2 and B1 level items. Specifically, B2 items represented the largest proportion at 31.88%, closely followed by B1 items at 25.12%.</i>
Keywords:	
English proficiency test	
Academic reading	
CEFR	
Higher education	
Test alignment	

INTRODUCTION

In light of the government's commitment to enhance Thailand's competitiveness and foster a more global perspective, the Ministry of Education has acknowledged the imperative for Thai people to attain a proficient level of English language skills. This recognition stems from the fact that English is widely regarded as an international and global language (Kirkpatrick, 2008). To support English language learning and teaching, the Ministry of Education has adopted the Common European Framework of Reference for Languages (CEFR) across all levels of the education system since 2014 (Office of the Basic Education Commission, 2014). The primary goals of implementing the CEFR policy were twofold: to improve the quality of English instruction and learning, with the expectation of enhancing the proficiency of both teachers and students and to establish a standardized measure of achievement for Thai students. Likewise, in 2016, the Office of the Higher Education Commission (OHEC) also embraced

the CEFR as a benchmark for comparing test results. This entailed the requirement for every higher education institution to administer an English standardized test to evaluate the English language proficiency of graduating students. Additionally, under the official announcement by OHEC, it was mandated that students' scores should be aligned with the CEFR or other equivalent standards (Office of the Higher Education Commission, 2015).

Aligning language tests with CEFR has not been new but has been increasingly accepted (e.g., Brunfaut & Harding, 2014; Hidri, 2021; Knoch & Frost, 2016; Shin et al., 2022; Wudthayagorn, 2018). Since its introduction in 2001 by the Council of Europe, the CEFR has prompted increasing interest among test developers who seek to align their English language tests with the CEFR levels. Notable high-stakes standardized tests that have been linked to the CEFR include the TOEFL iBT (Tannenbaum & Wylie, 2008) and TOEIC (Tannenbaum & Wylie, 2019). In Thailand, the OHEC has been encouraging higher institutions to set English competence exit requirements upon graduation and also to map their institutional test scores to the proficiency levels of the CEFR (Office of the Higher Education Commission, 2015). Several standardized English proficiency tests, developed by Thai universities, have been mapped to the CEFR, such as the Srinakharinwirot University-Standardized English Test (SWU-SET) (Ativorakun & Wudthayagorn, 2018), Chulalongkorn University-Test of English Proficiency (CU-TEP) (Wudthayagorn, 2018), the Test of English for Thai Engineers and Technologists (TETET) (Jaturapitakkul & Todd, 2018), the CMRU-TEP, by Chiang Mai Rajabhat University (Nakanitanon, 2021), and The Thammasat University General English Test (TU-GET) Computer-based test (CBT) (Shin et al., 2022). These tests have undergone mapping processes to align their proficiency levels with those specified in the CEFR. It has become apparent that mapping test scores to the CEFR framework contributes to meaningful interpretation. Score users can easily understand which score range corresponds to specific CEFR levels, as the CEFR scales provide a convenient framework for understanding and communicating the progression of language proficiency. Furthermore, they facilitate the assessment of an individual's proficiency concerning this progression (Council of Europe, 2001). Most importantly, the CEFR allows a globally recognized interpretation of language proficiency levels. Its descriptive scales offer a practical means to communicate an individual's improvement in language skills. Mapping English proficiency tests to the CEFR levels is a constructive method for assigning tangible significance to those scores (Kane, 2012).

The Khon Kaen University Academic English Language Test (KKU-AELT) is an academic English proficiency test developed by Khon Kaen University (KKU), Thailand. Target test takers are both Thai and international individuals who are interested in enrolling in a graduate program or applying for a lecturer position at KKU. The test includes reading and writing parts. The scores obtained from the KKU-AELT serve as an important indicator for test takers to assess their readiness for graduate-level study. Despite a few studies conducted to support the validity and reliability of the KKU-AELT (Poonpon, 2021; Srisawat & Poonpon, 2023; Thongyoi & Poonpon, 2020), these mostly relied on academic writing tests. More research on reading tests may be needed to ensure the validity and meaningful interpretation of reading test scores (Reynolds et al., 2021). Since the KKU-AELT reading tests were claimed to be designed based on the CEFR B level (Poonpon, 2021), validation of the test construct and content should be worth exploring. In particular, examining how well these reading tests align with the standard framework is crucial. Therefore, the present study aimed to validate the test construct of the KKU-AELT

reading tests in relation to the proficiency levels defined by the CEFR. In particular, the study explored the level of the academic reading tests based on the CEFR framework, as well as the prominent characteristics of both the reading texts and the test items in terms of their level and key features. This study was guided by the following research questions:

- 1) What are the CEFR level and the major characteristics of reading texts of the KKU-AELT reading tests?
- 2) What are the CEFR level and the major characteristics of test items of the KKU-AELT reading tests?

LITERATURE REVIEW

1. The Common European Framework of Reference (CEFR)

According to North (2007), an overview of the CEFR is presented, encompassing common reference levels along with a descriptive scheme. The common reference levels serve as a foundation for analyzing second language (L2) learning objectives and informing the design of teaching materials, learning activities, and assessment practices. The descriptive scheme encompasses two main dimensions: vertical and horizontal.

The vertical dimension of the CEFR outlines the proficiency levels of language learners in listening, reading, speaking, and writing skills. It categorizes these levels into ascending six levels from A1 to C2 with three broad bands: Basic User, Independent User, and Proficient User. Each level delineates what learners are expected to be capable of and their level of proficiency using positive "can do" statements. As Figueras (2012) emphasizes, while language education may vary across countries, programs, and classrooms, it commonly focuses on specifying what students 'can do' and what can be captured instead of what they are unable to do. Likewise, "can do" descriptors illustrate typical performances associated with a particular proficiency level, and they can be observed, measured, and described clearly (Jin et al., 2017).

The horizontal dimension of the CEFR involves the categorization of language use contexts into diverse elements, such as purposes, mental contexts, constraints, situations, domains, communicative themes, and communicative tasks. This dimension also encompasses the communicative language competences of learners and the strategies they utilize to bridge the gap between their linguistic resources (competences) and actual communicative activities (real ability).

The CEFR assessment benchmark is straightforward, determining the level of proficiency a test taker can achieve. It is widely recognized as an internationally standardized approach for assessing English language proficiency, alongside other renowned assessments like IELTS and TOEFL. Hence, the CEFR serves as a global framework for language development and testing, offering common reference levels that have been standardized as benchmarks for describing language proficiency.

Due to its significant international impact and positive reception from various researchers, many countries outside of Europe have embraced the CEFR and integrated it into their language education contexts. In Thailand, the OHEC, which operates under the Ministry of Education, introduced the CEFR to the basic education system in 2014. This initiative recommended the implementation of the CEFR, along with Communicative Language Teaching (CLT), as the latest policy to enhance the quality of English learning and teaching in Thai schools. The implementation of the CEFR in the Thai English learning context serves the purpose of establishing achievement benchmarks for Thai students. According to Wudthayagorn (2018), students are expected to achieve a minimum CEFR level of A1 by the end of grade 6, A2 upon completion of grade 9, and B1 upon concluding grade 12 or vocational college. In 2015, the Office of the Higher Education Commission (2015) also adopted the CEFR into higher education. Consequently, all higher education institutions were required to assess students' English language proficiency using the CEFR or similar standards as a reference upon graduation.

2. CEFR aligning methods

The Council of Europe (2009) has developed a comprehensive guide known as the Manual, which presents a recommended set of procedures for aligning language tests with the CEFR. The Manual consists of five interrelated sets of procedures: Familiarization, Specification, Standardization Training, Standard Setting, and Validation. The process of aligning tests to the CEFR typically begins with the initial stage called Familiarization, followed by the Specification, Standardization Training, Standard Setting, and finally, Validation. The primary objective of the Familiarization stage is to provide participants with a thorough understanding of the CEFR. The Specification stage, also referred to as construct congruence (Tannenbaum & Cho, 2014), is an essential component of the standard-setting process. Its primary purpose is to describe how the content of the test aligns with the CEFR. This stage is considered one of the initial steps in standard-setting procedures as it provides evidence that the test content and tasks are consistent with the desired framework. Typically, the Specification stage involves completing various Forms provided in the Manual. The General Description Forms (A1-A8) are completed by the test developer or coordinators who possess comprehensive knowledge of the test's development and the estimated proficiency levels in relation to the CEFR. The Standardization Training stage is designed to enhance understanding of the CEFR, while the Standard Setting stage focuses on determining the cut-off scores that define test-takers' borderline performances. The Validation stage, on the other hand, involves gathering evidence to establish the validity and reliability of the test.

The procedures outlined in the Manual have been extensively employed by numerous researchers to establish a connection between different tests and the CEFR. By thoroughly examining the test specifications and items, the Manual assists test providers in gaining a comprehensive understanding of test quality, which is vital for enhancing or reforming tests to meet international standards (Council of Europe, 2009). Wu and Wu (2007) have highlighted the Manual's comprehensive assessment of test quality and its utility as a valuable tool for re-evaluating language tests. Furthermore, North (2011) confirmed the Manual was designed to support ongoing improvement in the processes used to relate tests to the CEFR, ensuring that assessments were coherent and realistically reflect the proficiency levels described by the framework.

Given the focus of the present study to align locally developed academic reading tests onto the CEFR, two CEFR aligning procedures were employed: the Familiarization stage and the Specification stage. In the Specification stage, the CEFR Content Analysis Grid for Reading was utilized to analyze the KKU-AELT reading tests and establish its alignment with the CEFR.

3. Related studies

Several scholars have conducted studies concerning the alignment of tests with the CEFR in different academic contexts. For example, O'Sullivan (2010) conducted a study to establish empirical evidence connecting the Communicator, a comprehensive test of reading, listening, and writing skills in English, with CEFR level B2. This investigation followed the four interrelated steps outlined in the initial pilot version of the Manual. Familiarization activities involved the participation of two trainers and nine participants. During the Specification stage, participants completed Specification forms. The outcomes demonstrated that this alignment project has promoted a systematic reform of the Communicator. Similarly, Papageorgiou (2007) undertook a study to establish a connection between the Trinity College London GESE and ISE exams and the CEFR. This investigation followed the five procedures delineated in the Manual. The findings from the analysis of panelists' rating consistency and their comprehension of the CEFR were favorable, and the alignment of the exam with the CEFR was achieved through an examination of test content during the Specification phase.

In Asia, the alignment of tests with CEFR has been increasingly conducted. In Taiwan, Wu and Wu (2010) conducted a study to align the General English Proficiency Test (GEPT) with the CEFR, adopting a three-step process outlined in the Pilot Manual such as Familiarization, Specification, and Standardization. The participants consisted of 20 experienced professionals in English teaching and testing. During the Familiarization stage, the participants engaged in various activities, which encompassed presenting the CEFR to them, self-assessing their English proficiency using the CEFR's self-assessment grid, sorting CEFR descriptors, and comparing their outcomes with the CEFR level descriptors. In the Specification stage, the researchers employed the CEFR Content Analysis Grid for Reading to assess the different levels of the GEPT Reading Comprehension Tests, aiming to align them with the CEFR levels. A total of 49 texts and 200 items in the GEPT Reading Comprehension Tests were analyzed. The content analysis results revealed a trend of increasing complexity in GEPT reading texts as the GEPT level advanced. The study's authors noted the challenge of lacking precise and clear guidelines for making comparisons between constructs and difficulty levels across different testing systems. In 2016, Zou conducted a study with the goal of aligning the National Matriculation English Test (NMET) with the CEFR. This alignment focused specifically on the reading and listening sections of the test paper, and it followed the four procedures outlined in the Manual. The findings of this study demonstrated that the receptive skills of listening tasks in the NMET could be associated with the A2 level of the CEFR, while the reading tasks could be linked to the B1 level of the CEFR (Zou, 2016). Another study by Xinli (2019), who conducted a study to establish a connection between the College English Test Band four (CET-4) Reading Comprehension Tests in mainland China and the CEFR. To achieve this alignment, the study analyzed three sets of CET-4 Reading Comprehension Tests, aiming to identify the key characteristics of the texts and items within the CET-4 Reading Comprehension Test.

The study's outcomes revealed a general alignment of CET-4 Reading Comprehension Tests with the B1 level of the CEFR.

In Thailand, Wudthayagorn (2018) conducted a study aimed at aligning the CU-TEP with the CEFR. To accomplish this, the study engaged the expertise of 13 experts who assessed 120 CU-TEP items using the yes/no Angoff (YNA) technique. In this assessment, the experts determined whether a borderline student at A2, B1, B2, and C1 levels would provide a correct response to each item. These evaluations took place over three rounds. The study's findings revealed CU-TEP cut-off scores of 14, 35, 70, and 99 for A2, B1, B2, and C1 levels, respectively, out of a total of 120 points.

METHODOLOGY

1. Khon Kaen University Academic English Language Test (KKU-AELT)

KKU-AELT is an institutional English language proficiency test developed and administered by the Center for English Language Excellence (CELEX) at Khon Kaen University (KKU). It is offered to individuals seeking to assess their English proficiency and use their test scores for graduate program applications and is also open to those interested in applying for a lecturer position at KKU.

There are two parts in this three-hour test: a reading part (two hours) and a writing part (one hour). The KKU-AELT reading tests comprise six reading texts. Texts 1 to 4 feature multiple-choice test items with four options, while texts 5 and 6 include a combination of multiple-choice items and short answer questions where test takers are required to provide a brief response in a blank space. The test content of the KKU-AELT reading tests draws from a variety of text sources, including general English reading texts and academic reading texts. In the writing part of the KKU-AELT, test takers are required to write a minimum of 250 words in response to a given prompt. To ensure a reliable evaluation, each response is assessed by a minimum of two raters, with an inter-rater reliability level of at least .80 (Srisawat & Poonpon, 2023). The present study only focused on academic reading tests. The writing part was not included as it has been explored in subsequent studies by the authors.

The KKU-AELT test results are reported in proficiency levels, consisting of seven score bands (i.e., Band 1 – 7). These score bands are determined based on raw score ranges. These bands are defined as follows: Band 1 (0 – 18), Band 2 (19 – 35), Band 3 (36 – 43), Band 4 (44 – 61), Band 5 (62 – 73), Band 6 (74 – 87), and Band 7 (88 – 100). These KKU-AELT band descriptors offer a comprehensive insight into the competency associated with each level, encompassing both the overall test performance and individual skill assessments. Prospective applicants aiming to enroll in a Master's degree program at KKU are required to achieve a minimum score of Band 4 on the KKU-AELT. As for applicants aspiring to enroll in a Doctoral degree program, a minimum score of Band 5 is obligatory in both the reading and writing sections for eligibility to gain admission to KKU's Graduate School (Graduate School of Khon Kaen University, 2022).

2. Participants

After the study was approved by the ethics committee, three English language lecturers were recruited. They were informed of the consent requirements and gave their consent prior to participating in the study. They were selected through a purposive sampling method with three selection criteria, including a minimum of five years of teaching experience in universities, expertise in test development, and a comprehensive understanding of the KKU-AELT reading tests. Despite their initial lack of familiarity with the CEFR, the participants were provided with training activities through the Familiarization and Specification workshops to assist them. As a result, these individuals actively engaged in the CEFR Familiarization workshop and completed all the suggested activities. The participants were asked to engage in Familiarization activities designed to foster a comprehensive understanding of the CEFR. This understanding was to be achieved through a sequence of activities delineated in the Manual. Additionally, they were required to analyze the KKU-AELT reading tests and establish a connection with CEFR levels by employing the CEFR Content Analysis Grid for Reading.

3. Materials

3.1 Three sets of KKU-AELT reading tests

Three complete KKU-AELT reading tests administered in the first and second semesters of the academic year 2021 were randomly selected. Each set comprised 6 reading texts and 69 test items, resulting in a cumulative count of 18 reading texts and a total of 207 test items.

3.2 Materials in the familiarization activities

To support the participants' understanding of the CEFR, this study employed the Familiarization activities outlined in the Manual (Council of Europe, 2009). The provided materials encompassed: 1) the CEFR levels for interaction and production, 2) CEFR levels for reception, 3) a self-assessment grid for reading, and 4) the CEFR scales for reading (e.g., overall reading comprehension, reading for information and argument).

3.3 CEFR content analysis grid for reading

The CEFR Content Analysis Grid for Reading was designed as a framework based on the CEFR to analyze language test items, texts, tasks, and specifications (Alderson et al., 2006) (see Appendix). It comprises two key components to be examined: reading texts and test items. The reading texts encompass various characteristics, including text source, authenticity, discourse type, domain, topic, nature of the content, text length, vocabulary, grammar, and estimated CEFR level of text levels. Simultaneously, the characteristics of the test items consist of item type, operations, and estimated CEFR level of test items.

4. Data collection procedures

Aligning a test to the CEFR can be seen as a process of building an argument based on a theoretical rationale. The Manual outlines five interrelated sets of procedures, which encompass

Familiarization, Specification, Standardization Training, Standard Setting, and Validation (Council of Europe, 2009). Given the focus of the present study to align the coverage of the KKU-AELT reading tests onto the categories of the CEFR, two CEFR aligning procedures were employed: the Familiarization stage and the Specification stage. Figure 1 illustrates the data collection procedures with accompanying details.

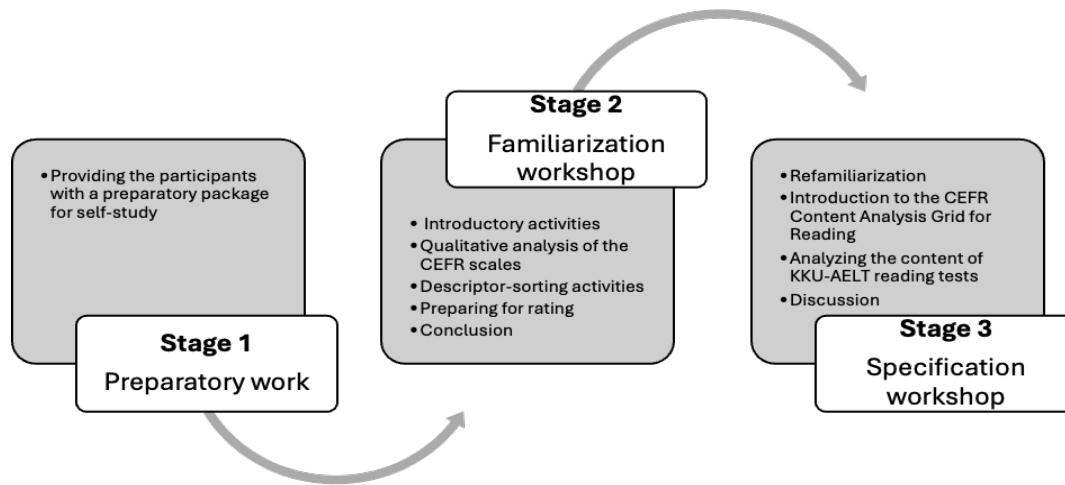


Figure 1 Data collection procedures

The data collection procedures consist of three clearly defined phases: Initial preparations, the Familiarization workshop, and the Specification workshop. Detailed explanations for each stage are presented below.

Stage 1: Preparatory work

Before the Familiarization workshop, the preparatory work involved providing participants with a self-study preparatory package. This package encompassed the following materials: 1) a simplified version of the salient features of CEFR levels for interaction and production, 2) a simplified version of the salient features of the CEFR levels for reception, 3) a self-assessment grid for reading, 4) the CEFR scales for reading (e.g., Overall Reading Comprehension, Reading for Information and Argument, and Reading for Orientation), and 5) a comprehensive written information sheet describing the overall linking procedures.

Stage 2: Familiarization workshop

The purpose of the Familiarization workshop was to provide participants with adequate comprehension of the CEFR through a series of activities outlined in the Manual. Table 1 illustrates an overview of the estimated time allocated to each activity during the Familiarization stage.

Table 1
Estimated time for familiarization activities

Activities	Time
1) Introductory activity	60 min
1.1 A presentation of CEFR and the Familiarization workshop	
1.2 Descriptor-sorting activity 1 (the CEFR levels for interaction and production)	
1.3 Self-assessment of participants' English reading proficiency by using a self-assessment grid	
2) Qualitative activity including group work	30 min
2.1 Descriptor-sorting activity 2 (the CEFR levels for reception)	
2.2 Reconstructing self-assessment grid	
3) Preparation for rating	30 min
3.1 Descriptor-sorting activity 3 (the scales related to reading skills)	

After completing these activities, their judgment sheets were collected, and intra-rater and inter-rater reliability were reported.

Stage 3 Specification workshop

The purpose of the Specification workshop was to analyze the examination content and establish a relationship with the CEFR levels based on coverage (Council of Europe, 2009). During this stage, the participants were asked to examine the content of the KKU-AELT reading tests by using the CEFR Content Analysis Grid for Reading as a framework. The Specification workshop comprised three steps: 1) re-familiarization and introduction of the CEFR Content Analysis Grid for Reading, 2) analysis of test papers, and 3) discussion.

Following the completion of re-familiarization activities, which mirrored those in the Familiarization stage, an introduction to the CEFR Content Analysis Grid for Reading was provided, and then a sample text was analyzed. Subsequently, participants individually engaged in analyzing the contents of the KKU-AELT reading tests (one set of papers). These analyses were conducted in three separate sessions, and at the end of each session, all judgment sheets rated by the participants were collected for further analysis. The time allocated for the Specification stage is estimated in Table 2 below.

Table 2
The estimated time for the specification workshop

Activities	Time
1. Re-familiarization & introduction of the CEFR Content Analysis Grid for Reading (one sample text was analyzed)	3 hours
2. Participants individually analyze one test paper	3 hours
3. Participants individually analyze one test paper	3 hours
4. Participants individually analyze one test paper	3 hours
5. Discussion of the results	1 hour

5. Data analysis

After collecting the data from the descriptor-sorting activities and calculating the scores of the participants, all the data were analyzed to examine the inter- and intra-rater reliability.

The analysis of intra- and inter-rater reliability findings (0.86 for intra-rater reliability and 0.90 for inter-rater reliability) allowed us to infer that all participants exhibited a solid comprehension of the CEFR scale and that they were capable of accurately evaluating the test items by utilizing the relevant can-do statements from the CEFR. Descriptive statistics (i.e., mean, median, SD, and percentage) were also used to analyze the data from the CEFR Content Analysis Grid for Reading. Finally, the characteristics and the CEFR levels of the KKU-AELT reading tests and test items were reported.

RESULTS

1. The CEFR level and the major characteristics of reading texts of the KKU-AELT reading tests

The analysis of the reading texts in this study was conducted based on ten key characteristics specified in the CEFR Content Analysis Grid for Reading. The results of each characteristic's analysis are presented below.

1) Text sources

In the first analysis of the KKU-AELT reading tests, the focus was on the text source. A total of 40 text sources were available, as indicated in the Grid. However, from the 18 reading texts examined, only 5 categories were identified. As illustrated in Table 3, the five categories of text sources in the KKU-AELT reading tests were: public announcements and notices (27.78%), journal articles (27.78%), magazine/online articles, and newspapers (22.22%), newspapers (16.67%), and conference programs (5.55%). Hence, the majority of text sources in the KKU-AELT reading tests were public announcements, notices and journal articles, accounting for 10 out of 18 reading texts.

Table 3
Text source of KKU-AELT reading texts

Text source	Frequency & Percentage			
	Set 1	Set 2	Set 3	Total
Public announcements & notices	2	1	2	5 (27.78%)
Magazine/online articles	2	1	1	4 (22.22%)
Journal articles	2	1	2	5 (27.78%)
Conference Programs	0	1	0	1 (5.55%)
Newspaper	0	2	1	3 (16.67%)
Total	6	6	6	18 (100%)

2) Authenticity

The Grid categorized texts into three types: genuine, adapted, and pedagogic. Table 4 shows that out of the 18 texts in the KKU-AELT reading tests, 12 are genuine, six are adapted/simplified, and none are pedagogic.

Table 4
Authenticity of KKU-AELT reading texts

Authenticity	Frequency & Percentage			
	Set 1	Set 2	Set 3	Total
Adapted/simplified	3	2	1	6 (33.33%)
Genuine	3	4	5	12 (66.66%)
Pedagogic	0	0	0	0 (0.00%)
Total	6	6	6	18 (100%)

3) Discourse types

The third characteristic of the KKU-AELT reading tests was the discourse type. According to Table 5, of the five types, 38.88% of the texts are classified into mainly expository type, 22.22% are mainly argumentative and 16.67% are mainly descriptive and mainly instructive argumentative. Only one input text is classified as mainly narrative

Table 5
Discourse types of KKU-AELT reading texts

Discourse type	Frequency & Percentage			
	Set 1	Set 2	Set 3	Total
Mainly Descriptive	1	2	0	3 (16.67%)
Mainly Narrative	1	0	0	1 (5.55%)
Mainly Expository	2	3	2	7 (38.88%)
Mainly Instructive	1	0	2	3 (16.67%)
Mainly Argumentative	1	1	2	4 (22.22%)
Total	6	6	6	18 (100%)

4) Domains

The KKU-AELT reading tests consisted of four domains: personal, public, occupational, and educational. Table 6 reveals that the most prevalent domain among all the texts was public (38.88%). The personal and educational domains accounted for the same proportion, 27.78%. The occupational domain had the lowest frequency.

Table 6
Domains of KKU-AELT reading texts

Domain	Frequency & Percentage			
	Set 1	Set 2	Set 3	Total
Personal	2	0	3	5 (27.78%)
Public	1	4	2	7 (38.88%)
Occupational	1	0	0	1 (5.55%)
Educational	2	2	1	5 (27.78%)
Total	6	6	6	18 (100%)

5) Topics

Based on the participants' analysis, Table 7 presents 10 distinct topics that were categorized across the three sets of KKU-AELT reading tests. Among these topics, education had the highest

representation, with 8 out of 18 texts making up 44.44% of the total. The environment topic followed closely at 11.11%. Notably, the topics of food and drink, free time, entertainment, places, job ads, technology, news, history, and relationships with other people each accounted for the same proportion of 5.55%.

Table 7
Topics of KKU-AELT reading texts

Topics	Frequency & Percentage			
	Set 1	Set 2	Set 3	Total
Food and Drink	1	0	0	1 (5.55%)
Free time, entertainment	0	0	1	1 (5.55%)
Places	1	0	0	1 (5.55%)
Education	3	3	2	8 (44.44%)
Job ads	1	0	0	1 (5.55%)
Technology	0	1	0	1 (5.55%)
News	0	1	0	1 (5.55%)
Environment	0	1	1	2 (11.11%)
History	0	0	1	1 (5.55%)
Relationships with other people	0	0	1	1 (5.55%)
Total	6	6	6	18 (100%)

6) Nature of content

There were four categories of the nature of content outlined in the CEFR Content Analysis Grid for Reading: mostly concrete content, only concrete content, fairly abstract content, and mainly abstract content. As depicted in Table 8, the majority of texts fell under the 'mostly concrete' category, representing 77.78%. The remaining two categories, only concrete content and fairly abstract content, had equal proportions of 11.11%.

Table 8
Nature of content of KKU-AELT reading texts

Nature of content	Frequency & Percentage			
	Set 1	Set 2	Set 3	Total
Only concrete content	0	2	0	2 (11.11%)
Mostly concrete content	5	4	5	14 (77.78%)
Fairly abstract content	1	0	1	2 (11.11%)
Mainly abstract content	0	0	0	0 (0.00%)
Total	6	6	6	18 (100%)

7) Text length

Text length was determined by counting the number of words in each reading text. Overall, no notable difference was observed among the three sets of KKU-AELT reading tests in terms of text length (Table 9). Table 9 provides the specific text lengths for reference. In each set of the KKU-AELT reading tests, the first and fourth texts had similar word lengths. However, the second, third, fifth, and sixth texts varied in length, with the second text ranging from 291 to 419 words, the third text from 334 to 416 words, the fifth text from 338 to 549 words, and the sixth text from 669 to 783 words. Notably, the sixth text was considerably longer than the other five, with a word count ranging from 669 to 783.

Table 9
Text lengths of KKU-AELT reading texts

Test/Text	Text 1 (total words)	Text 2 (total words)	Text 3 (total words)	Text 4 (total words)	Text 5 (total words)	Text 6 (total words)
Set 1	207	294	358	208	402	689
Set 2	213	291	334	231	338	669
Set 3	180	419	416	215	549	783
Average length	200	335	370	220	430	715

8) Vocabulary

The result from the vocabulary analysis, using the English Vocabulary Profile, an online tool supported by the Council of Europe, was reported. Table 10 shows that the highest proportion of vocabulary in the three sets was at the A1 level, accounting for 58.95% in total, followed by the A2 level at 12.45%. The B1 level was only 2% less than A2 level, with 10.91%. Predictably, the vocabulary at C1 and C2 levels was the least frequent, accounting for 2.38% and 1.31%, respectively. Furthermore, approximately 6.72% of the vocabulary was unlisted. These words included such as proper nouns, compound words, and less common words found in texts.

Table 10
CEFR level of vocabulary in KKU-AELT reading texts

CEFR level of Vocabulary	Frequency & Percentage (%) of words			
	Set 1	Set 2	Set 3	Total
C2	28	15	41	84 (1.31%)
C1	50	58	44	107 (2.38%)
B2	173	128	164	465 (7.28%)
B1	265	188	244	697 (10.91%)
A2	258	254	283	795 (12.45%)
A1	1,215	1,195	1,355	3,765 (58.95%)
Unlisted	125	150	154	429 (6.72%)

9) Grammar

The analysis of the CEFR level of grammar in the Grid was carried out by referencing the English Grammar Profile Online. Table 11 illustrates the distribution of grammar structures across the three sets of KKU-AELT reading tests, with percentages for each CEFR level as follows: A1 (4.68%), A2 (13.50%), B1 (34.16%), B2 (41.33%), and C1 (6.33%). A1 level had the fewest grammar structures, totaling only 17 grammar structures. Conversely, the B2 level exhibited the highest proportion at 41.33%, followed by the B1 level at 34.16%. Notably, no C2-level grammar structures were identified.

Table 11
CEFR level of grammar in KKU-AELT reading texts

CEFR level of Grammar	Frequency & Percentage (%) of grammar types			
	Set 1	Set 2	Set 3	Total
C2	0	0	0	0 (0%)
C1	8	10	5	23 (6.33%)
B2	49	37	64	150 (41.33%)

CEFR level of Grammar	Frequency & Percentage (%) of grammar types			
	Set 1	Set 2	Set 3	Total
B1	36	39	49	124 (34.16%)
A2	20	15	14	49 (13.50%)
A1	2	9	6	17 (4.68%)
Total	115	110	138	363 (100%)

10) Estimation of CEFR level of the reading texts

The final characteristic of reading texts analyzed in the Grid involved estimating the CEFR levels of the texts. Based on the analysis, the texts in the KKU-AELT reading tests were determined to consist of four levels. These four levels are presented in Table 12. The findings revealed that 50% of the reading texts were estimated at the B2 level. Specifically, there were three texts from the first set, four from the second set, and two from the third set falling within this category. Conversely, texts estimated to be at both B1 and C1 levels were distributed evenly, each accounting for 22.22%. Nevertheless, out of the 18 texts analyzed, only one was categorized at the A2 level, while none of the reading texts were identified as A1 or C2 levels.

Table 12
Levels of texts estimated for KKU-AELT reading texts

Texts level estimated	Frequency & Percentage (%)			
	Set 1	Set 2	Set 3	Total
C2	0	0	0	0 (0.00%)
C1	1	1	2	4 (22.22%)
B2	3	4	2	9 (50.00%)
B1	2	1	1	4 (22.22%)
A2	0	0	1	1 (5.55%)
A1	0	0	0	0 (0.00%)
Total	6	6	6	18 (100%)

2. The CEFR level and the major characteristics of test items of the KKU-AELT reading tests

The analysis of three aspects of test items included item types, operations, and the estimated CEFR levels of the test items. Each characteristic's analysis results were described as follows.

1) Item types

The first aspect of the characteristics of test items dealt with the item types. Based on the participants' analysis of the three sets of KKU-AELT reading tests, two item types were identified: multiple choices and short answers. As seen in Table 13, both item types were consistently distributed in the three sets of KKU-AELT reading tests (57 multiple-choice items and 12 short answers in each set). The majority of test items were of the multiple-choice type, representing 82.60%, while short answers accounted for 17.40%.

Table 13
Item types of KKU-AELT reading tests

Item type	Frequency & Percentage (%) of test items			
	Set 1	Set 2	Set 3	Total
Multiple choices	57	57	57	171 (82.60%)
Short answers	12	12	12	36 (17.40%)
Total	69	69	69	207 (100%)

2) Operations

The second aspect covered the operations of the items. Table 14 displays the frequency of each operation. Identifying factual information had the highest occurrence, accounting for 18.35%. Summarizing information in a passage and guessing vocabulary from context followed closely, with frequencies of 17.40% and 15.45%, respectively. On the other hand, understanding rhetorical purpose and simplifying information were less frequent, with frequencies of 5 and 4 out of 204 items, respectively. Interestingly, the operation of making inferences about the author's meaning, commonly found in other academic English tests like TOEFL iBT, also appeared in 21 out of 204 items, representing 10.15% (Milanović, 2011).

Table 14
Operations of KKU-AELT reading tests

Operations	Frequency & Percentage (%) of test items			
	Set 1	Set 2	Set 3	Total
Detect the topic and main idea/gist	7	7	5	19 (9.18%)
Simplify information	0	2	2	4 (1.94%)
Guess vocabulary from context	10	11	11	32 (15.45%)
Make inferences about what author means	7	8	6	21 (10.15%)
Identify a reference	8	7	8	23 (11.11%)
Identify factual information	14	10	14	38 (18.35%)
Identify negative factual information	9	11	9	29 (14.00%)
Understand rhetorical purpose	2	1	2	5 (2.42%)
Summarize information in a passage	12	12	12	36 (17.40%)
Total	69	69	69	207 (100%)

3) Item level estimated

The last characteristic analyzed was the estimated CEFR level of the test items. Table 15 displays the distribution of item level estimated. Based on the analysis, the highest proportion of items was classified at the B2 level, representing 31.88%, followed by B1 at 25.12%. A1, A2, and C1 items were nearly equal, with percentages of 12.56%, 16.43%, and 14.00%, respectively. Notably, no test items were categorized as C2 levels.

Table 15
Estimated CEFR level of test items

Item level estimated	Frequency & Percentage (%) of test items			
	Set 1	Set 2	Set 3	Total
C2	0	0	0	0 (0.00%)
C1	10	10	9	29 (14.00%)
B2	21	21	24	66 (31.88%)
B1	17	18	17	52 (25.12%)
A2	12	12	10	34 (16.43%)
A1	9	8	9	26 (12.56%)
Total	69	69	69	207 (100%)

DISCUSSION

This section discusses the key characteristics of the academic reading texts and test items identified in the KKU-AELT reading tests, which can depict clear CEFR levels of the tests.

1. Characteristics of the KKU-AELT academic reading texts

The results of the study revealed four dominant characteristics of the academic reading texts in the tests, which can depict clear CEFR levels of the tests. These included text sources, domains, topics, text length, vocabulary and grammar, and the estimated CEFR level.

The text sources identified in the present study were journal articles and public announcements or notices. Out of 40 types of text sources in the Grid, only five types were identified, i.e., public announcements/ notices, magazine/online articles, journal articles, conference programs, and newspapers. However, journal articles and public announcements or notices were the most prevalent. The KKU-AELT reading texts were derived from two primary sources: general English reading texts and academic reading texts. The KKU-AELT reading tests served as an academic English reading assessment and were used as an entrance exam for graduate students (Poonpon, 2021). This likely explained why the main text sources identified in the present study were journal articles and public announcements or notices. The predominant use of journal articles and public announcements in the KKU-AELT reading tests has significant implications for test developers. The emphasis on journal articles, for example, aligns with the academic orientation of the test. This implied that test developers prioritized content that reflected the academic reading skills necessary for success in graduate-level studies. In addition, journal articles typically exhibited a higher level of text complexity in terms of language, structure, and concepts. Relying heavily on this source could increase the difficulty level of the reading test, potentially making it more challenging for test takers who are less familiar with academic-style writing. Meanwhile, including public announcements or notices helped incorporate real-world text types into the test. This authenticity was important for assessing test takers' ability to comprehend and interpret practical, everyday information, which is a valuable skill in academic and professional settings. It helped them develop the necessary skills to tackle academic literature and formal announcements effectively.

In terms of text domains, the findings revealed that all four domains in the CEFR framework were found, including personal, public, occupational, and educational domains. Nevertheless, the distribution of text domains differed. Among the domains, the public domain had the highest frequency, accounting for 38.88%, followed by the personal and educational domains, each with an equal frequency of 27.78%. These findings aligned with prior research conducted by Wu and Wu (2010), demonstrating that the GEPT reading texts were predominantly situated within the public and educational domains, and indicating a decrease in the proportion of texts within the personal domain as the GEPT level progresses.

The distribution of text domains in the KKU-AELT reading tests raised some important considerations. While the CEFR emphasized the significance of the personal domain for language teaching, learning, and testing (Council of Europe, 2001), the findings indicated that the text in this domain only accounted for 27.78% of the test. This discrepancy suggested the need for test developers to align the test more closely with international standards by giving greater emphasis to texts in the personal domain. As per the Council of Europe (2001, p. 45), each language use act was situated within a specific context, organized in one of the social life domains. This domain selection had significant implications for the choice of situations, purposes, themes, and texts in teaching and testing materials. It was essential for test developers to consider the motivational effects of selecting domains with present relevance concerning their future utility.

When text domains were taken into account, the relevance of the test can be significantly improved. For example, including texts from the personal domain, which test takers may find more engaging and relevant to their experiences, can boost their motivation and involvement. Additionally, this approach helped in contextualizing the usage of language. As every act of language usage occurred within a specific context, it was crucial for test developers to integrate a range of text domains. This ensured that reading tests reflected a variety of real-life situations and purposes. This variety helped in assessing the test taker's ability to understand and interpret texts across different domains of social life. The results of the present study underscored the importance of considering the distribution of text domains in the KKU-AELT reading tests. By aligning the tests more closely with international standards, incorporating more texts from the personal domain, and recognizing the motivational impact of relevant and relatable materials, test developers could enhance the effectiveness and validity of the language assessment process.

Moreover, the results concerning text topics in the KKU-AELT reading tests exhibited a diverse range of topics, aligning well with the requirements of the Grid. These topics covered areas such as education, food and beverage, free time and entertainment, places, job classifieds, news, environment, and history. Notably, the topics related to education received particular emphasis in this study, confirming that the tests emphasize academic contexts.

When considering text length, vocabulary, and grammar, the results revealed consistent patterns in text length of the KKU-AELT reading tests, with similar patterns with the approximate words in paragraphs one to six at 200, 300, 350, 200, 400, and 700 long in each set of test papers, respectively. Regarding vocabulary, A1 level had the highest proportion, followed

by A2 level, while B1 vocabulary was slightly less than A2. However, vocabulary at C1 and C2 levels showed the lowest frequency. In terms of grammar structures, the majority of them were challenging at the B2 level, which had the highest proportion, followed by the B1 level. Notably, the A1 level had the fewest grammar structures among all the CEFR levels in the KKU-AELT reading tests. These findings offered both contributions and challenges to understanding the construct validity of the test. In this case, it concerned the KKU-AELT's effectiveness in assessing academic English reading proficiency as defined by the CEFR levels. Regarding the text length consistency, the consistent pattern in text lengths across the tests contributed to the construct validity by a standardized measure of reading complexity. This consistency supported the test's reliability, as test takers were exposed to similarly structured reading texts across different test versions. In light of grammar structure alignment, the predominance of B2-level grammar structures, followed closely by B1, suggested that the test was designed to challenge and assess independent users. This aligned well with the goal of evaluating academic readiness for graduate studies, as such levels of proficiency were often required. Meanwhile, the distribution of vocabulary levels, with a higher frequency of A1 and A2 levels and lesser emphasis on C1 and C2 levels, indicated a potential misalignment with the intended CEFR B level. This could challenge the construct validity by suggesting that the vocabulary may not be sufficiently challenging for a test aimed at assessing readiness for graduate-level study. The findings suggested an opportunity to enhance the test's content validity, particularly by increasing the representation of higher-level vocabulary (C1 and C2) and ensuring a broader range of grammar structures were assessed. This would align the test more closely with the advanced proficiency levels expected of graduate students and academic professionals.

In addition, the estimated CEFR level of the academic reading texts was found to be at A2, B1, B2, and C1 levels. Among these levels, B2 had the highest frequency as per the participants' analyses. According to the B2 level, learners possess high independence in reading. They can adapt their reading style and speed to various texts and purposes. B2-level learners can extract information, ideas, and opinions from specialized sources within their field and comprehend specialized articles beyond their field, as well as articles and reports on contemporary issues with specific viewpoints (Council of Europe, 2001). The finding confirmed the construct designed by the test developer using the CEFR B level as the framework (Poonpon, 2021).

2. Characteristics of KKU-AELT academic reading test items

The Grid categorized test items into three main aspects: item types, operations, and estimation of CEFR levels. The two types of items found were multiple-choices and short-answers. Throughout the three sets of KKU-AELT Reading tests, these two item types were consistently represented, with multiple-choice items having the highest proportion.

The participants' analyses of the operations of the test items revealed a diverse range of operation types. Among operation types outlined in the Grid, identifying factual information obtained the highest frequency, followed by summarizing information in a passage, guessing vocabulary from context, understanding rhetorical purpose, and making inferences about what the author means. The alignment of these findings with other academic English tests, such as

the TOEFL iBT, indicated a similarity in the skills being assessed. The reading section of the TOEFL iBT, known for assessing non-native speakers' ability to understand university-level academic texts and passages, employed input texts from real university-level textbooks and included questions to measure test-takers' basic comprehension, ability to make inferences, and reading for detail (Milanović, 2011). Moreover, the topics covered by the passages in the TOEFL iBT may be unfamiliar to the test takers. Likewise, Wu and Wu (2010) observed that, with increasing GEPT levels, the test items became progressively more cognitively demanding. The GEPT Advanced level reading comprehension test incorporated both careful reading and expeditious reading elements. The expeditious reading tasks encompassed activities such as targeted searches to grasp the main ideas of the texts and efficient scanning to pinpoint specific details.

These similarities suggested that the KKU-AELT reading tests were designed to measure similar reading operation types, aligning with other internationally recognized academic English exams. The incorporation of diverse operation types in the KKU-AELT reading tests indicated a comprehensive approach to evaluating language proficiency, including the ability to comprehend, analyze, and interpret various types of text. By assessing a wide range of reading operations, the test could provide a more accurate evaluation of a test taker's reading skills, ensuring that the results reflected their ability to navigate academic content effectively (Cohen & Upton, 2007). We may assume that test tasks designed for KKU-AELT reading tests were those corresponding to tasks test takers may be faced with in target language situations.

Regarding the estimation of CEFR levels of test items, the results revealed a significant presence of B2 and B1 level items. B2 items represented the highest proportion at 31.88%, followed closely by B1 items at 25.12%. A1, A2, and C1 level items were nearly equal, with percentages of 12.56%, 16.43%, and 14.00%, respectively. There were no test items categorized as C2 level. These findings are consistent with Wudthayagorn's (2018) research, where the study also analyzed the CEFR levels of test items in determining cut-off scores between CEFR levels. Wudthayagorn's study revealed that the widest range of the CU-TEP scores in relation to the CEFR level was at B1, followed by B2, C1, and A2. This suggests that language assessments often focus on higher representation of B1 and B2 level items. Likewise, the absence of C2 level items in both studies could be attributed to the specific nature of the test takers and the purpose of the test. For instance, within the context of the KKU-AELT reading tests, a noteworthy observation was that the target test-taker demographic consisted of graduate students. This observation may be attributed to the fact that there was a relatively small proportion of A1 and A2 level test items, which might not align with their language proficiency requirements. Conversely, C2 level items, indicative of near-native language competence, may not be essential for the specific goals and purposes of the test.

In addition, the higher proportion of B2 and B1 level items in the present and Wudthayagorn's studies suggest that these levels were considered crucial for language learners' academic and professional development. The B2 level indicated a high level of language proficiency and independence in reading, while the B1 level signified that learners could read and comprehend texts on familiar topics with some level of autonomy. These levels were often targeted in language assessments to assess test takers' ability to handle academic and real-life situations

effectively (Cohen & Upton, 2007). The findings of both studies emphasized the importance of considering the targeted proficiency levels when designing language assessments. By focusing on relevant CEFR levels, test developers could ensure that the test aligned with the specific needs of test takers and accurately measured their language proficiency in a given context.

IMPLICATIONS

This study aimed at aligning the KKU-AELT reading tests to the CEFR by employing two standard-setting procedures outlined in the CEFR Manual: the Familiarization procedure and the Specification procedure. The alignment procedures are highlighted as practical methods for critically reviewing and evaluating the content and statistical characteristics of examinations. This process is essential for improving the test's validity and reliability. The study offers some implications as follows.

The alignment procedures can serve as valuable guidelines for other test developers seeking to align their tests with the CEFR framework. The Familiarization procedure plays a very crucial role in providing the participants with a deep understanding of the CEFR, enabling them to perform analysis tasks effectively in aligning test content with the CEFR framework. The Content Analysis Grid for Reading used in the Specification step is very effective in facilitating the alignment of the test content to CEFR levels. Additionally, the use of online text analysis tools, the English Vocabulary Profile (Bax, 2012) and English Grammar Profile (Green, 2008), to help the analysis in this Specification procedure proves that they can be very objective and helpful in estimating the CEFR levels for vocabulary and grammatical features in the reading texts of the KKU-AELT reading tests. This application not only reduces the complication of the comparison of construct and difficulty levels across different testing systems (Wu & Wu, 2010), but also offers a more nuanced approach to test alignment with the CEFR framework. Based on these robust analyses the construct validity and the CEFR alignment of the academic reading tests can be more warranted. Therefore, the study suggests such standard-setting procedures as alternatives for future test alignment studies.

CONCLUSION AND LIMITATIONS

The study aimed to align the KKU-AELT reading tests to the CEFR to increase the construct validity of the test. Following the Familiarization and the Specification procedures provided in the CEFR Manual (Council of Europe, 2001), the raters were well trained and the CEFR Content Analysis Grid for Reading was utilized to explore the CEFR level of the academic reading tests and the prominent characteristics of the reading texts and the test items in terms of their level and key features. The results indicated that most of both reading texts and test items of the KKU-AELT reading tests correspond to the B2 CEFR level. Texts estimated as B1 and C1 levels were evenly distributed, and none of the reading texts were classified as A1 or C2 levels.

This study has a few limitations. First, the participants of the present study encountered a prominent difficulty in determining boundaries between different CEFR levels due to the

ambiguity of expressions in certain descriptors. This is consistent with the findings from Alderson et al. (2006), Huang et al. (2018), and Papageorgiou's (2010) studies. Alderson et al. (2006) explored the challenges of using descriptors in the CEFR framework. They found that the level descriptors lacked clear and precise definitions due to vague wording and lack of specific criteria, making it difficult for participants to make consistent judgments about language proficiency levels. Similar to the current study, expressions like "long, short, simple, and complex" created confusion among participants. In addition, Huang et al. (2018) found that the participants found difficulty in determining boundaries such as in distinguishing between B2 and C1. Therefore, different interpretations can lead to deviation in the judgment of test items. Since ambiguity in descriptors can lead to inconsistencies in assessing language learners' abilities and make it challenging to establish clear boundaries between CEFR levels, additional support, e.g., scoring aids or exemplars, is suggested to help raters better understand and apply the descriptors in practice. This method could potentially address the issue of vague descriptions in the CEFR.

Another limitation is related to the experts' judgments. In this study, the estimated CEFR level was estimated by the three experts' agreed judgments. This may be considered subjective. To obtain more evidence-based results, further investigation of the complexity and difficulty of the input texts may help. For example, the study conducted by Wu and Wu (2010) utilized readability formulas from Chall & Dale (1995) and Fry (1968) to assess the complexity and difficulty of the input texts employed in the GEPT reading comprehension tests. By looking at an expansion in vocabulary range and an increase in text difficulty, their findings indicated the level of GEPT advances. This suggests that the CEFR Manual should provide clear guidelines for each level of text, enabling a more evidence-based approach in the aligning process. Establishing clearer criteria would enhance objectivity and consistency in alignment studies.

THE AUTHORS

Sivakorn Tangsakul is a PhD candidate in the Applied Linguistics Program at the Faculty of Humanities and Social Sciences, Khon Kaen University, and an English language lecturer at Nakhon Phanom University, Thailand. His research interests include language testing and assessment and corpus linguistics.

sivakorn_tk@npu.ac.th

Kornwipa Poonpon is an assistant professor of English and Applied Linguistics at the Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand. Her research interests include second language assessment, corpus linguistics, and EAP and ESP pedagogy.

korpul@kku.ac.th

REFERENCES

Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2006). Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR construct project. *Language Assessment Quarterly*, 3(1), 3–30.

Ativorakun, C., & Wudthayagorn, J. (2018). Mapping Srinakharinwirot University – Standardized English Test (SWU-SET) onto the Common European Framework of Reference (CEFR). *Suranaree Journal of Social Science*, 12(2), 69–84. <https://doi.org/10.55766/CTUU4836>

Bax, S. (2012). *Text inspector*. <https://textinspector.com/>

Brunfaut, T. & Harding, L. (2014). *Linking the GEPT listening test to the Common European Framework of Reference* (LTTC-GEPT Research Report No. RG-05). The Language Training and Testing Center.

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Brookline Books.

Cohen, A. D., & Upton, T. A. (2007). “I want to go back to the text”: Response strategies on the reading subtest of the New TOEFL. *Language Testing*, 24(2), 209–250. <https://doi.org/10.1177/0265532207076364>

Council of Europe. (2001). *Common European Framework of Reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Council of Europe. (2001). *Common European Framework of Reference for languages: Learning, teaching, assessment – Structured overview of all CEFR scales*. <https://rm.coe.int/168045b15e>

Council of Europe. (2009). *Relating language examinations to the ‘Common European Framework of Reference for Languages’: Learning, teaching, assessment (CEFR)*. A Manual. <http://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680667a2d>

Figueras, N. (2012). The impact of the CEFR. *ELT Journal*, 66(4), 477–485. <https://doi.org/10.1093/elt/ccs037>

Fry, E. (1968). A readability formula that saves time. *Journal of Reading*, 11(7), 265–271.

Graduate School of Khon Kaen University. (2022). *Graduate student manual 2022*. <https://app.gs.kku.ac.th/images/img/news/announcement/2-1GS-77-66.pdf>

Green, A. (2008). English profile: Functional progression in materials for ELT. *Research Notes*, 33, 19–25.

Hidri, S. (2021). Linking the International English Language Competency Assessment suite of examinations to the Common European Framework of Reference. *Language Testing in Asia*, 11(1), 1–24. <https://doi.org/10.1186/s40468-021-00123-8>

Huang, L., Kubelec, S., Keng, N., & Hsu, L. (2018). Evaluating CEFR rater performance through the analysis of spoken learner corpora. *Language Testing in Asia*, 8(1), 1–17. <https://doi.org/10.1186/s40468-018-0069-0>

Jaturapitakkul, N. & Todd, R. W. (2018). *Test of English for Thai engineers and technologists (TETET)*. Technical Report. https://sola.kmutt.ac.th/tetet/doc/TETET_Technical_report_NJ_RT_Complete.pdf

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), 1–19. <https://doi.org/10.1186/s40468-017-0032-5>

Kane, M. T. (2012). Validating score interpretations and uses. *Language Testing*, 29(1), 3–17. <https://doi.org/10.1111/jedm.12000>

Kirkpatrick, A. (2008). English as the official working language of the Association of Southeast Asian Nations (ASEAN): Features and strategies. *English Today*, 24(2), 27–34. <https://doi.org/10.1017/S0266078408000175>

Knoch, U., & Frost, K. (2016). *Linking the GEPT writing sub-test to the Common European Framework of Reference (CEFR)* (LTTC-GEPT Research Report No. RG-08). The Language Training and Testing Center.

Milanović, M. (2011). The construct of reading and its operationalization in the internet-based test of English as a foreign language. *Philologia*, 9(1), 73–82. <https://doi.org/10.18485/philologia.2011.9.9.8>

Nakanitanon, P. (2021). Linking an English proficiency test to the CEFR: Setting validity cut scores. *The New English Teacher*, 15(1), 53–64.

North, B. (2007). The CEFR: Development, theoretical and practical issues. *Babylonia*, 1, 22–29.

North, B. (2011). Putting the Common European Framework of Reference to good use. *Language Teaching*, 47(2), 1–22.

Office of the Basic Education Commission. (2014). *Guidelines for practices by Ministry of Education: English education policy reform*. <http://upload.sru.ac.th/download.aspx?NFILE=24569561651071773219718420214022421089204114>

Office of the Higher Education Commission. (2015). *Policy of upgrading English education standards of higher education institutions*. http://www.dqe.mhesi.go.th/front_home/Data%20Bhes_2559/04052559.pdf

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 University Press.

Papageorgiou, S. (2007). *Relating the Trinity College London GESE and ISE examinations to the Common European Framework of Reference* (Final Project Report, February 2007). Trinity College London. <http://www.trinitycollege.co.uk/resource/?id=2261>.

Papageorgiou, S. (2010). Investigating the decision-making process of standard setting participants. *Language Testing*, 27(2), 261–282.

Poonpon, K. (2021). Test takers' perceptions of design and implementation of an online language testing system at a Thai university during the COVID-19 pandemic. *PASAA*, 62, 1–28.

Reynolds, C. R., Altmann, R. A., & Allen, D. N. (2021). The meaning of test scores. In C. R. Reynolds, R. A. Altmann & D. N. Allen (Eds.), *Mastering modern psychological testing* (pp. 91–131). Springer. http://dx.doi.org/10.1007/978-3-030-59455-8_3

Shin, S. Y., Sanonguthai, S., & Tangkiengsirisin, S. (2022). *Aligning the CEFR levels with the TU-GET CBT scores*. Thammasat University. <https://litu.tu.ac.th/wp-content/uploads/2023/09/Aligning-the-CEFR-Levels-with-the-TU-GET-CBT-Scores.pdf>

Srisawat, C., & Poonpon, K. (2023). Revision of an academic English writing rubric for a graduate school admission test. *PASAA*, 65, 234–262.

Tannenbaum, R. J., & Cho, Y. (2014). Critical factors to consider in evaluating standard-setting studies to map language test scores to frameworks of language proficiency. *Language Assessment Quarterly*, 11, 233–249. <https://doi.org/10.1080/15434303.2013.869815>

Tannenbaum, R. J., & Wylie, E. C. (2008). Linking English-language test scores onto the Common European Framework of Reference: An application of standard-setting methodology. *ETS Report Series*, 2008(1), 1–75.

Tannenbaum, R. J., & Wylie, E. C. (2019). *Mapping the TOEIC Tests on the Common European Framework of Reference*. Educational Testing Service. <https://www.ets.org/pdfs/toeic/toeic-mapping-cefr-reference.pdf>

Thongyoi, K., & Poonpon, K. (2020). Phrasal complexity measures as predictors of EFL university students' English academic writing proficiency. *rEFLections*, 27(1), 44–61. <https://doi.org/10.61508/refl.v27i1.241750>

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, 1–20.

Wu, J. R., & Wu, R. Y. (2010). Relating the GEPT reading comprehension tests to the CEFR. In W. Martyniuk (Ed.), *Aligning tests with the CEFR* (pp. 204–224). Cambridge University Press.

Wudthayagorn, J. (2018). Mapping the CU-TEP to the Common European Framework of Reference (CEFR). *LEARN Journal*, 11(2), 163–180.

Xinli, D. (2019). *Linking college English reading comprehension tests band four to the Common European Framework of Reference for Languages (CEFR)* [Unpublished master's thesis]. Suranaree University of Technology.

Zou, Y. Y. (2016). *Aligning National Matriculation English Test (NMET) with the Common European Framework of Reference (CEFR)* [Unpublished master's thesis]. Guangdong University of Foreign Language Studies.

Appendix

The CEFR content analysis grid for reading

The CEFR Content Analysis Grid for Reading (also known as the Dutch CEFR Construct Grid) was designed as a framework based on the CEFR to analyze language test items, texts, tasks, and specifications (Alderson et al., 2006). It comprises two key components to be examined: reading texts and test items (see below). The reading texts encompass various characteristics, including text source, authenticity, discourse type, domain, topic, nature of the content, text length, vocabulary, grammar, and estimated CEFR level of text levels. Simultaneously, the characteristics of the test items consist of item type, operations, and estimated CEFR level of test items.

Characteristics	Definition	Measurement
<i>Analysis of Reading texts</i>		
1. Text sources	Text sources in the reading tests drawn from a wide range source	Identifying text sources, e.g., personal letters, stories, brochures, journal articles, magazines/newspapers, advertising material, reference books, notices, regulations, business letters, reports/memorandums, and more
2. Authenticity	Reading texts that manifest themselves in various forms, presenting readers with a range of options to choose from based on their specific objectives and requirements	Categorizing into three types. 1. Genuine texts remained unchanged in their original form. 2. Adapted texts underwent modifications, such as additions, deletions, or changes in wording. 3. Pedagogic texts were simplified for instructional purposes.
3. Discourse types	Classification of reading texts based on their inherent characteristics and communicative purposes	Classifying texts into types: descriptive-oriented, narrative-oriented, expository-oriented, instructive-oriented, and argumentative-oriented
4. Domains	Categorization of reading texts based on the specific context or field to which they belong	Identifying a text domain, e.g., personal, public, educational, or occupational
5. Communication topics	Themes that are addressed within reading texts, focusing on various aspects of human interaction and exchange	Classifying topics of the texts, e.g., daily life, education, weather, culture/customs, science, history, literature, and fine arts
6. Nature of content	Level of abstractness	Classifying texts, e.g., "only concrete," "mostly concrete," "fairly abstract," and "mainly abstract"
7. Text length	Length of reading texts	Number of words in each reading text
8. Vocabulary	Range of vocabulary utilized	Using English Vocabulary Profile Online
9. Grammar	Variety of grammatical features used	Using English Grammar Profile Online
10. Text level estimated	Estimation of CEFR levels of the input texts	Estimating 6 levels: A1, A2, B1, B2, C1 or C2

Characteristics	Definition	Measurement
<i>Analysis of Reading texts</i>		
<i>Analysis of test items</i>		
11. Item types	Diverse range of question formats employed in the reading tests	Identifying formats, e.g., multiple-choice questions, matching exercises, short-answer questions, and summary completion tasks
12. Operations	Description of the mental processes involved in completing each test item	Identifying mental process, e.g., recognizing information, making inferences, drawing a conclusion, or evaluating information
13. Item level estimated	Estimation of the CEFR levels of each test item	Identifying A1, A2, B1, B2, C1, or C2 CEFR level to the test items