PASAA Journal
Special Issue 2025/2026, 93–127

Mapping the KKU-AELT Reading Test to the CEFR: A Standard-Setting Study

Sivakorn Tangsakul and Kornwipa Poonpon*

Khon Kaen University, Thailand
**Corresponding author:* korpul@kku.ac.th

Article information	
Abstract	This standard-setting study aimed to align the Khon Kaen University Academic English Reading Test (KKU-AELT), a locally developed reading test, to the CEFR. The study adopted the staged approach (Council of Europe, 2009), involving familiarization, specification, standardization training, standard setting, and validation. The item-descriptor-matching method, an item response theory-based and test-centered approach (Ferrara et al., 2008), was employed. Five panelists participated in the standard setting process. The instruments included a judgment sheet for the ordered-item booklet, a judgment sheet for the booklet, and three feedback questionnaires. The data obtained were analyzed using classical statistics, i.e., frequency, percentage, mean, standard deviation, and standard error judgment. The intraclass correlation coefficient and Cronbach's alpha were also employed for rater agreement. A high level of inter-rater consistency was found across three rounds of judgment. The results revealed the final cut-off scores for the reading test, with the score ranges corresponding to the CEFR levels delineated as A1 (1-15), A2 (16-31), B1 (32-54), B2 (55-83), and C1 (84-100). The CEFR interpretation of academic

E-ISSN: 2287-0024

	reading scores provides key insights for the ongoing improvement of students' English proficiency. This study can serve as a valuable guide for future research endeavors aimed at aligning locally developed language tests with the CEFR.
Keywords	academic reading, language test, CEFR, Thai higher education, KKU-AELT
APA citation:	Tangsakul, S. & Poonpon, K. (2025/2026). Mapping the KKU-AELT Reading Test to the CEFR: A standard-setting study [Special Issue]. *PASAA, 73*, 93–127.

1. Introduction

The Common European Framework of Reference for Languages (CEFR) is a widely recognized framework for language teaching, learning, and assessment. It provides a standardized method with six reference levels across three proficiency bands: A1-A2 (Basic), B1-B2 (Independent), and C1-C2 (Proficient). These levels describe proficiency in reading, writing, listening, and speaking. The foundational document, the "Common European Framework of Reference for Language: Learning, Teaching and Assessment" (Council of Europe, 2001), includes detailed descriptors for each level, aiding teachers and learners in language development and assessment.

In Thailand, the Ministry of Education adopted the CEFR in 2014, aiming to enhance the English proficiency among Thai students and teachers in the basic education system. The CEFR-based policy initiative set specific targets for students at all school levels: A1 by grade 6, A2 by grade 9, and B1 by grade 12 and vocational college (Office of the Basic Education Commission, 2014). Additionally, the Office of the Higher Education Commission (OHEC) adopted CEFR in 2016, setting benchmarks for undergraduate students to achieve B2 and graduate students to reach C1 by 2024. Higher education institutions must now conduct standardized English tests in order to ensure that graduates meet these CEFR standards (Office of the Higher Education Commission, 2024).

Many schools and universities require CEFR-based test scores for admission and graduation, and job seekers use these scores, such as the TOEIC, to demonstrate their English language proficiency. This has led test providers to align their tests with the CEFR, as shown by various studies (e.g., Brunfaut & Harding, 2014; Knoch & Frost, 2016; Tannenbaum & Baron, 2011; Tannenbaum & Wylie, 2005; Tannenbaum & Wylie, 2008; Wu & Wu, 2010). In order to facilitate this alignment, the Council of Europe introduced recommended procedures in 2003, which was formalized in the 2009 publication, "Relating Language Examinations to the CEFR" (Council of Europe, 2009). Following the introduction of a pilot version of the Manual in 2003, numerous major international standardized tests, primarily offered by the Educational Testing Service (ETS) and the British Council, have aligned their test scores with the six CEFR proficiency levels—A1, A2, B1, B2, C1, and C2 (e.g., Tannenbaum & Baron, 2011; Tannenbaum & Wylie, 2005; Tannenbaum & Wylie, 2008). Examples of standardized English tests mapped to the CEFR include the Test of English as a Foreign Language, the Test of Spoken English (TSE), the Test of Written English (TWE) (Tannenbaum & Wylie, 2005), the TOEFL iBT, the TOEIC, the TOEIC Bridge (Tannenbaum & Wylie, 2008), the TOEFL Junior Comprehensive Test (Tannenbaum & Baron, 2015), the TOEFL ITP (Tannenbaum & Baron, 2011), the IELTS, and the Aptis (O'Sullivan, 2015).

In addition to the internationally recognized standardized tests mentioned above, various Thai universities have recently aligned their institutional English proficiency tests with the CEFR framework. Examples include the Srinakharinwirot University Standardized English Test (SWU-SET) (Ativorakun & Wudthayagorn, 2018), Chulalongkorn University's Test of English Proficiency (CU-TEP) (Wudthayagorn, 2018), the Test of English for Thai Engineers and Technologists (TETET) by King Mongkut's University of Technology Thonburi (Jaturapitakkul & Todd, 2018), Chiang Mai Rajabhat University's CMRU-TEP (Nakanitanon, 2021), and Thammasat University's TU-GET CBT (Shin et al., 2022).

The Khon Kaen University Academic English Language Test (KKU-AELT) is an academic English proficiency test developed by Khon Kaen University (KKU), Thailand. The test includes reading and writing skills. The scores obtained from the KKU-AELT serve as an important indicator for test takers for assessing their readiness for graduate-level studies. Despite a few studies conducted to support the validity and reliability of the KKU-AELT (Poonpon, 2021; Srisawat & Poonpon, 2023; Thongyoi & Poonpon, 2020), research on reading tests is needed in order to ensure the validity and meaningful interpretation of reading test scores (Reynolds et al., 2021). Since the 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. Mapping the KKU-AELT with CEFR scales would enhance its standardization and clarity in assessing language proficiency. Therefore, the present study aimed to map the KKU-AELT reading test scores with the CEFR levels, facilitating the meaningful mutual recognition and interpretation of KKU-AELT reading test scores in alignment with the CEFR. This study is guided by the following research questions:

1) What CEFR-mapped cut-off scores can be established for the KKU-AELT reading test?

2) To what extent does the CEFR mapping process build accurate evidence and judgments to support validity arguments for an academic reading test?

2. Literature Review

2.1 The CEFR for Languages and Its Application in Language Assessment

The CEFR was developed by the Council of Europe as a comprehensive framework for language teaching, learning, and assessment (Council of Europe, 2001). Its common reference levels and descriptive framework have been widely recognized as fundamental tools for establishing second language (L2) learning objectives and for guiding the development of teaching materials, learning activities, and assessment instruments (North, 2007). As an internationally accepted standard, the CEFR provides a shared basis for describing language

proficiency and is comparable in function to global English language assessments such as the IELTS and TOEFL.

The CEFR descriptive framework is structured along two main dimensions: vertical and horizontal. The vertical dimension defines proficiency levels across the four language skills—listening, reading, speaking, and writing—ranging from A1 to C2 and grouped into three broad categories: basic user, independent user, and proficient user. Each level is characterized by positive "can do" descriptors that articulate what learners are able to accomplish in real-life communicative situations. As emphasized by Figueras (2012), despite differences in language education systems across countries and contexts, the CEFR consistently focuses on learners' communicative abilities rather than their limitations. These descriptors represent typical observable and measurable performances associated with each proficiency level (Jin et al., 2017). The horizontal dimension categorizes language use into various elements, including purposes, mental contexts, constraints, situations, domains, communicative themes, and tasks. It also covers the communicative language competences of learners and the strategies they use to align their linguistic resources (competences) with actual communicative tasks (real ability). Together, these dimensions provide a comprehensive representation of language ability as both competence and performance.

Beyond its pedagogical value, the CEFR has had a substantial impact on language testing and assessment. In order to support test developers in aligning examinations with CEFR levels, the Council of Europe introduced a structured framework for relating language tests to the CEFR (Figueras et al., 2005). Following a seminar held in Helsinki in 2002, an initial manual was published in 2003, outlining four key stages: familiarization, specification, standardization, and empirical validation. This framework was later refined in a revised version of the manual (Council of Europe, 2009), which defined five interrelated procedures: familiarization, specification, standardization training (benchmarking), standard setting, and validation.

Familiarization ensures that judges develop a thorough understanding of CEFR levels, scales, and descriptors. Specification involves analyzing test content in relation to CEFR levels, resulting in an initial content-based alignment claim. Standardization training uses CEFR-calibrated exemplars of test tasks and learner performances to establish a shared interpretation of proficiency levels among judges, and a standard setting assigns test takers to CEFR levels based on their performance, while validation employs both internal and external evidence to confirm the accuracy and reliability of the alignment claims.

These stages are not strictly linear. However, the familiarization stage is a prerequisite for both the specification and standardization stages, as a comprehensive understanding of the CEFR scales and descriptors significantly influences the overall quality of the mapping process (Council of Europe, 2009). The manual also details judge-training procedures, using CEFR-calibrated exemplars to ensure consistent, theory-based interpretations of proficiency levels during CEFR mapping.

2.2 The Item-Descriptor Matching Method in Mapping a Language Test and Its Application

The item-descriptor (ID) matching method, proposed by Ferrara et al. (2008), was employed to align KKU-AELT reading test scores with CEFR levels for several reasons. First, the ID matching method is versatile and applicable across various item types, making it ideal for the KKU-AELT reading test, which includes both multiple-choice and short-answer items. Secondly, this method avoids the need for panelists to estimate a student's likelihood of answering correctly, reducing biases and errors in judgment, as supported by Nickerson's (2004) study. Thirdly, it simplifies the panelists' task by focusing on aligning specific test items with CEFR descriptors, rather than imagining a student's capabilities, thus enhancing the reliability and validity of their assessments (Ferrara et al., 2008). Lastly, the data-driven approach of the ID matching method uses actual test scores to support decision-making, allowing panelists to focus on aligning each

item's requirements with the appropriate CEFR levels. The ID matching method is explained in the subsequent section.

The ID matching method is a judgment-based standard-setting procedure used to establish cut-off scores by aligning test items with performance level descriptors (Ferrara et al., 2008). The method relies on expert judgments to determine where score thresholds should be located along a proficiency continuum. First, judges work with an Ordered Item Booklet (OIB), in which test items are arranged from easiest to most difficult based on their IRT difficulty estimates rather than their original test order. Each item is presented individually with its text, any accompanying visuals, and the scoring key. This ordering enables judges to evaluate items progressively across proficiency levels. Secondly, judges use performance-level descriptors, such as CEFR levels, to define the knowledge and skills expected at each proficiency level. These descriptors serve as the primary reference for interpreting item demands and guide judges in categorizing item-level performance. Thirdly, judges consult an item map that provides essential information for each item, including item identification numbers, item type, and difficulty estimates. The item map supports systematic comparison between item requirements and the expectations outlined in the performance level descriptors. Fourthly, judges make two main decisions; they first match each item's knowledge and skill requirements to the most appropriate performance level descriptor, and they then analyze the pattern of these alignments across the ordered items to identify threshold regions, where the distinction between adjacent performance levels is unclear. These regions typically feature items that fluctuate between two adjacent levels, indicating a transition point in proficiency. Finally, cut-off scores are placed within these identified regions. According to Ferrara et al. (2008), these cut-off scores can be determined by the judges reaching an agreement, or calculated by psychometricians using statistical summaries of the judges' judgments. Since the distinction between levels is unclear in these regions, locating cut-off scores within them provides a defensible transition point between proficiency levels. Overall, the ID matching method offers

a systematic and transparent approach to standard setting by combining expert judgment, item difficulty information, and performance level descriptors in order to establish defensible cut-off scores.

3. Methodology

3.1 Panelists

Careful selection of experts is essential in standard-setting procedures. Previous studies have recommended involving a large number of experts in order to enhance the quality and validity of the standard-setting process (Fulcher, 2010). According to Fulcher, a larger number of qualified and experienced experts can improve the reliability of their judgments. However, owing to practical, logistical, and time constraints, the present study included a panel of five experts. These experts were English language lecturers from universities in northeastern Thailand, purposively selected based on the following criteria: 1) a minimum of five years' teaching experience in higher education; 2) experience in English language test development, ensuring familiarity with test items and their objectives; and 3) a sufficient understanding of CEFR descriptors to make informed and accurate judgments. They were recruited and informed of the study's objectives and procedures before signing an informed consent form.

All of the participants were male; two held doctoral degrees, two were Ph.D. candidates, and one held a master's degree. All had extensive experience in English language teaching and assessment. Although the experts were not initially fully familiar with the CEFR, they completed CEFR familiarization workshops conducted in accordance with the official CEFR manual in order to ensure a shared and comprehensive understanding of the framework. Following this training, the panelists analyzed the KKU-AELT reading tests and established cut-off scores. Informed consent was obtained from all of the participants prior to their involvement in the study.

3.2 Materials and Instruments

3.2.1 The KKU-AELT Reading Test

The Khon Kaen University Academic English Language Test (KKU-AELT) is an institutional English proficiency test administered by the Center for English Language Excellence (CELEx) at Khon Kaen University. The test targets individuals assessing their English proficiency, particularly applicants to graduate programs or candidates for lecturer positions. The researchers requested official permission from CELEx to access a sample set of KKU-AELT reading tests and their corresponding test scores for use in the study. After approval, CELEx provided one complete KKU-AELT reading test, administered for 326 test takers in May 2021; all of the identifying information was removed in order to ensure anonymity.

The test consisted of six reading texts, covering both general and academic reading. Reading texts 1 through 4 are centered around short texts or news articles, presented in a multiple-choice format with four choices. Reading texts 5 and 6 emphasize comprehension of longer texts through a combination of multiple-choice and short-answer questions. There are 69 test items in total, each contributing to a maximum possible score of 100. It is noteworthy that these items vary in point value, with some worth 1 point and others 2 points. The test demonstrated a high level of reliability, with a coefficient of 0.88.

3.2.2 Materials in the Familiarization Activities

In order 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 included: 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, etc.).

3.2.3 Ordered Item Booklet

The Ordered Item Booklet was used as a tool in the standard setting workshop in order to determine cut-off scores. The KKU-AELT test items were calculated and arranged in the Order Item Booklet (OIB). The test items started with the least difficult item and progressed to the most difficult item based on the difficulty value (*p*-value), which differed from the order in the original test book. It displayed one item per page. Each page included a reading passage, a test item, and the scoring key.

3.2.4 A Judgment Sheet for the Ordered Item Booklet

A judgment sheet for the OIB was also developed for use in the standard setting workshop in order to determine the CEFR level of each test item. The panelists used this judgment sheet to determine matches between the knowledge and skill requirements of the test items and the descriptions of knowledge and skills in this judgment sheet and then assigned a CEFR level to each test item. This judgment sheet for the OIB was developed based on the CEFR descriptors, including overall reading comprehension, reading for information and argument, and reading for orientation (Council of Europe, 2001).

The judgment sheet for the OIB consisted of two main parts. Part I contained the CEFR reading descriptor scales, focusing on relevant constructs such as overall reading comprehension and reading for information and argument (Council of Europe, 2001). These descriptors provided a common reference framework to guide judges in interpreting the language demands of each item. The judges were encouraged to consult them throughout to ensure their ratings aligned with the intended proficiency definitions. Part II was the judgment recording section, where the judges documented their decisions for each test item; the judges evaluated each item using the guiding question: “Which CEFR level most closely matches the knowledge and skills required to respond successfully to this item?” and assigned a CEFR level (A1–C2) accordingly.

Each row in the judgment sheet corresponded to a single OIB item and included essential information to support informed judgments: the OIB page number, item type (i.e., multiple-choice), the item number from the original test, and the item difficulty value (*p*-value). The final column was reserved for the judge's estimated CEFR level. This structured layout allowed the judges to consider both qualitative aspects (language skills and cognitive demands) and quantitative information (item difficulty) when assigning levels.

3.2.5 Feedback Questionnaires

Three feedback questionnaires were developed in order to measure the procedural validity of the study. One questionnaire was administered at the end of the familiarization stage, another at the end of the standardization training stage, and the third at the end of the standard setting stage. These questionnaires were in order to assess the extent to which the panelists understood each stage and the overall standard-setting procedures, as well as their familiarity with the relevant CEFR descriptors and the clarity of the instructions provided at each stage. All of the questionnaires were designed using a 5-point Likert scale (Likert, 1932) and included items evaluating the clarity of instructions, procedures, and materials used in each session.

Construct validity was examined using the Index of Item-Objective Congruence (IOC) (Rovinelli & Hambleton, 1997). The questionnaires were evaluated by two experts in English language teaching and assessment that had extensive experience in aligning tests with the CEFR. The experts rated each item using the IOC method and provided comments on both language and content. Based on their feedback, the questionnaires were revised and refined. The results showed that the IOC values for the familiarization stage questionnaire, the standardization training stage questionnaire, and the standard setting stage questionnaire were 0.92, 0.92, and 0.89, respectively, indicating a high level of content validity.

3.3 Standard-setting Procedures

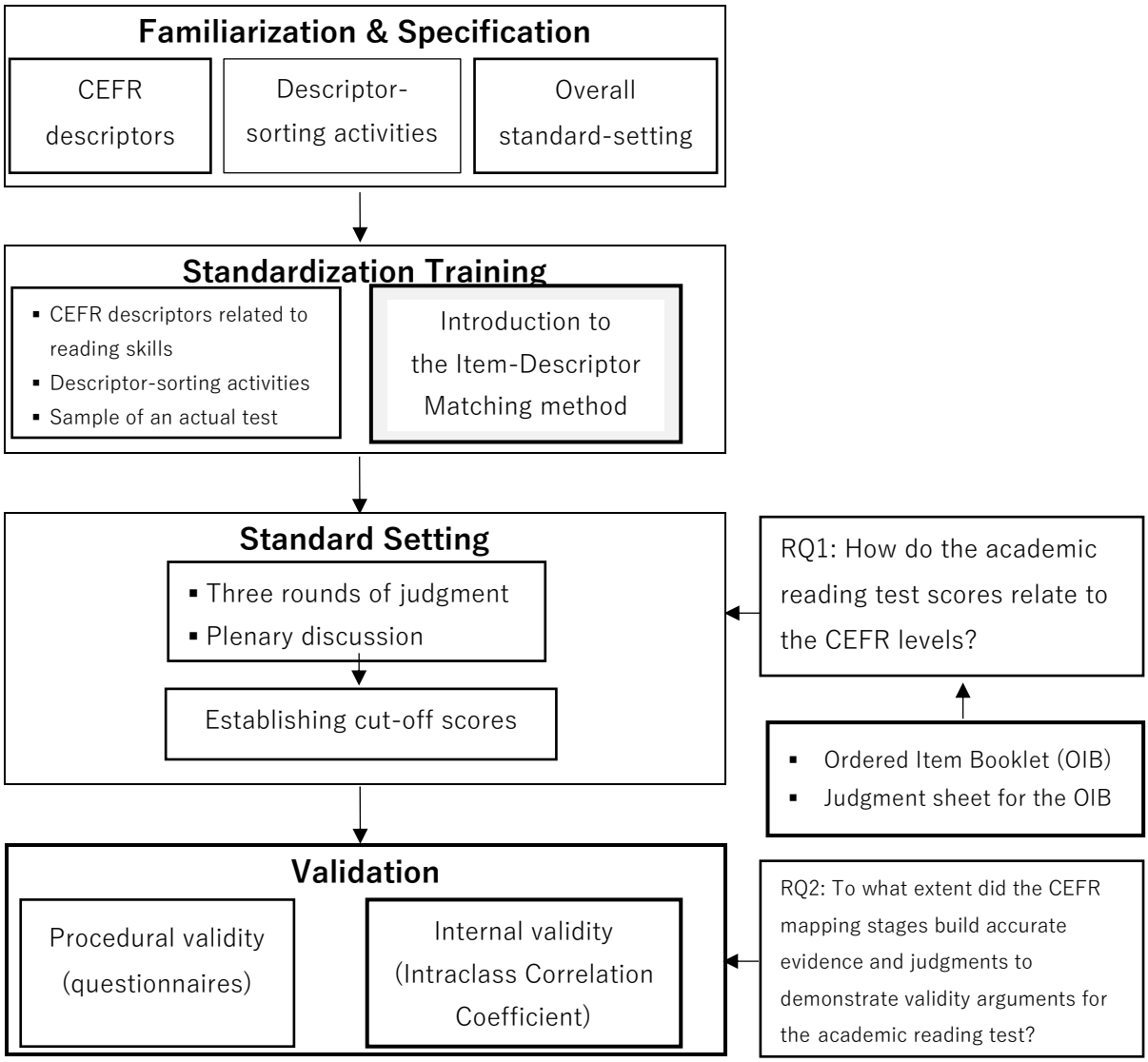
The standard-setting procedures used in the present study strictly followed the stages outlined in the manual (Council of Europe, 2009): familiarization and specification, standardization, standard setting, and validation (Figure 1). The details for each stage—objectives, participants, materials and instruments, activities, and expected outputs—are described below.

3.3.1 The familiarization and specification stage familiarized participants with the CEFR reading descriptors, standard-setting procedures, and ethical considerations. The CEFR training was conducted by the researchers: one was an expert in the CEFR and language testing; the other had undergone extensive CEFR training through careful study of the official CEFR Manual and related documentation. Before conducting the training, the researchers reviewed the materials in detail and undertook self-training to ensure that they were adequately prepared to serve as trainers. During this stage, the panelists participated in a series of activities in order to familiarize themselves with the CEFR reading descriptors and the overall standard-setting procedures, and to acknowledge the ethical procedures and consent requirements.

3.3.2 At the standardization training stage, the panelists were prepared for the standard setting stage. They were introduced to the ID matching method (Ferrara et al., 2008), the OIB, and the judgment sheet. All of the panelists were trained in applying ID matching in order to determine a cut-off score and to practice assigning CEFR levels to the sample test items. The panelists were also encouraged to discuss areas of argument and confusion in their judgment training. Following this, they completed a feedback questionnaire for evaluating the standardization training stage.

Figure 1

Standard-setting Procedures for Mapping the Academic Reading Test Scores to the CEFR

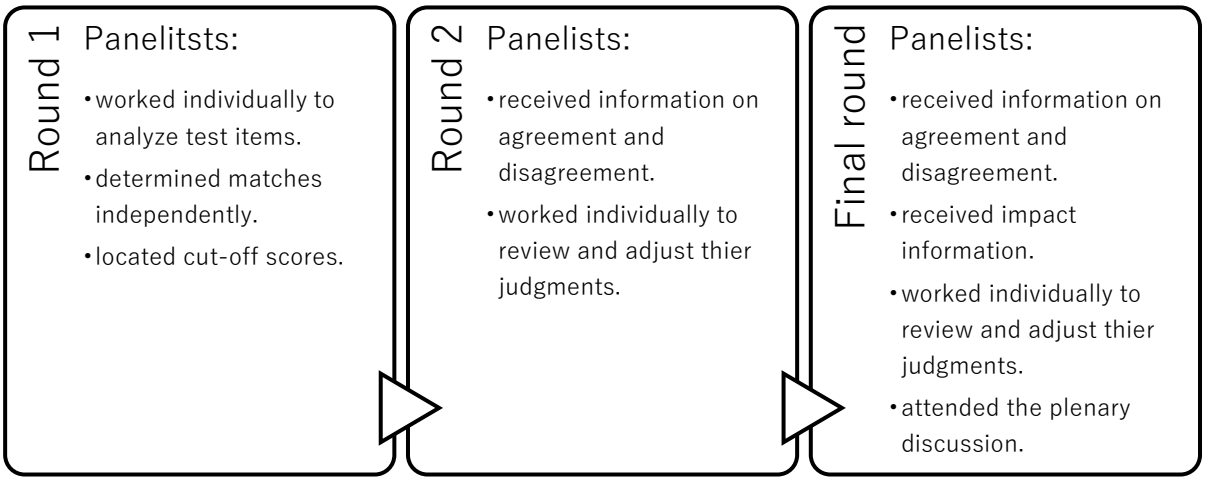


3.3.3 The next stage was the standard setting stage. The objective of this stage was to establish a final cut-off score. The standard setting was divided into three rounds, taking three days. Figure 2 depicts the details of each session. Round 1 involved individual analysis by the panelists. They were required to examine each test item by considering the question, “Which CEFR level most closely matches the knowledge and skills required to respond successfully to this item?” The panelists then provided their judgments for each test item (i.e., A1, A2, B1, B2, C1, and C2) on the judgment sheet for the OIB Concurrently, the panelists independently examined item-descriptor matches and subsequently identified cut-off scores

within the threshold regions. In Round 2, the panelists received feedback detailing their levels of agreement and disagreement with other panelists regarding item-descriptor matches and threshold regions. Then they independently reviewed and adjusted the locations of their cut-off scores, revisited their matches between items and descriptors, and adjusted as needed. Round 3 began with panelists receiving feedback detailing areas of agreement and disagreement regarding the placement of cut-off scores. Then they were provided with impact data, specifically the percentage of test takers at or above each CEFR level, calculating using the median cut-off score. The panelists individually reviewed and adjusted the placements of their cut-off scores, ultimately settling on the final locations. The round ended with a plenary discussion in which the panelists reviewed the outcomes and reached a final consensus. The expected output from this stage is a final cut-off score for each adjacent level of the CEFR (e.g., between A1/A2, A2/B1, B1/B2, and B2/C1). After these three rounds, the panelists were required to complete the questionnaire feedback in order to evaluate the standard setting workshop.

Figure 2

Standard Setting Sessions



3.3.4 The validation stage was to provide procedural evidence and inter-rater reliability to support the final cut-off scores, and procedural validity and

internal validity (Kaftandjieva, 2004) were evaluated. Procedural validity was evaluated by applying clear, systematic, and rigorous criteria to the selection of panelists, ensuring that all possessed extensive English language teaching experience, experience in developing English language tests, and familiarity with the CEFR. Meanwhile, the panelists' judgments from the descriptor-sorting activities in the familiarization stage and the standard setting stage were analyzed in order to provide internal validity. Inter-rater agreement was analyzed to determine the extent to which panelists reached consensus across the various rounds of judgment during the standard-setting sessions.

3.4 Data Analysis

The study used qualitative and quantitative analyses to address the two research questions. The qualitative analysis centrally examined the characteristics and CEFR alignment of each test item. Each item was reviewed in order to identify its underlying constructs, language skills, cognitive demands, and correspondence with the relevant CEFR descriptors. The items were systematically categorized by reading subskills (i.e., identifying main ideas, making inferences, understanding details, and interpreting vocabulary in context) and by CEFR proficiency levels. This interpretive process ensured that the test content reflected the intended proficiency constructs prior to quantitative analysis. For RQ1 (establishing final cut-off scores), the mean, median, standard deviation (SD), and standard error of judgment (SEJ) were calculated. For RQ2, questionnaire data were analyzed; means and SD were computed, and Cronbach's alpha and intraclass correlation coefficient (ICC) were used to assess the inter-rater reliability among the panelists.

4. Findings

4.1 Established CEFR-mapped Cut-off Scores for the KKU-AELT Reading Test

Throughout the judgment process in the standard setting stage, the findings showed four distinct cut-off scores for the KKU-AELT reading test, intended to

align the KKU-AELT reading test scores with the five CEFR levels (A1, A2, B1, B2, and C1). Notably, the distinction between the C1 and C2 levels was omitted from this analysis, primarily due to the test's exclusive focus on a receptive skill (e.g., reading skills) and the constraints imposed by a limited number of test items and the scarcity of individuals within the target test takers that would fall into the C2 level. Hence, this segment provides the cut-off scores for the KKU-AELT reading test for the CEFR levels A1/A2, A2/B1, B1/B2, and B2/C1, as determined through three rounds of panelist evaluation.

The findings from the round 1 judgement (Table 1) showed that the median cut-off scores from the OIB page for A1/A2, A2/B1, B1/B2, and B2/C1 were 16, 23, 37, and 61, respectively. The smallest standard deviation was the boundary between B2 and C1 levels, indicating a high level of agreement among panelists regarding the OIB page for this adjacent level. Conversely, the largest standard deviation was evident at A2/B1, suggesting that the panelists had varying opinions on the A2/B2 cut-off scores. The standard error of judgment (SEJ) followed a pattern similar to the standard deviation, with the smallest SEJ at the B2/C1 level and the largest at the A2/B1 level.

Table 1

Score Boundaries of Round 1 Judgment

Panelists	OIB page cut-off scores for each CEFR boundary			
	A1/A2	A2/B1	B1/B2	B2/C1
P01	16	27	43	58
P02	12	23	41	61
P03	16	23	37	59
P04	20	29	39	63
P05	14	17	37	61
Highest	20	29	43	63
Lowest	12	17	37	58
Median	16	23	37	61
SD	2.97	4.60	2.61	1.95
SEJ	1.33	2.06	1.17	0.87

The findings from the round 2 judgement (Table 2) showed that the median cut-off scores from the OIB page for A1/A2, A2/B1, B1/B2, and B2/C1 were 13, 23, 39, and 61, respectively. The smallest standard deviation remained at the boundary between B2 and C1 levels, indicating a high level of agreement among the panelists on the cut-off scores for this adjacent level. Conversely, the largest standard deviation shifted from the boundary between A2/B1 to B1/B2, suggesting that the panelists held varied opinions on the B1/B2 cut-off score on the OIB page. The SEJ followed a pattern similar to the standard deviation, with the smallest SEJ observed at the B2/C1 level and the largest at the B1/B2 level.

Table 2

Score Boundaries of Round 2 Judgment

Panelists	OIB page cut-off scores for each CEFR boundary			
	A1/A2	A2/B1	B1/B2	B2/C1
P01	15	24	42	59
P02	11	21	37	61
P03	13	23	39	60
P04	12	26	39	61
P05	13	23	38	61
Highest	15	26	42	61
Lowest	11	21	37	59
Median	13	23	39	61
SD	1.48	1.82	1.87	0.89
SEJ	0.66	0.81	0.84	0.40

Table 3

Score Boundaries of Round 3 Judgment

Panelists	OIB page cut-off scores for each CEFR boundary			
	A1/A2	A2/B1	B1/B2	B2/C1
P01	12	21	41	59
P02	11	22	37	60
P03	13	22	39	59
P04	11	24	39	60

Panelists	OIB page cut-off scores for each CEFR boundary			
	A1/A2	A2/B1	B1/B2	B2/C1
P05	11	23	38	60
Highest	13	24	41	60
Lowest	11	21	37	59
Median	11	22	39	60
SD	0.89	1.14	1.48	0.55
SEJ	0.40	0.51	0.66	0.24

The findings of round 3 (Table 3) showed that the OIB page median cut-off scores for the boundaries of the CEFR levels for A1/A2, A2/B1, B1/B2, and B2/C1 were 11, 22, 39, and 60, respectively. The median cut-off scores for the boundaries between A1/A2, A2/B1, and B2/C1 were lower than those for round 2, while the B1/B2 median cut-off score remained the same. The smallest and largest standard deviations in round 3 also followed the same pattern as in round 2: the smallest standard deviation was at the boundary between B2 and C1, and the largest was at the boundary between B1 and B2 CEFR levels. The SEJ followed a similar pattern to the standard deviation, with the smallest SEJ observed at B2/C1 and the largest at B1/B2 level. During round 3, the final judgment round, the panelists engaged in a plenary discussion in order to establish the final cut-off scores. Therefore, the KKU-AELT reading test cut-off scores formally reported in this study were based on the panelists' judgment in this final round.

After establishing OIB page cut-off scores for each CEFR boundary, the next step in standard setting was to determine recommended cut-off scores by converting the raw scores into scaled scores. There were 69 test items, each contributing to the raw score, with a maximum possible score of 100. The items varied in point value, with some worth 1 point and others 2 points, as mentioned earlier. For instance, within the range of the OIB page cut-off scores for the A1 level, spanning OIB pages 1–10, some of the items were valued at 1 point, while others within this range could be assigned a value of 2 points. These designated score points served to describe test takers that possessed the knowledge and

capability specified at a particular CEFR level from those that lacked such proficiency. Table 4 presents the final cut-off scores for the KKU-AELT reading test at CEFR levels based on round 3 findings.

Table 4

Total KKU-AELT Reading Test Cut-off Scores for the CEFR Levels

CEFR levels	KKU-AELT reading test OIB page (69 items)	KKU-AELT reading test scaled scores (100 points)	Score intervals
C1	60 – 69	84 – 100	17
B2	39 – 59	55 – 83	29
B1	22 – 38	32 – 54	23
A2	11 – 21	16 – 31	16
A1	1 – 10	1 – 15	15

As shown in Table 4, upon converting the OIB page cut-off scores, the initial raw score of 69 was transformed into a scaled score of 100. The findings revealed the final cut-off scores for the KKU-AELT reading test based on Round 3 results. The CEFR score ranges were defined as follows: A1 (1 to 15), A2 (16 to 31), B1 (32 to 54), B2 (55 to 83), and C1 (84 to 100). In other words, the test takers achieving scores between 1 and 15 were classified as belonging to the A1 level; conversely, those scoring between 16 and 31 were placed in the A2 level. Similarly, test takers scoring 32–54 were assigned to the B1 level. Similarly, test takers scoring 55–83 were classified as B2 level, and those scoring 84–100 were classified at the C1 level. Moreover, the largest score interval was observed at the B2 level (29 scores), followed by the B1 level (23 scores); meanwhile, score intervals for the A1, A2, and C1 levels were evenly distributed, encompassing scores of 15, 16, and 17 respectively.

After establishing the final cut-off scores, the next step was to classify test takers based on them. Table 5 displays the test takers classified by final cut-off

score ranges according to CEFR levels, based on the test scores from the 326 test takers that took the KKU-AELT reading test administered in May 2021. The total test score was 100, based on 69 test items. Notably, the test demonstrated high reliability, with a coefficient of 0.88. According to Table 5, on the final cut-off scores, out of 326 test takers' scores, most of the test takers' scores were classified into B1 level (48.46%), followed by the A2 level (27.00%), the B2 level (23.62%), and the A1 level (0.92%), respectively. It is important to note that none of the test takers' scores in the study were classified at the C1 level (0.00%).

Table 5

Test Takers Classified by Cut-off Score Ranges with Respect to the CEFR Levels

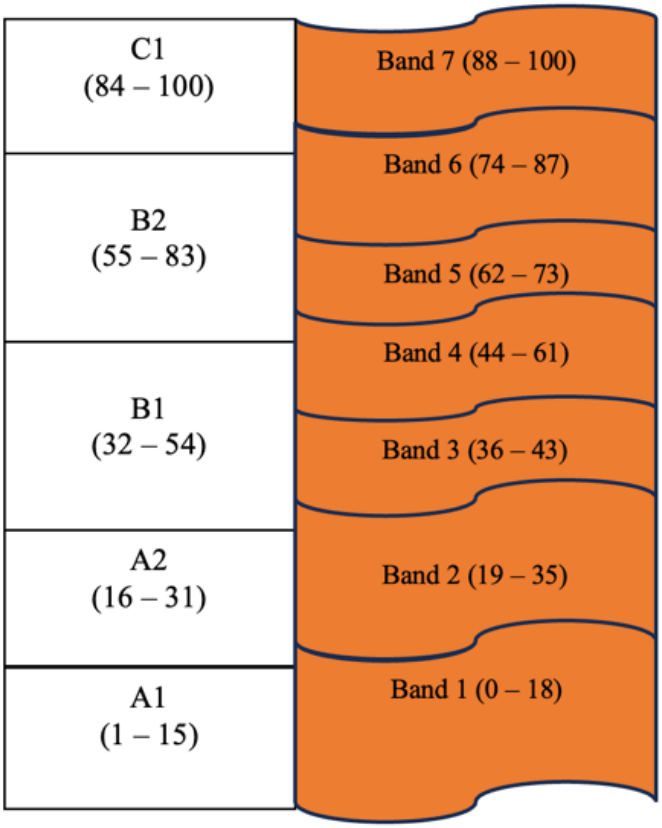
CEFR levels	KKU-AELT reading test scores	Number of test takers (N=326)
C1	84–100	0 (0.00%)
B2	55–83	77 (23.62%)
B1	32–54	158 (48.46%)
A2	16–31	88 (27.00%)
A1	1–15	3 (0.92%)

In addition to reporting test takers classified by the CEFR cut-off score ranges, it was important to compare CEFR levels with KKU-AELT band scores, since the CEFR provides an internationally standardized scale for interpreting English proficiency. Figure 3 shows this comparison. The KKU-AELT results were reported according to seven proficiency levels, defined by scaled score ranges: 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) (Graduate School of Khon Kaen University, 2022). These band descriptors summarize competencies for each level across overall performance and individual skills. Mapping the CEFR reading scores onto the KKU-AELT bands showed substantial overlap between the two systems, A1 level scores (1–15) correspond to Band 1; A2 scores (16–31) fall between Bands 1 and 2; B1 scores (32–54) span Band 2-4; B2 scores (55–83) span Bands 4–6, and C1 scores (84–100) lie between Bands 6 and 7. Thus, most CEFR cut-off ranges

overlapped two or three KKU-AELT bands. For example, a B1-level score could place a test taker in Band 2, 3, or 4, depending on the exact scaled score.

Figure 3

Score Comparison Between the CEFR Levels and the KKU-AELT Band Scores



4.2 The CEFR Mapping Process and Its Validity Arguments for the Academic Reading Test

Procedural and internal validity were examined in order to ensure that the mapping process was robust and trustworthy.

4.2.1 Procedural validity was evaluated using feedback questionnaires administered at three points: after familiarization, after standardization training, and after the standard setting. The panelists' feedback from the familiarization stage (Table 6) was largely positive; they reported a good understanding of the CEFR descriptors and the KKU-AELT reading test, found the instructions clear, and considered the descriptor-sorting activities useful. In addition, most valued

the group discussions for aligning their interpretations of the CEFR levels. One panelist cited time constraints, and a few noted that some descriptors seemed unrelated to the test content or were difficult to apply.

Table 6

Feedback for the Familiarization Sessions

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. After the familiarization, I had a good understanding of the CEFR.	-	-	-	2 (40%)	3 (60%)
2. After the familiarization, I had a good understanding of the KKU-AELT reading test.	-	-	-	2 (40%)	3 (60%)
3. The instructions were clear and easy to follow.	-	-	-	2 (40%)	3 (60%)
4. The familiarization allowed me to understand the CEFR scales and descriptors.	-	-	-	1 (20%)	4 (80%)
5. The relevant CEFR levels and descriptors were clear to me.	-	-	-	1 (20%)	4 (80%)
6. The descriptor-sorting activities were useful for standard-setting procedures.	-	-	-	1 (20%)	4 (80%)
7. I was able to complete the session within the suggested time required.	-	-	1 (20%)	3 (60%)	1 (20%)

b) The panelists' feedback from the standardization training revealed positive responses to the sessions (Table 7). This stage was found to be informative and clearly presented. The examples of the test items and judgment

forms enhanced their understanding of the CEFR, enabling them to assess CEFR levels proficiently. The panelists understood the purpose and process and completed the training activities. They agreed that the discussions facilitated consensus and that the training tasks were effective in familiarizing them with the CEFR descriptors and understanding of the reading skills assessed in the test. However, one panelist reported insufficient time to complete all of the activities. The open-ended responses also highlighted the benefits of discussion, with one panelist noting clearer mapping procedures resulting from active involvement in the training activities. Another panelist mentioned that analyzing the illustrative samples and CEFR descriptors required significant concentration, thus placing a cognitive load on them.

Table 7

Feedback for the Standardization Training Stage

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. The session was informative and clearly presented.	-	-	-	1 (20%)	4 (80%)
2. The illustrative examples helped me to understand the CEFR levels.	-	-	-	1 (20%)	4 (80%)
3. The set of judgment forms (such as the OIB, the Judgment Sheet for OIB, and CEFR descriptors) helped me to determine the cut-off scores of the KKU-AELT reading test.	-	-	-	2 (40%)	3 (60%)
4. The discussion among the judges was helpful in terms of reaching consensus among the judges.	-	-	-	1 (20%)	4 (80%)

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
5. After the training, I felt familiar with the CEFR descriptors.	-	-	-	-	5 (100%)
6. The tasks in the training session helped me to understand the reading skills tested in the test.	-	-	-	2 (40%)	3 (60%)
7. I was able to complete the session within the suggested time required.	-	-	1 (20%)	2 (40%)	2 (40%)

c) The panelists' feedback from the standard setting stage was predominantly positive (Table 8); they clearly understood the standard-setting process, and the tasks of reviewing and responding to the KKU-AELT items improved their comprehension of the assessments. Further, the judgment sheet was considered to be easy to follow, and feedback highlighting disparities in the judgments was valuable. Furthermore, the panelists felt confident in assigning CEFR levels to the test items. Most open-ended comments were positive, and many said that their confidence increased in the second round after feedback and discussion. One panelist described the mapping procedures as well-organized, informative, and enjoyable though intensive; however, another noted that the ID matching method was time-consuming.

Table 8

Feedback for the Standard Setting Stage

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. After I attended the standard setting process, I had a clear understanding of the purpose of the standard setting process.	-	-	-	1 (20%)	4 (80%)

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
2. The process instructions were clear and easy to follow.	-	-	-	1 (20%)	4 (80%)
3. Reviewing and responding to the KKU-AELT reading items helped me to understand the assessment.	-	-	-	2 (40%)	3 (60%)
4. The relevant CEFR levels and descriptors were clear to me.	-	-	-	2 (40%)	3 (60%)
5. I was able to follow instructions and complete the judgment sheet as required.	-	-	-	1 (20%)	4 (80%)
6. Information showing how my judgments differed from the judgments of other participants was helpful.*	-	-	-	1 (20%)	4 (80%)
7. During the standard setting process, I felt confident about my decision on the CEFR levels of the test items.	-	-	-	2 (40%)	3 (60%)
8. I was able to complete the session within the suggested time required.	-	-	-	1 (20%)	4 (80%)

*Note: The panelists were asked to review their judgments where they deviated significantly from other judgments.

4.2.2 Internal validity was assessed via internal consistency measures—Cronbach's alpha and the interclass correlation coefficient—with coefficients above 0.7 considered acceptable (Kaftanjieva & Takala, 2002). Inter-rater consistency was evaluated for two judgment tasks: descriptor sorting during familiarization and cut-off score setting during standard setting. The Cronbach's alpha increased from 0.899 in round 1 to 0.981 and 0.985 in rounds 2 and 3 (Table 9). This indicated improved accuracy in placing the descriptors in the second

round. The ICC values showed a strong shared understanding of the common reference levels, with intra-rater ICC at 0.92 and inter-rater ICC at 0.95 (Table 10). Both exceeded the 0.70 threshold (Kaftanjieva & Takala, 2002), indicating higher rater consistency across activities. Taken together, the reliability evidence suggests that the panelists had a solid grasp of the CEFR scales and could reliably evaluate the test items using the relevant descriptors. Thus, their reliable judgments supported the validity of the findings.

Table 9

Inter-rater Consistency in the Standard Setting Judgments

Rounds	Alpha	ICC
Round 1	0.899	0.887
Round 2	0.981	0.981
Round 3	0.985	0.985

5. Discussion

This section discusses the findings, including the final cut-off score ranges, the CEFR target levels for graduate students, and the comparison between CEFR levels and KKU-AELT band scores.

5.1 Ranges of the Final Cut-off Scores

The findings showed varying score intervals for each CEFR level: B2 (29 points), B1 (23 points), C1 (17 points), A2 (16 points), and A1 (15 points) (Table 4). This variation was expected, as the CEFR levels do not represent equal intervals. These findings were consistent with the work of Wudthayagorn (2018), Shin et al. (2022), and Nakanitanon (2021), who also determined cut-off scores for CEFR levels. Wudthayagorn (2018) reported different cut-off ranges: B1 (35 points), B2 (29 points), C1 (22 points), and A2 (21 points). Shin et al. (2022) found the widest score ranges at B2 and B1 levels, followed by C1/C2, A1, and A2. Nakanitanon (2021) similarly found the widest range at B2 and B1, followed by A2, C1, and A1. In these studies, the cut-off score ranges for the lowest and highest

scales (A and C levels) were the smallest, since they were easier to observe and judge than the middle scale. This consequence stemmed from the CEFR's purposive design in order to allow flexibility for local adaptation of the scales, enabling their application across multiple contexts and use in all languages (Council of Europe, 2001).

However, the findings revealed varying score intervals, with the highest standard error of judgment (SEJ) in round 3 observed between B1 and B2 levels. This was likely due to ambiguous CEFR descriptors for reading skills between these levels. The high SEJ between B1 and B2 levels highlights the challenge of defining clear boundaries within the CEFR's language proficiency continuum, given the overlap in reading skills required (Alderson et al., 2006). This overlap requires extensive learning activities for both levels (Harsch & Martin, 2012), making it difficult for panelists to make precise judgments and increasing SEJ rates at this transition.

The diverse score ranges in these tests indicated that advancing between CEFR levels presents varying challenges and requires different levels of effort and time. For instance, in the KKU-AELT reading test, moving from A1 to A2 requires 15 points, while progressing from B2 to C1 requires 29 points. This significant point difference at the B2 level reflects the difficulty of transitioning between these levels. B2 users need to extract information from specialized sources, whereas C1 users must understand lengthy, complex texts and identify nuanced details (Council of Europe, 2001). This shows that language assessments pose different challenges and efforts as test takers progress through proficiency levels.

5.2 CEFR Target Levels for Graduate Students

In line with Thailand's CEFR policy framework, the Office of the Higher Education Commission (2024) mandates that undergraduate students achieve at least a B2 level, while graduate students must advance to a C1 level. As previously mentioned, this study analyzed KKU-AELT reading test scores from 326

candidates, revealing insights into the English proficiency of prospective graduate students. The largest proportion of test-takers was classified at the B1 level (48.46%), followed by A2 (27.00%), B2 (23.00%), and A1 (0.92%); no test-takers achieved the C1 level (0.00%). This highlights a significant gap between current proficiency levels and the OHEC's (2024) target of C1 for graduate students. The prevalence of B1 scores underscores the challenges in bridging the gap to C1, which requires mastering a broader range of skills and competencies (Alderson et al., 2006). Achieving the C1 level requires rigorous commitment and practice, in line with the latest CEFR standards (Office of Higher Education Commission, 2024).

This finding signals the need for crucial attention to action for higher education institutions. It emphasizes the need for curriculum design, instructional materials, an English-speaking environment, English language training programs, and other aspects in order to encourage students to meet this requirement. Such initiatives are essential not only to align with the ambitious CEFR C1 standard set by the OHEC but also to empower graduate students with the linguistic competencies required to succeed academically and professionally in an increasingly globalized context.

5.3 Score Comparison Between the CEFR levels and the KKU-AELT Band Scores

The KKU-AELT uses a seven-score band system (Bands 1–7), where each band corresponds to specific competencies through scaled score ranges (Graduate School of Khon Kaen University, 2022). This detailed band descriptor system provides nuanced insights into the abilities associated with each level. However, when aligning these bands with the CEFR's six-level framework (A1–C2), as delineated in its 2001 publication, certain complexities arise, particularly due to the broad nature of the CEFR levels, especially at the B1 and B2 levels (Harsch & Martin, 2012). The present study's findings illustrate that the CEFR level cut-off scores often overlap with multiple KKU-AELT bands. For example, scores

qualifying for the CEFR B1 level stretch from Band 2 to Band 4. This variance means that test takers with CEFR B1-level scores may be categorized in Bands 2, 3, or 4, reflecting a range of proficiencies within the same CEFR level due to the extensive skill set required at these stages (Council of Europe, 2001).

The broad descriptors for B1 and B2 levels in the CEFR encompass a wide range of skills, making it challenging to precisely map these onto a fixed band score without some overlap (Alderson et al., 2006). The introduction of the CEFR's 2020 version, which includes plus levels such as B1+ and B2+, could refine this alignment process (Council of Europe, 2020). By adopting these updated CEFR scales, the KKU-AELT could offer more distinctions in test takers' proficiency levels. For instance, under the revised CEFR framework, test takers scoring between 55 and 83 based on the final cut-off scores under the present study might be further divided into B2 and B2+ categories, with the latter more accurately identifying those at the stage of advancing to the C1 level. This highlights the need to refine language assessment tools in order to better reflect diverse competencies. Integrating nuanced levels such as B1+ and B2+ can cater to varied abilities, offering clearer progression pathways and targeted learning strategies.

6. Limitations and Recommendations

Despite considerable effort to ensure validity and reliability of the study findings, it is imperative to acknowledge potential limitations. A key limitation is that the study was conducted within a single institutional and national context (Khon Kaen University, Thailand). The panel judgments, the test sample, and the mapping between KKU-AELT scores and CEFR levels, therefore, reflect only local expectations. As a result, the specific cut-off ranges and band correspondences reported here should not be assumed to be able to be generalized automatically to other countries, institutions, or test forms without caution. In addition, the study involved a relatively small group of panelists to determine the cut-off scores, potentially introducing subjectivity. It also focused solely on internal validity, which is essential for comprehensively validating the findings; external validity is also

essential. This would involve recruiting a diverse test-taking population, including those that have taken both the KKU-AELT reading test and an external examination aligned with the same CEFR level.

Despite the constraints, the study offers several recommendations. The study acknowledges that panel size, panelist selection, and the specific test sample may have influenced the findings. Future research should use larger or alternative panels and an independent test sample for comparative analyses in order to strengthen robustness. This study employed the ID matching (test-centered) method in order to establish cut-off scores; exploring alternative standard-setting methods would help to determine whether different approaches would yield distinct cut-off scores. In order to establish external validity, future research may triangulate panel judgements with empirical data (e.g., student performance distributions or outcomes on external CEFR-aligned tests) or conduct cross-validation studies (across different panels, in additional countries) to test the stability of cut-off scores across settings.

7. Conclusion

The present study mapped the KKU-AELT reading test with the CEFR using the standard-setting procedures in the Manual for Relating Exams to the CEFR (Council of Europe, 2009), based on the judgments from a panel of five experts whose consistency and understanding of CEFR descriptors were verified. The panelists' ratings and their grasp of the CEFR showed satisfactory reliability, supporting the validity of the alignment claim as required by the manual. Cut-off scores linked to CEFR levels were established, enabling the KKU-AELT reading test scores to be interpreted in terms of CEFR proficiency levels. These mappings provided actionable feedback on the test takers' English reading and helped the teachers to interpret how learner performance by score range to inform instruction and assessment. The standard-setting procedures are potentially transferable to institutions in comparable contexts (e.g., universities in other non-native English-speaking settings), provided that key validation steps are replicated. The study's

documentation can also benefit test takers, users, and developers and serves as a practical guide for future CEFR alignment projects in similar contexts.

8. About the Authors

Sivakorn Tangsakul is a Ph.D. graduate from the Applied Linguistics Program, Faculty of Humanities and Social Sciences, Khon Kaen University, and is an English language lecturer at Nakhon Phanom University, Thailand. His research interests include language testing and assessment and corpus linguistics. He can be reached at sivakorn_tk@npu.ac.th.

Kornwipa Poonpon is an associate professor of Applied Linguistics in the Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand. Her research interests include second language assessment, corpus linguistics, and technology-enhanced language learning. She can be reached at korpul@kku.ac.th.

9. Acknowledgement

The authors would like to express their gratitude to the Center for English Language Excellence at Khon Kaen University for granting permission to collect the data.

10. Declaration of AI Use

The authors declare that Grammarly, Gemini, and ChatGPT5 were used only to check spelling and grammar in the preparation of the manuscript.

11. 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. https://eprints.lancs.ac.uk/id/eprint/3909/

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

Brunfaut, T. & Harding, L. (2014). *Linking the GEPT listening test to the Common European Framework of Reference*. Taiwan: Language Training and Testing Centre. https://www.lttc.ntu.edu.tw/lttc-gept-grants/RReport/RG05.pdf

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

Council of Europe. (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment: Manual*. Council of Europe, Language Policy Division.

Ferrara, S., Perie, M. & Johnson, E. (2008). Matching the judgmental task with standard setting panelist expertise: The item-descriptor (ID) matching method. *Journal of Applied Testing Technology, 9*(1). 1–20.

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

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. https://doi.org/10.1191/0265532205lt308oa

Fulcher, G. (2010). *Practical language testing*. Hodder Education.

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

Harsch, C., & Martin, G. (2012). Comparing CEF-related and independent writing proficient ratings: An empirical study. *Language Testing, 29*(3), 345–366

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 Standard 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

Kaftandjieva, F. (2004). *Standard setting*. Section B of the Reference Supplement to the preliminary version of the Manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment. Council of Europe.

Kaftandjieva, F., & Takala, S. (2002). *Relating the Finnish Matriculation Examination English Test results to the CEF scales*. Paper presented at the Helsinki Seminar on Linking Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment, Helsinki, Finland.

Knoch, U., & Frost, K. (2016). *Linking the GEPT writing sub-test to the Common European Framework of Reference (CEFR).* Language Training and Testing Centre. https://www.lttc.ntu.edu.tw/lttc-gept-grants/RReport/RG08.pdf

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1–55.

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

Nickerson, R. S. (2004). *Cognition and chance: The psychology of probabilistic reasoning.* Lawrence Erlbaum Associates.

North, B. (2007). The CEFR illustrative scales. *Modern Language Journal, 94*(4), 656–659. http://www.jstor.org/stable/4626092

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. http://old.drs.ac.th/ext/tch_data/tch_ 02.pdf

Office of the Higher Education Commission. (2024). *Policy of upgrading English education standards of higher education institutions*. Office of the Higher

Education Commission, Ministry of Higher Education, Science, Research and Innovation. https://www.ops.go.th/th/ches-downloads/edu-standard/item/9625-2567

O'Sullivan, B. (2015). *Linking the Aptis reporting scales to the CEFR*. Technical Report. https://www.britishcouncil.org/exam/aptis/research/publications/reporting-scales

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. https://doi.org/10.58837/CHULA.PASAA.62.1.1

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. https://doi.org/10.1007/978-3-030-59455-8_3

Rovinelli, R.J. & Hambleton, R.K. (1997). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal of Educational Research, 2*, 49–60.

Shin, S. Y., Sanonguthai, S., & Tangkiengsirisin, S. (2022). *Aligning the CEFR levels with the TU-GET CBT scores* [Technical report]. Thammasat University. https://drive.google.com/drive/folders/1XDcUnxZwBpMaCdtjuZpn8i6bHx489ynM?fbclid=IwAR0UW-09Zu0uwWTYlfzz9etUAXCKk2kcqggCAMAWfKyDWBiAVGpB5q22SA

Srisawat, C. & Poonpon, K. (2023). Revision of an academic English writing rubric for a graduate school admission test. *PASAA, 65*, 234–262. https://doi.org/10.58837/CHULA.PASAA.65.1.9

Tannenbaum, R. J., & Baron, P. A. (2011). *Mapping TOEFL® ITP scores onto the Common European Framework of Reference* (Research Memorandum ETS RM-11-33). ETS.

Tannenbaum, R. J., & Baron, P. A. (2015). *Mapping scores from the TOEFL Junior Comprehensive Test onto the Common European Framework of Reference* (TOEFL Research Report No. RM-15-13). ETS.

Tannenbaum, R. J., & Wylie, E. C. (2005). *Mapping English language proficiency test scores onto the Common European Framework* (TOEFL Research Report No. RR-80). ETS.

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* (TOEFL iBT Research Report iBT-06). ETS.

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. F. (2010). Relating the GEPT reading comprehension tests to the CEFR. In W. Martyniuk (Ed.), *Aligning tests with the CEFR* (*Studies in Language Testing*, Vol. 33, pp. 204–224). Cambridge University Press.

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. https://so04.tci-thaijo.org/index.php/LEARN/article/view/161641