

## Mapping Standardized English Proficiency Test onto the Common European Framework of Reference (CEFR)

Kwanhathai Choedchoo<sup>1\*</sup>, Nutthaporn Owatnupat<sup>1</sup>, and Sutsawad Jandum<sup>1</sup>

<sup>1</sup>Faculty of Humanities and Social Sciences, Suan Dusit University, Thailand

\*Corresponding author, E-mail: choedchoo584@gmail.com

### Abstract

The process of formulating and structuring the Standardized English Proficiency Test involves the development of the Suan Dusit University Standardized English Proficiency Test (SDU-SEPT) in alignment with the Common European Framework of Reference (CEFR). The primary objective of the research was to create a standardized English proficiency test that meets the CEFR standards specifically within the context of Suan Dusit University. To ensure the content validity of the test specifications, a panel of experts was tasked with evaluating them using the Item Objective Congruence (IOC) index. Test topics that received an IOC score exceeding 0.60 were carefully selected for inclusion in the test. The test was structured into two sections, designed as an objective test format. The first section focused on assessing language competency, covering areas such as vocabulary, grammar, and sentence structure. The second section aimed to evaluate language skills, including reading, and listening comprehension, language usage, and functional language, comprising a total of 100 items. In order to measure the quality of the language test, its evaluation was benchmarked against the CEFR global scale at the A2-B2 levels. This calibration provided a standardized framework for assessing the proficiency levels of the test-takers in alignment with internationally recognized language proficiency standards.

**Keywords:** Mapping CEFR, English Proficiency Test, Standardized test

### Introduction

Thailand's Ministry of Education has adopted a policy to enhance the proficiency of English as a means to access global knowledge, align with international standards, and boost Thailand's competitive edge in the future. This directive, outlined in the Office of the Higher Education Commission (OHEC) No. 3/2016 meeting on March 22, 2016, emphasizes rigorous implementation and evaluation. In accordance with the OHEC's mandate to elevate English standards across various levels of higher education institutions, one of the key strategies involves institutions incorporating extracurricular activities for assessing English language skills. These assessments can either align with the institution's standardized test or be deemed appropriate for evaluating English proficiency by referencing international standards, such as the Common European Framework of Reference (CEFR) used by the European Union. This initiative aims to gauge the English language capabilities of each student. Institutions are encouraged to integrate the test results into transcripts or certificates, effective from the academic year 2016.

## Origin of the Study

Dynamic and rapidly changing situations and social contexts have an effect on education arrangement at all levels. The university continues to impose clear directions and guidelines for operations of the university to achieve the goals and objectives that aim to integrate with the new way of life by improving its systems and techniques focus on the 2020 to 2024 period. The focus is on interweaving learning both in and outside the classroom through the online learning system and University learning area. Suan Dusit University initiated this with the announcement in 2020 on the subject of mandating English language proficiency tests for students. This requires students and staff to improve their English language proficiency by doing the TOEIC test that matches the CEFR standard to meet the criteria set by the university and announced by the Thailand Professional Qualification Institute committee. In this regard, the research team study is focused on developing a means to raise and improve students' abilities and improve the student English proficiency level at all levels of study with a low cost to evaluate the student's English language proficiency. Therefore, to accomplish this, the research team has studied and designed a test model (Test Specification). This Test Specification was developed and studied in the context of Suan Dusit University and matched to the Common European Framework of Reference (CEFR). This included preparation for comparing scores on the SDU-SEPT (Standardized English Proficiency Test) as well as developing standards and a cut-off score in preparing for the English language proficiency test (Alderson, 2000).

## Journal of Curriculum and Instruction, CMU

### Research scope

The formulation of a Test Specification played a crucial role in the initial stages of defining and designing a proficiency electronic test. The test in question, known as the Suan Dusit University – Standardized English Proficiency Test (SDU-SEPT), serves as an assessment tool for evaluating the English language skills of students at Suan Dusit University. It aligns with the Common European Framework of Reference (CEFR). The CEFR provides the framework for the test specification, drawing on the expertise of English teaching and the collaborative efforts of university faculty, which includes Thai teachers and native English speakers from both internal and external sources.

This collective effort ensures the incorporation of key features into the test and validates that the test's development adheres to the established test specification. The primary focus is on assessing high competency receptive skills, aligning with the overarching research objective to evaluate English language proficiency for international communication. The study involves the creation of an online test tailored to assess English proficiency within the context of Suan Dusit University, guided by the standards set by the Common European Framework of Reference (CEFR).

### **English language proficiency range level refers to the European Union's International Framework**

The English language proficiency range level is based on the Common European Framework of Reference (CEFR), as utilized by the European Union. The development of the English Proficiency SDU-SEPT test at Suan Dusit University involved collaboration with English teaching experts and native speakers, both internal and external to the university. This collaborative effort aimed to establish the characteristics of the language proficiency test, ensuring alignment with the CEFR standard.

The test focuses on evaluating key English language skills related to receptive skills and English usage as an international language. These skills include reading, listening, and writing, along with grammatical knowledge. The collaborative input from experienced teachers in English courses, linguistics, and related fields was instrumental in crafting a test that meets the specifications. Upon careful consideration, both researchers and experts concluded that the test aligns with the European international standard framework (CEFR). It encompasses levels A2, B1, and B2, as outlined in the Test Specification.

### **Common European Framework of Reference for Languages**

The creation of the Common European Framework of Reference for Languages (CEFR) by the Council of Europe in 2001 marked a significant milestone in the realm of language assessment and proficiency standards. This comprehensive framework was developed to provide a unified and standardized benchmark for evaluating language skills across European nations. By offering a common reference point, the CEFR not only facilitates consistent assessment practices but also promotes greater transparency and comparability in language proficiency evaluation. Language instructors and learners alike benefit from the guidelines outlined in the CEFR, which serve as a valuable tool for understanding and evaluating English proficiency at a functional level. These standardized guidelines enable educators to tailor language learning experiences to meet the diverse needs of learners, while helping learners gauge their own language skills and progress more effectively. Central to the CEFR are the detailed descriptors and proficiency levels articulated in the framework, particularly within the Global Scales designed to classify proficiency in the English language. By aligning language proficiency with these established descriptors, individuals can better articulate and demonstrate their language abilities in various contexts, from educational settings to professional environments.

Moreover, the CEFR holds particularly relevance in the realm of higher education, where the enhancement of English language skills is crucial for academic success and global communication. By providing a framework that is widely recognized and understood, the CEFR contributes to the cultivation of language proficiency in higher education settings, fostering a more integrated and inclusive approach to language learning and assessment.

### English proficiency test development

The English proficiency test is designed to assess individuals' competence in the English language across various personalities. Its focus lies in evaluating the test taker's proficiency level, which in turn determines their ability to use English effectively. Proficiency, in this context, refers to the mastery of the language. Broadly speaking, there are two main formats for the proficiency exam. Firstly, it evaluates the English language skills of individuals who possess sufficient knowledge and abilities to pursue successful careers or further academic studies. Secondly, it assesses whether test takers can meet the required language standards. These proficiency tests serve as valuable tools for evaluating individuals in both academic and professional settings (Hughes, 2020). Orozco et al. (2019) emphasize the importance of thorough preparation and planning in creating an effective English proficiency test that meets the needs of its users. Using language proficiency levels as a foundational framework during test development ensures the test's validity. Consequently, the Common European Framework of Reference for Languages (CEFR) has been adopted as a conceptual framework for crafting the test, aligning with these principles.

The CEFR offers a descriptor that signifies a clear and comprehensive level of proficiency in English, implemented through the Orozco et al. (2019) test. This assessment utilizes the CEFR as its conceptual framework. The initial crucial step involves focusing on the test specification, encompassing tasks, skills, or competencies targeted for evaluation, test types, content resources (text sources), topic areas, and the intended competency level. Additional considerations incorporate test duration, the variety of item types in the exam, the length of the reading section, examination format, and the scoring system used to define and design the features of the Test Specification.

## Methodology

### *Test developing cycle*

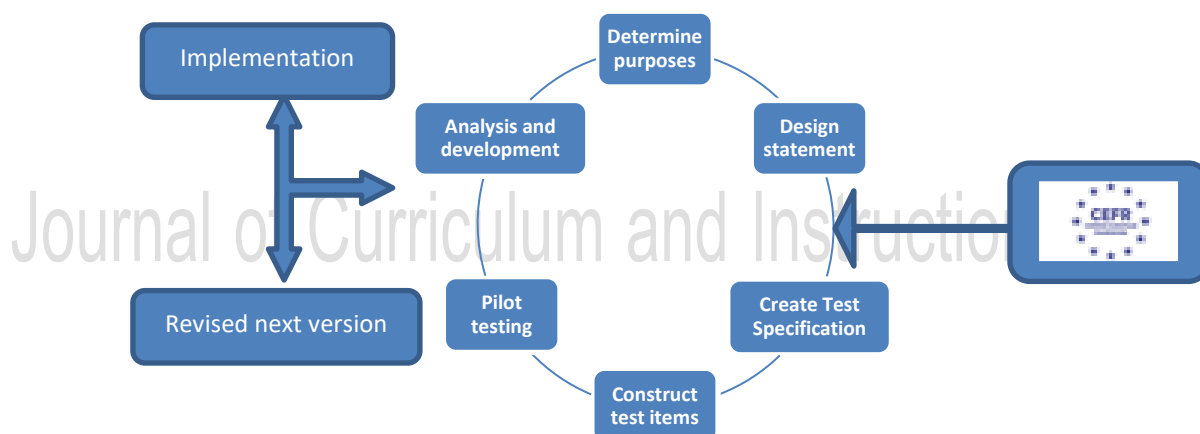
The research project is focused on the creation and development of a comprehensive language proficiency test. The initial phase involves setting clear test objectives and crafting a detailed design statement that encompasses various aspects of the test, including the target test-takers and stakeholders who will benefit from the test outcomes. This stage also involves defining the decisions to be made and the specific areas that will be evaluated as part of the testing process (Bachman & Palmer, 1996; 2010).

Following the establishment of the design statement, the next step is to delineate the test requirements to provide a structured outline of the testing procedure. This includes tasks such as creating a test list, designing the test structure, and establishing a scoring mechanism to evaluate the responses of the test-takers (Fulcher & Davidson, 2007). The test requirements serve as a roadmap for the development and implementation of the language proficiency test.

Depending on the unique design and specifications of the test, experienced language testing developers, equipped with proficiency in this domain, collaborate to either collect or generate a list of prompts that will be used during the testing process. These prompts play a crucial role in assessing the

language proficiency of the test-takers and ensuring that the test accurately measures their language skills based on the defined objectives and criteria.

In summary, the research project involves a systematic approach to creating a language proficiency test, starting from defining objectives and formulating a design statement to outlining test requirements and developing a set of prompts for evaluation purposes. This meticulous process ensures that the test is robust, reliable, and aligns with the intended goals of assessing language proficiency effectively. Following this, a pilot test is imperative before the full-fledged test is implemented. In this study, a pilot test involving 183 participants was conducted through an online system. Post-testing, statistical analyses of the items were performed to identify any problematic elements that require rectification or enhancement in subsequent testing phases. It's crucial to emphasize that the entire test development process is iterative, where tasks are performed in cycles. Consequently, design directives and test requirements must continuously evolve, be well-documented, and undergo constant adaptation and changes to better align with the needs of the stakeholders, as illustrated in the figure below (Sims, 2015).



**Figure 1** Test developing cycle

Source: Orozco & Shin (2019)

### Content Validity of Test Specification of English Proficiency Test

The competency level of a test hinges on the content being assessed. For instance, when gauging interest, the questions within a test or exam should align with topics of interest. Achieving this involves evaluating whether the content of the text effectively mirrors the intended concept. Content precision holds particular significance, especially in appraising academic accomplishments. Any misalignment between the evaluation of teaching and learning outcomes through the test and the covered content is indicative of a discrepancy in the test's content integrity (O'Loughlin, 2013). To assess this, one can scrutinize the test creation process, ensuring it effectively measures the targeted criteria. Alternatively, examining responses to questions, such as observing subsequent behaviors corresponding to the answers, helps gauge the alignment. The verification of the test creation process by content experts is essential. They evaluate

whether the questions in the test accurately represent the content in question. This involves comparing the test content with the expected questions to gauge their consistency. This specific testing method for content validity, conducted by experts, is referred to as the Index of Item-objective Congruence (IOC).

- i. The process of formulating and outlining the Test Specification for the English proficiency test involves the following procedural steps:
- ii. Examining the criteria standards outlined in the European Framework of Reference (CEFR) and understanding its relevance in the context of English language usage and the management of English teaching at Suan Dusit University.
- iii. Defining the test, including optional formats, determining choices, and specifying the English language skills to be evaluated.
- iv. Constructing the test involves creating a Test Specification and conducting trials with a sample group of students based on documents related to the development of the English proficiency test. With the prerequisites outlined in the Test Specification, the research team establishes the procedural steps for crafting the Suan Dusit University English proficiency test, as outlined below:

- i. ***Defining attributes of the test specification.***

The group of experts thoroughly examined the delineations of English proficiency levels as outlined in the Common European Framework of Reference (CEFR) at A2, B1, and B2 levels, establishing the test's objectives. This encompasses language components (structure, vocabulary, concepts, functions) and language skills to assess the examination format, content scope, activity types, number and weighting of questions, testing methods, descriptors, evaluation criteria, scoring, and specifications for each English proficiency level, thereby determining the proficiency spectrum of the SDU-SEPT English Language Proficiency Test.

- ii. ***Meeting the committee to create the test.***

The aim of this phase was to provide a comprehensive overview of the test features and the procedures involved in its execution to the test creation committee. Additionally, it involved a collaborative exploration of the broader context in which students utilize English, encompassing aspects such as the content of teaching and learning, the utilization of educational materials, the management model for teaching and learning within the English language subject group in the general education category, and the assessment and evaluation of students in English Language. The development of tests was guided by the Common European Framework of Reference for Languages (CEFR) (Brown, H. D., & Abeywickrama, P. 2010).

- iii. ***Committee created the test based on the attribute of the test specification.***

The determination of content validity involves researcher-led testing overseen by a committee of experts, specifically evaluating the consistency, measured through the Index of Item Objective Congruence (IOC), of the test with the proficiency levels outlined in the CEFR. This process aims to establish a conformance index, serving as an expert assessment of accuracy (North, B. 2014). The evaluation entails



assessing whether each exam or question aligns with the intended content or learning objectives. This is achieved by utilizing assessment criteria and incorporating expert opinions to determine the correspondence between each question and its purpose or content, as indicated by the Index of Item-Objective Congruence (IOC). The sum of expert review scores (N) is considered, with three experts, each possessing over 10 years of experience at the Ph.D. level, holding positions as Assistant Professors in English Applied Linguistics and teaching English as an International Language. These experts are recognized figures well-versed in the use of English in the educational context of Suan Dusit University.

The IOC Judgment Criteria, with a threshold of 0.60 or higher, indicates that the question set in the Test Specification can be precisely measured for its intended purpose and aligns well with the content. This suggests that the questions can be utilized and further enhanced based on expert advice. In Version 1 of the test, all items surpassed the 0.60 threshold, indicating that the test details meet the required standards according to the results of the study.

**Table 1** SDU-SEPT-Vocabulary Session

20 Items Timing 30 minutes	Vocabulary level at A2-B2, reference profile on Cambridge Dictionary online	12 items	12 tasks	Multiple Choices: a word or phrase which completes the sentence with four choices	Understanding lexical appropriacy in the discrete sentences
	Vocabulary level at A2-B2, reference profile on Cambridge Dictionary online	12 items	12 tasks		

**Table 2** SDU-SEPT- Part 2 Grammar and Structure

20 Items Timing 30 minutes	Parts of speech	8	8	Multiple Choices: an error correction item with four choices	Using English at the word or sentence level, including the use of correct structural words and forms; correct and appropriate words and sentences; variety of forms in
	1.1 Adjectives – comparative, – use of “than” and definite articles	(1 item)	(1 task)		
	1.2 Adverbial phrases of time, place, and frequency – including word order Adverbs of frequency	(1 item)	(1 task)		
	1.3 Conditionals, 2nd and 3 <sup>rd</sup> Connecting words expressing cause and effect, contrast, etc.	(2 items)	(2 tasks)		
	1.4 Connecting words expressing cause and effect, contrast, etc.	(1 item)	(1 task)		



	1.5 Modals – can/could Modals – have to/should	(2 items)	(2 tasks)		expressing similar meaning;
	1.6 Punctuations	(1 item)	(1 task)		
	Tenses	7	7	Multiple Choices: a sentence and ask them to choose which of four alternatives has the same meaning	application of word derivation.
	2.1 Past continuous and Past simple	(2items)	(2 tasks)		
	2.2 Present perfect continuous and Present perfect/past simple	(2 items)	(2 tasks)		
	2.3 Will and going to, for prediction	(1 item)	(1 task)		
	2.4 Reported speech (range of tenses)	(1 item)	(1 task)		
	2.5 2Future perfect	(1 item)	(1 task)		
	3. Active and Passive voices	(2 items)	(2 tasks)	Multiple Choices: a sentence and ask them to choose which of four alternatives	
	Phrasal verbs, extended	(2 items)	(2 tasks)	Multiple Choices: a sentence and ask them to choose which of four alternatives	
	Wish/Would expressing habits, in the past	(2 items)	(2 items)	Multiple Choices: a sentence and ask them to choose which of four alternatives has the same meaning	





**Table 3** SDU-SEPT- Part 3 Reading skill

20 Items  Timing 40 minutes	A longer text with six questions	(1 item)	(6 tasks)	Twenty discrete four-option with	Reading to identify: <ul style="list-style-type: none"><li>• main message</li><li>• purpose</li><li>• detail</li></ul> Showing detailed comprehension of a text
	An Email text with five questions	(1 item)	(5 tasks)	Multiple-choice items.	
	A short text each with 2 questions	(1 item)	(4 tasks)		
	NEWS text each with 5 questions	(1 item)	(5 tasks)		

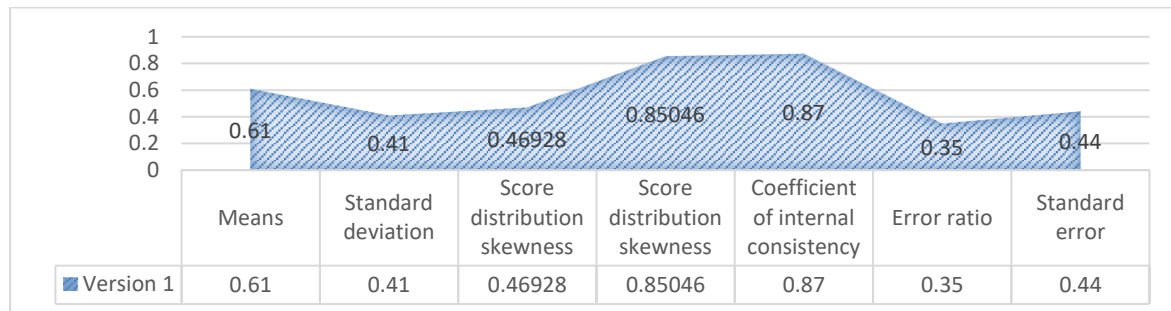
**Table 4** SDU-SEPT- Part 4 Usage and Functional Language

20 Items	Discuss work with a colleague	(1 item)	(5 tasks)	Multiple choices: A word or a phrase, then fill in the blank to complete the conversation with four choices	Synthesizing information in a piece of correct and appropriate extended Spoken language.
Timing	Enquire about and negotiate special treatment (prices and conditions of sale, rates, terms and conditions)	(1 item)	(5 tasks)	Multiple choices: A word or a phrase, then fill in the blank to complete the conversation with four choices	Synthesizing information in a piece of correct and appropriate extended Spoken language.
30	Complain and negotiate redress (poor service e.g. returning faulty, inappropriate or unwanted goods and negotiating for a replacement or refund)	(1 item)	(5 tasks)	Multiple choices: A word or a phrase, then fill in the blank to complete the conversation with four choices	
minutes	Express regrets and negative wishes or intentions	(1 item)	(5 tasks)	Multiple choices: A word or a phrase, then fill in the blank to complete the conversation with four choices	

### SDU-SEPT Testing Quality

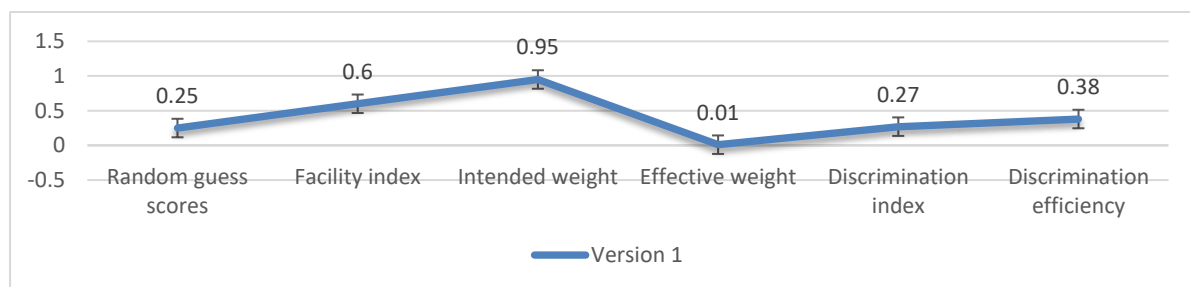
The purpose of utilizing single-question exams is to integrate the test series into the standardization process. Additionally, it involves comparing the scores obtained from the SDU-SEPT test with the Common European Standards Framework of Reference (CEFR). The subsequent steps will delve into the structure of study and development.

Descriptive statistical analysis to determine minimum value, maximum value, distribution, percentage, mean, standard deviation, error ratio, and standard error



**Figure 2** The scores obtained from the SDU-SEPT test

Examining the test items involves assessing fundamental details, identifying challenges, gauging their discriminatory capability, and estimating the overall test's guessing probability. The internal consistency, measured by Cronbach's alpha reliability coefficient, was determined to be  $r = 0.87$ , as stated by Hughes (2003).



**Figure 3** The internal consistency

### Mapping the SDU-SEPT and the criterion Standards (CEFR)

Tannenbaum & Wylie, (2013) said, there are several methods for comparing the scores of the English proficiency test with the standard frameworks. The researchers selected the Modified Angoff method as a suitable method for the Receptive skills test. Wudthayagorn, (2018) used an optional test format (Multiple Choice) that is based on the opinions of 12 language teaching and linguistics experts. There is a consensus on which test-takers will fall into the A2 to B2 grade and whether they can pass the exam and what percentage of their accuracy will be.

### **Panelist**

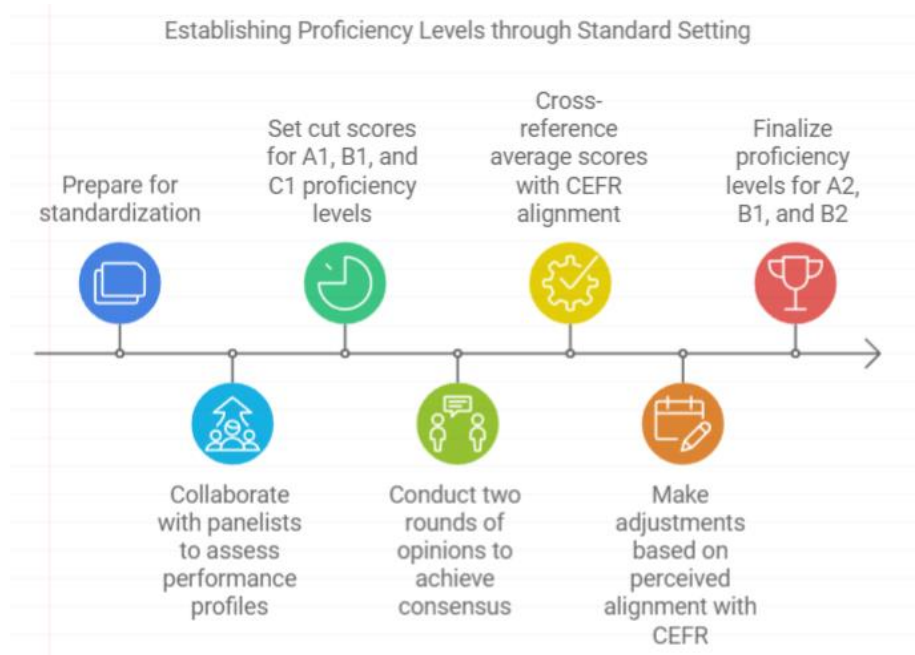
Twelve educators served on the standard-setting panel. Twelve educators reported being either full-time ESL/EFL teachers; they reported having at least 10 years of experience teaching ESL/EFL. As the criteria set for consideration are: faculty member in a Bachelor of Arts program English Language Business English Language and Communication and other related fields and/or has at least 5 years of experience teaching and designing English language tests at a higher education PhD.

### **Preparation**

To compare test scores with the European Framework of Reference for International Standards. The research team adopted the Modified Angoff-Based Predictions of test Items Performance method because it is suitable for the multiple-choice test and only 3 intersection points or cutscores are specified, namely A2, B1, and B2. The researchers describe the overall characteristics of the standard framework of European International Reference (CEFR) Council of Europe. (2001) and an overview of students' English proficiency (Plake, et al. 2000). Suan Dusit University and provide opportunities for discussion and exchange teaching experience of teachers who have taught students of different faculties that different English proficiency is possible. This is to understand the nature and abilities of students of Suan Dusit University and visualization of the lowest competency of that level (Just Qualified Candidate) match. Key questions in the implementation of Modified Angoff-Based Predictions of test scores are compared to CEFR; A2, B1 and B2 levels (Baron & Tannenbaum, 2011; Bonnet, 2007)

### **Standard setting**

The primary objective in standard-setting judgments was for each panelist to assess the performance profiles and determine the total score most likely to correspond to the A2, B1, and B2 proficiency levels. Consistent with the process, the A1, B1, and C1 cut scores were established. Achieving consensus involved two out of three rounds of opinions, and the average score from each test over these rounds was cross-referenced to analyze the alignment with each grade level according to the CEFR. Standardization was a four-step process: 1) Preparation, 2) Panelists' collaboration, 3) Standard setting, and 4) Adjustments based on perceived alignment with the CEFR (Alderson et al., 2006; Athiworakun et al., 2018; Plake et al., 2000; Tannenbaum & Baron, 2008; 2010; 2013; 2015).



**Figure 4** Standard setting

The panelists' standard-setting task was to determine the cut-off scores aligned with the CEFR. This involved estimating what percentage of students with the lowest A2-B1-B2 proficiency, in Round 1, would answer each test item correctly to reach the targeted CEFR levels. The increments ranged from 0, 5, 10, 15, 20, 25, up to 100. Subsequently, the 12 panelists deliberated on the Round 2 results and classification estimates, culminating in a final set of judgments during Round 3. In this phase, they were tasked with making holistic cut-score decisions for the overall test section. Specifically, the 12 panelists reviewed the JQC definitions for A2, B1, and B2 CEFR levels and adjusted their Round 2 A2, B1, and B2 cut scores accordingly (Tannenbaum & Baron, 2008; 2010; 2013; 2015).

**Table 5** Mapping SDU-SEPT onto CEFR level A2 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> round

Round	Ex1	Ex2	Ex3	Ex4	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10	Ex11	Ex12	SD
1=49.25	52.19	48.10	48.24	35.00	59.90	69.65	40.65	44.70	47.90	35.25	41.00	68.45	0.96
2=49.00	51.87	47.51	47.96	35.36	59.79	68.87	39.90	45.45	47.61	35.62	39.82	68.23	0.95
3=48.80	51.79	47.18	47.64	35.26	59.59	68.71	39.70	45.46	47.28	35.52	39.41	68.06	0.95
<b>X=49.02</b>	51.95	47.60	47.95	35.21	59.76	69.08	40.08	45.20	47.60	35.46	40.07	68.25	0.95

According to the results of the SDU-SEPT test at level A2, the initial round revealed a mean cut-off score of 49.25 points with a standard deviation of 0.96 points. The subsequent rounds showed consistency, with the mean for the 2nd round at 49.00 points and the 3rd round at 48.80 points, both accompanied by a standard deviation of 0.95.

In summary, there was a slight decrease in both the mean and standard deviation from the 1st round to the 2nd round and the 3rd round. Expert opinions remained constant, with an overall mean for all three rounds at 49.02 and a standard deviation of 0.95.

**Table 6** Mapping SDU-SEPT onto CEFR level B1 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> round

Round	Ex1	Ex2	Ex3	Ex4	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10	Ex11	Ex12	SD
<b>1=67.51</b>	76.07	63.08	62.72	52.70	58.63	85.01	59.14	78.86	75.67	57.87	58.90	81.54	0.92
<b>2=67.49</b>	76.02	62.90	62.70	52.67	58.56	84.96	59.18	79.01	75.89	57.90	58.67	81.45	0.88
<b>3=67.43</b>	75.98	62.72	62.52	52.65	58.34	84.80	59.23	79.00	76.11	57.93	58.45	81.46	0.88
<b>X=67.84</b>	76.02	62.90	62.64	52.67	58.51	84.92	59.18	78.95	75.89	57.90	58.67	81.48	0.89

As the results of the SDU-SEPT test at level B1, the first round found that the mean cut-off score was 67.51 points and the standard deviation was 0.92 points. The second round was at 67.49 points by the standard deviation was 0.88. The third round was 67.43 points, and the standard deviation was 0.88 respectively. Overall, the mean and the standard deviation decreased and were closed from round 1 to round 2. The third round was at 67.84 points with a standard deviation of 0.89 consistently.

**Table 7** Mapping SDU-SEPT onto CEFR level B2 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> round

Round	Ex1	Ex2	Ex3	Ex4	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10	Ex11	Ex12	SD
<b>1= 81.76</b>	93.70	75.50	78.15	68.84	73.15	91.55	78.70	86.45	76.5	79.75	88.5	90.30	0.67
<b>2= 81.71</b>	93.64	75.36	78.03	68.93	73.28	91.47	78.49	86.31	76.37	79.95	88.39	90.30	0.64
<b>3= 81.67</b>	93.57	75.31	78.01	68.97	73.36	91.38	78.27	86.18	76.33	80.10	88.27	90.31	0.63
<b>X= 81.71</b>	93.64	75.39	78.06	68.91	73.27	91.47	78.49	86.31	76.40	79.93	88.38	90.30	0.62

As the results of the SDU-SEPT test, level B2, the first found that the mean cut-off score was 81.76 points and the standard deviation was 0.67 points. The second round was 81.71 points the standard deviation was 0.64. The third round was 81.71 points by the standard deviation was 0.62 respectively. Overall, the mean and standard deviation decreased from round expert opinions were consistent at 81.67 with a standard deviation of 0.62 consistently.

## Discussion

However, the difficulty in establishing connections is also influenced by the nature of the tests, particularly in terms of how well they align with the Common European Framework of Reference for Languages (CEFR). Tests explicitly designed to correspond to the CEFR are more likely to present fewer challenges in linking compared to tests that adopt a post-hoc approach, as seen in the current situation. Although the tests examined in this research covered the four primary language skills outlined by the CEFR,

the test items and tasks were not specifically crafted to assess these skills in alignment with the CEFR. While this didn't prevent suggesting cut scores for certain levels, it did contribute to the difficulties encountered with the C1 and C2 levels. Emphasizing the use of level descriptors to guide test development can enhance alignment and the potential meaningfulness of cut scores.

## Implications and Limitations

Mapping standardized English proficiency tests onto the CEFR can lead to a better alignment between language assessment practices and internationally recognized language proficiency standards. This can facilitate a more standardized and transparent evaluation of language skills. By aligning English proficiency tests with the CEFR, it becomes easier to compare language proficiency levels across different assessment tools and educational contexts. This can enhance the comparability of language proficiency assessments on a global scale. However, the essential guidelines and procedures outlined in this paper incorporate many of the recommendations found in the literature review for developing a 'single-use' proficiency test, primarily designed as a placement exam. Nevertheless, these guidelines can also serve as a model for creating other proficiency exams. For instance, the language center, where this study was conducted, has been tasked by the university for over a year to produce various 'specific purposes' English exams: exit transfer exams, summer transfer exams, admission entrance exams, extension entrance exams, staff promotion exams, etc. After each administration, these exams are made available online, rendering them non-reusable, and necessitating the creation of new exams each year. The implementation of CEFR-aligned English proficiency tests may vary across different testing organizations and educational institutions. Variability in the interpretation and application of the CEFR descriptors can lead to inconsistencies in assessment outcomes. Consequently, through the strategic adjustment of the difficulty level of the test content to cater to multiple uses, the program has discovered that the methodologies detailed in this paper serve as a highly effective model for the development of language proficiency tests. As highlighted by Baron and Tannenbaum (2011), these outlined procedures have proven instrumental in shaping the creation of language assessment tools that are not only versatile but also align closely with the intended objectives and standards.

## Further research recommendation

In the area of mapping Standardized English Proficiency Tests onto the Common European Framework of Reference (CEFR) could focus on several key aspects to enhance the alignment and effectiveness of language proficiency assessments:

1. Comparative Analysis: Conduct a comparative analysis of existing standardized English proficiency tests with the CEFR levels to identify areas of alignment and discrepancies. This analysis can provide insights into how well current tests reflect the CEFR descriptors and can guide the development of more accurate proficiency assessments.

2. Validation Studies: Carry out validation studies to evaluate the relationship between test scores on standardized English proficiency tests and the corresponding CEFR levels. This research can help establish the validity and reliability of the tests in measuring language proficiency in alignment with the CEFR standards.

3. Item Development: Explore strategies for developing test items that closely align with the language competencies outlined in the CEFR descriptors. Focus on creating test items that accurately assess the language skills specified in the CEFR levels, such as listening, reading, writing, and speaking abilities.

## Conclusion

The discussion highlighted the challenges in establishing connections between test results and proficiency levels, particularly in the context of aligning tests with the Common European Framework of Reference for Languages (CEFR). Tests that are explicitly designed to correspond to the CEFR standards are more likely to facilitate this alignment compared to tests that adopt a post-hoc approach, as observed in the study. While the tests in question covered the primary language skills outlined by the CEFR, the test items and tasks were not specifically tailored to assess these skills in alignment with the CEFR descriptors. This lack of alignment posed challenges, especially in determining cut scores for higher proficiency levels such as C1 and C2. Moving forward, it is crucial to emphasize the use of level descriptors provided by the CEFR to guide test development, ensuring that test items and tasks are designed to assess language skills in a manner that aligns with internationally recognized standards. By enhancing this alignment, future standardized English proficiency tests can provide more meaningful and accurate assessments of test-takers' language proficiency levels in accordance with the CEFR framework.

## References

- Alderson, J. C. (2000). Technology in testing: The present and the future. *System*, 28(4), 593-603.
- Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2006). *Analyzing 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: An International Journal*, 3(1), 3-30.
- Athiworakun, C., Vathanalaoha, K., Thongprayoon, T., Rajprasit, K., & Yaemtui, W. (2018). SWU-SET as a CEFR Standardized English Test. *Journal of Language Teaching and Research*, 9(2), 261-267.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice*. Oxford, UK: Oxford University Press.
- Baron, P. A., & Tannenbaum, R. J. (2011). *Mapping the TOEFL® Junior™ Test onto the Common European Framework of Reference*. *Research Memorandum No. RM-11-07*. Princeton, NJ: Educational.
- Bonnet, G. (2007). *The CEFR and education policies in Europe*. *Modern Language Journal*, 669-672.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices (Vol. 10)*. White Plains, NY: Pearson Education. Council of Europe. Council for Cultural Co-operation.



- Council of Europe. *Council for Cultural Co-operation. Education Committee. Modern Languages Division.* (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Cambridge University Press.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment.* New York: Routledge.
- Hughes, A. (2020). *Testing for language teachers.* Cambridge university press.
- North, B. (2014). *The CEFR in practice.* Cambridge, UK: Cambridge University Press.
- Noijons, J., & Kuijper, H. (2010). *Mapping the Dutch foreign language state examinations onto the CEFR.* Aligning Tests with the CEFR: Reflections on Using the Council of Europe's Draft Manual, 33, 247.
- Office of the Basic Education Commission. (2014). *Ministry of Education's Guidance on Teaching and Learning English Language Reform.* [Online]. Available at: file:///C:/Users/LC09/Downloads/5705140101%20(3).pdf
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363-380.
- Orozco, R. A. Z., & Shin, S. Y. (2019). Developing and Validating an English Proficiency Test. *MEXTESOL Journal*, 43(PUBART).
- Plake, B. S., Impara, J. C., & Irwin, P. M. (2000). Consistency of Angoff-Based Predictions of Item Performance: Evidence of Technical Quality of Results from the Angoff Standard Setting Method. *Journal of Educational Measurement*, 37(4), 347-355.
- Sims, J. M. (2015). A valid and reliable English proficiency exam: A model from a university language program in Taiwan. *English as a Global Language Education (EaGLE) Journal*, 2(1), 91-93.
- Tannenbaum, R. J., & Baron, P. A. (2010). *Mapping TOEIC® test scores to the STANAG 600 language proficiency levels.* Research Memorandum No. RM-10, 11.
- Tannenbaum, R. J., & Baron, P. A. (2015). *Mapping scores from the TOEFL Junior Comprehensive test onto the Common European Framework of Reference (CEFR).* <https://www.ets.org/Media/Research/pdf/RM-15-13.pdf>.
- Tannenbaum, R. J. (2010). *Setting standards on the TOEIC® Writing and speaking assessments for internationally trained nurses.* Research Memorandum No. RM-10-15. Princeton, NJ: Educational Testing Service.
- 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 Research Report Series, 2008(1), i-75.
- Tannenbaum, R. J., & Wylie, E. C. (2013). *Mapping TOEIC® and TOEIC Bridge™ test scores to the Common European Framework of Reference.* The research foundation for the TOEIC tests: A compendium of studies, 2, 6-1.
- 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.