JEL Journal of English Language and Linguistics
Vol. 6 No. 2 (May–August) 2025
ISSN:2730-2431 (Print)
ISSN:2821-952x (Online)

Developing a Standardized English Proficiency Test Based on the CEFR and Language Test Development Manual

Jhonel M. Balleras[1], Chakrit Yippikun[2*], Ruth B. Castro[3], Anuthida Prasertsak[4]
[1]*Christian University of Thailand, Nakhon Pathom, Thailand*
[2]*Thaksin University, Thailand*
[3]*Christian University of Thailand, Nakhon Pathom, Thailand*
[4]*Christian University of Thailand, Nakhon Pathom, Thailand*

APA Citation:
Balleras, J. M., Yippikun, C., Castro, R. B., & Prasertsak, A. (2025). Developing a standardized English proficiency test based on the CEFR and language test development manual. *Journal of Language and Linguistics, 6*(2), 169-189. https://doi.org/10.62819/jel.2025.1007

Received: April 10, 2025
Revised: May 15, 2025
Accepted: June 24, 2025

Abstract

This study was conducted to construct a Council of Europe Common European Framework of Reference-aligned Standardized English Proficiency Test to be reliable and valid. It utilized the instrument development research design participated in by 217 students from one university in Thailand during the second semester of academic year 2024. The actual test material was made based on the Council of Europe's (2011) Manual for Language Test Development and Examining, with its four major stages: planning, designing, try-out, and informing stakeholders. The listening and reading tests were adopted from the existing references and examination utilized by the university. The findings revealed that the English Proficiency Test was valid, with *ds=0.35 for the* Listening Test and *ds=0.35 for the* Reading Test, both at a moderate level. The test reliability result was r=1.35>0.05. Therefore, the test can be considered good and valid since the committee carefully designed and considered the processes of development aligned with the Council of Europe's (2011) Manual for Language Test Development and Examining. An English proficiency test was developed based on the Council of Europe's (2011) Manual for Language Test Development and Examining. Findings showed that the test is valid and reliable, making it a useful tool for measuring English proficiency.

Keywords: CEFR, language test development manual, standardized English proficiency test

Introduction

It is widely known that English is an essential component in intercultural communication. In 2014, the Common European Framework of Reference CEFR (2001) was mainly the reference

* Corresponding author.
E-mail address: chakrit.yip@gmail.com

for creating and crafting a standardized English assessment, which has also been introduced to Thai higher education institutions. Furthermore, the Office of the Higher Education Commission directed policy in enhancing the English standards of Thailand, which was stated in clause 5 of the directive: each higher education institution should require the students to take a standardized English proficiency test aligned to the CEFR standards. In connection with the existing directives and references stated above, the importance of English in today's generation should be focused on for the reason that, according to the English Proficiency Index or English Standard Test last 2022, language enables this connection, and when it comes to connecting beyond borders, English often takes center stage. English research and multinational projects, enjoy media from abroad, travel, engage with new research, and participate in global communities.

According to the study conducted by Wudthayagorn (2022), one of the many educational reforms of Thailand is the strategy on how the university students achieve fluency in the English language. It was also stated that, before the students are qualified for graduation, they must take a standardized English language examination that is aligned to the CEFR standards and results. Despite the Thai government's policy, there remains a dearth of research on implementing English examination and benchmarking in the Thai context. Although this policy has been in effect for several years, it is still unclear how universities have carried out their exit examinations. It was clearly stated that this study was conducted mainly to clarify how Thai public universities have implemented the national English test policy and established benchmarks, given the lack of research on its actual execution and effectiveness.

Christian University decided to construct its own English Proficiency Test to measure the English skills of the students in terms of reading and listening. This examination addresses the needs of the university to comply with the instructions from the above-stated directives of the Office of Higher Education through the initiative of Christian University of Thailand.

Christian University of Thailand English primarily led the development of the test, which was later called CUT-EPT, or Christian University of Thailand English Proficiency Test. Cooperatively, each college participated and cooperated for the success of this study. By carefully analyzing, planning, and studying the entire process of this research, test implementation would be possible and be successful. This task will be having one strong team to develop and execute a standardized-based examination, or the CUT-EPT. The team is needed to ensure that the materials will meet all the requirements of a good test, such as validity, reliability, and practicality.

Despite the growing importance of English proficiency in the academe, the university lacks a localized English proficiency test tailored to students' linguistic experiences and educational needs. Standardized tests often do not reflect the specific challenges faced by local students (Papageorgio et al., 2022; Park et al., 2022). Recent studies also emphasize that contextualized assessment aligned with the curriculum and local context improves validity and student acceptance (Chen, 2024; Hendrick & Smith, 2024). Therefore, this study aims to explore the feasibility of developing a localized English proficiency test suited to the institution's needs.

Literature Review

1. Developing Test Material for English Proficiency Test

Effective English proficiency test development involves a systematic process based on established theories, including test planning, designing, item writing, piloting, reviewing, and revision to ensure validity and reliability (Alderson et al., 2006). Aligning test content with frameworks like the CEFR is essential for meaningful assessment of language ability. Validity ensures the test measures the intended skills, while reliability guarantees consistent results, both often evaluated through statistical analysis during piloting (Messick, 1989; Brown, 2014). Building on these principles, Wudthayagorn (2022) explored how Thai universities implement English exit examinations in line with official guidelines. Cheewasukthawoen (2022) emphasized the CEFR-based stages of test development supported by statistical validation. Sridhanyarat et al. (2021) also stressed the importance of aligning tests to CEFR standards for fairness and consistency. Together, these studies demonstrate the need to integrate theoretical foundations with practical steps to develop valid, reliable, and contextually relevant English assessments.

Developing valid and reliable English proficiency tests requires careful alignment with internationally recognized frameworks such as the Common European Framework of Reference for Languages. Tannenbaum (2024) emphasizes that aligning test content to the CEFR's six-level scale ensures comprehensive assessment across the core language skills of reading, writing, speaking, and listening. This alignment provides clear benchmarks that facilitate accurate measurement of the learner's language proficiency and supports comparability across different contexts.

In addition, Yu (2022) conducted research within an international academic setting, demonstrating the effectiveness of standardized language proficiency descriptions for general English courses. By categorizing learners into advanced, intermediate, and basic proficiency levels, Yu showed that such classifications aid in the reliable implementation and measurement of standardized tests, contributing to their validity and utility in diverse educational environments. The study further discussed pedagogical implications and proposed directions for future research, highlighting the dynamic relationship between test design and instructional practices.

Orozco et al. (2019) studied the authenticity and qualities of language tests, which have been determined by their usefulness in terms of reliability, construct validity, authenticity, impact, and practicality. In addition, Splunder et al. (2022) developed an Interuniversity Test of Academic English linked to CEFR and validated by an independent audit commission. The test was considered a political tool of the government to enforce the language policy. The test was accepted after a few years of its introduction and validated and widely accepted. This research showed the context of the test construction and creation, addressing reliability and validity and, most importantly, the implication of the test, including its pedagogical and societal relevance.

The necessity of a systematic test development process is also well-documented. Gani (2020) highlights the critical role of pilot testing in enhancing test validity and reliability. Pilot testing allows researchers necessary modifications before final administration. This process not only strengthens the quality of the test instrument but also fosters ongoing improvement and adaptation to specific learner populations. Sims (2015) contributes a broader perspective by addressing challenges associated with commercially available language proficiency tests. While many such tests are widely used, Sims points out issues of high costs and limited contextual relevance, which can make them unsuitable for some language programs. Consequently, Sims argues that universities are encouraged to develop their own proficiency tests tailored to their students' needs and institutional goals. However, to maintain quality and credibility, Sims stresses that these locally developed tests should follow standardized procedures and established models for test development. To summarize, these studies underscore the importance of integrating theoretical frameworks with practical test development procedures. By aligning test content to established proficiency standards and rigorously piloting test instruments, institutions can create valid, reliable, and context-sensitive English proficiency assessments that meet both academic and pedagogical demands.

2. Reliability and Validity of English Proficiency Test

A fundamental aspect of language test development is ensuring the validity and reliability of the instrument, as these three qualities determine the accuracy and consistency of the results. Validity refers to whether a test measures what it intends to measure, while reliability indicates the consistency of test scores across different administrations. Brown (2024); Bachman and Palmer (2010). Jing (2029) underscored the importance of construct validity in evaluating English proficiency, demonstrating that correlations with established assessments confirmed the test's acceptability and relevance for language learners. This affirms the critical role of validity in ensuring test results truly reflect learners' competencies.

In support of this, Sugianto (2027) highlighted that a high-quality test should meet specific standards—it must be reliable, valid, objective, practicable, and economical. These characteristics ensure that the test not only serves its academic purpose but also contributes to effective educational decision-making. In his study, the summative test developed met these criteria, reinforcing its role in supporting valid judgments about student performance.

In the same way, Rosaroso (2015) assessed the use of reliability measures in the test validation. The author stated that this is an essential part of test standardization; the test was used as an admission and placement test in English, mathematics, and science proficiency. Validation, revision, and further evaluation need to be done in the process of test development through the help of the examination committee.

Jayanti (2019) measured the validity and reliability of the English Nation Final Examination, which fulfilled the criteria of a good test. Hence, the test developers are expected to focus more of their attention on identifying a good test while in the process of test construction. The author further emphasized that the quality of the test also reflects the quality of education in a specific

place. Furthermore, the educators' skills in constructing or developing tests will be enhanced and developed as well, leading to a motivation to construct and develop more tests in a specific area. To support this, Setiabudi (2019) emphasized the revision and improvement of the standardized test following the result of the reliability and validity test in order to come up with a good test.

Collectively, these studies emphasize that achieving and maintaining test validity and reliability is not a one-time effort but a sustained process involving careful design, evaluation, and refinement. These principles serve as foundational pillars for developing English proficiency tests that are fair, accurate, and educationally impactful.

Conceptual Framework of the Study

The framework of the study demonstrates how we applied the test development processes and reviewed how the researchers determined how far the instrument development process could be covered in one study based on the Council of Europe's (2011) Manual for Language Test Development and Examining. The framework of this study was presented below.

Figure 1
CUT English Proficiency Test Development Process

Planning: In this phase, gathering information, the team initiates a series of meetings and brainstorming in order to have clear ideas and processes of what would be the exact and achievable objectives in developing this test material. The test development team must identify and specify the purpose and design of the test prior to the initial phase of writing or the actual development of the exit examination test. This part also addresses the participants from different programs and courses and their language skills level and competencies. Since this test aims to measure the English proficiency level of the students, alignment to the approaches and strategies of teaching as well as the learning objectives of the lesson was given. The test result was used and aligned to the CEFR descriptors for future interventions to both teaching and learning processes. Lastly, the main objective of the test is to have a standardized test that is owned by the Christian University of Thailand to be used as an exit examination of all students leading to their graduation. In this stage, the name of the test was CUT-EPT, Christian University of Thailand English Proficiency Test, which will be taken by the college undergraduate students. It was anchored with its objective—to assess the level of the student's English proficiency; the basis for placement and other purposes of the test aligned with the cut-off scores aligned from the CEFR for cut-off scores corresponding to A1, A2, B1, B2, C1, and C2 of the CEFR, respectively. It has two major parts: listening and reading comprehension, with a total of 100 items with 1 corresponding point in each item. The test should be conducted in 1 hour and thirty minutes inside the testing venue or on-site administration.

Designing: At this stage, the test development team will now start to think of the total features of the test, such as type, number of items, test duration, number of respondents, and the skills that the focus of the test aims to measure. In addition, the physical appearance and the technicalities of the test will also be studied and finalized. In order to come up with a good and objective test, validity, reliability, authenticity, and practicality will also be planned. In designing the test tool, the team decided to lift and adopt questions from various English references that the English department used in testing the proficiency of the students. The specifications or references will be found on the actual testing tool. It covers the test components and presents the skill, part, focus, points, item placement, and duration of the test. Listening was composed of four major parts: 1 (Photographs), 2 (Question-Response), 3 (Conversations), and 4 (Talks), with 5 points, item placement 1 to 5, and the shortest duration, which is 3 minutes and 40 seconds. The second part, which is the question-response, has 15 points in total, placed on items 6 to 20, having a duration of 7 minutes and 29 seconds. Next is Conversation, which obtained the same points, which is 15, with its item placement 21-35 within 8 minutes and 20 seconds time duration, and Talks was also composed of 15 points, item placement 36-50, and the longest time duration of 9 minutes and 11 seconds. The listening test was completed as the committee agreed to have its total of 50 points to be answered and completed within 30 minutes. The reading comprehension test was composed of three major parts: Part 1, Incomplete Sentences, with 10 points and placed on the item numbers 1 to 10; Part 2, Text Completion, with 16 points and placed on the item numbers 11 to 26; and Part 3, Reading Comprehension, with the highest points of 24 and placed on the item numbers 27 to 50. This part of the test should be completed within 60 minutes depending on the ability of the test takers.

Tryout: In order to measure the quality of the test, pilot testing should be done to know the possible revisions and changes after the first test administration. Additionally, test piloting allows the team to identify the possible recommendations and suggestions in terms of test administration in terms of place of examination and number of students who will take the exam. Consultation with colleagues or stakeholders can be repeated until the final version of test specifications is approved. Up until the final version of the test requirements is accepted, the pilot and consultation with colleagues or stakeholders may be repeated. This stage was administered in 4 different classrooms inside the Christian University of Thailand, which was composed of 217 undergraduate students from different courses and programs. The test takers, or participants, were mixed with different levels, as the committee identified beforehand as beginner, intermediate, and advanced. Before taking the test, the participants were informed 2 weeks before and requested to participate in the study. All participants were officially registered and enrolled during the academic year 2567.

Informing Stakeholders: At this stage, information dissemination should be done in order for those who will be the test takers to be prepared and informed that there will be a need for them to take the English exit examination. At this time, technical features of the test will be introduced and communicated to the prospective test-takers. The committee undertook and planned the careful execution and implementation of the test by conducting a series of meetings and publishing and announcing the conduct and availability of the test. The purpose was

communicated by posting the announcements and publishing all the entire details of the English Proficiency Test on the official website of Christian University of Thailand. Moreover, directing memorandums and official statements from the committee approved by the academic office were released and posted through different means, such as the line group, the official Facebook page of Multidisciplinary College and the University, as well as proper and official communications on the Undergraduate Committee of Christian University of Thailand.

To develop the CUT-EPT, we utilized and patterned the processes on the Manual for Language Test Development and Examining for use with the CEFR. The manual was produced by ALTE on behalf of the Language Policy Division, Council of Europe. Based on the manual, the stages of test development consisted of planning, designing, trying out, and informing stakeholders.

To ensure its validity and reliability, statistical analysis was conducted on both listening and reading comprehension components. The result indicated that the test items were effective in distinguishing between high- and low-performing students, based on the accepted standards of the test evaluation. Both components were categorized at a moderate level, meaning the items were appropriate and could be retained.

Theoretical Framework

Since the goal of this study is to establish the stages in developing the English Proficiency Test of Christian University of Thailand, this research will be aligned to the theory of Language Assessment Literacy as stated by Taylor (2013), which proves that language testing theory, principles, and concepts, as well as technical item-writing skills, are suggested as the most important knowledge areas for those with test-writing responsibilities.

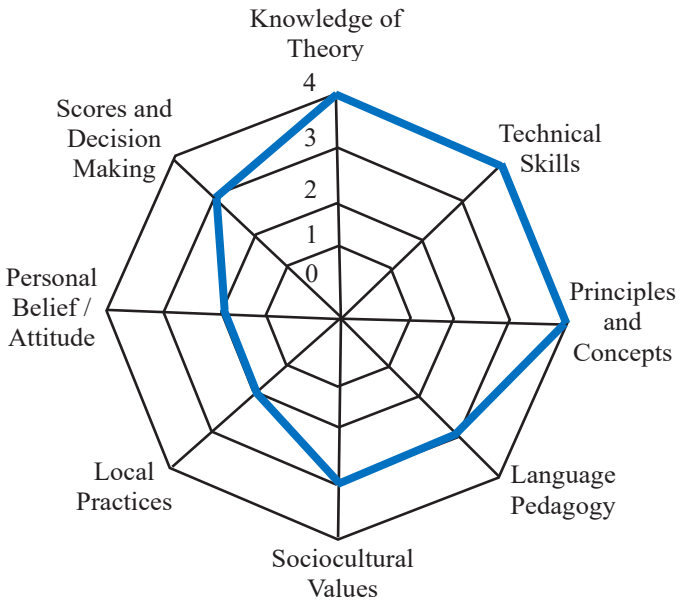
Furthermore, designing and implementing good assessments that reflect the needs of particular social contexts requires the establishment of appropriate criteria and standards for language assessment. In fact, educational programs and resources should be developed to equip the assessment community with the ability to understand and engage in quality assessments.

The figure presented below represents the language assessment literacy processes and principles. Those skills and knowledge are necessary in order to come up with valid and reliable tests because evaluating language is strongly tied to LAL skills and concepts. In particular, because test design, usage, and interpretation of statistics, as well as test grading, are necessary for evaluating language ability, they are considered to be a part of LAL. Principles are seen the same in LAL and assessment literacy; specifically, they pertain to rules of conduct for ethics, fairness, and assessment outcomes.

Looking over Inbar-Lourie's list of LAL's components for language teachers will help us to get a sense of the wide span of the field of study. The author argues that LAL is "a unique complex entity," similar to general assessment literacy for instructors but yet being different from it. The following are the components of LAL for language instructors, according to the author:

1. Understanding of the social role of assessment and the responsibility of the language tester.
2. Understanding of the political and social forces involved, testing power and consequences.
3. Knowing the knowledge of how to write, administer, and analyze tests; report test results; and ensure test quality.
4. Understanding of large-scale test data.
5. Proficiency in Language Classroom Assessment.
6. Mastering language acquisition and learning theories and relating them in the assessment process.
7. Matching assessment with language teaching approaches. Knowledge about current language teaching approaches and pedagogies.
8. Awareness of the dilemmas that underlie assessment: formative vs. summative; internal vs. external; and validity and reliability issues, particularly with reference to authentic language use.
9. LAL is individualized, the product of the knowledge, experience, perceptions, and beliefs that language teachers bring to the teaching and assessment process.

Figure 2
Theoretical Paradigm of the Study



Given the idea and elements of language assessment literacy, it is now in this current study that skills and attributes in validating the reliability of the test are one of the most important concerns of this present study. Knowing the ability of the test validator is very important to come up with the examination to be administered at Christian University of Thailand. The output of this research would be the actual test examination tool, which will soon be called CUT-EPT, or Christian University of Thailand English Proficiency Test.

Research Objectives

To construct a Council of Europe Common European Framework of Reference-aligned Standardized English Proficiency Test to be reliable and valid

Research Questions

1. How was the English proficiency test of Christian University of Thailand developed and validated as a standardized assessment tool aligned to the CEFR based on the process outlined by the Manual for Language Test Development and Examining: planning, designing, try-out, and informing stakeholders?

2. Is the developed English Proficiency Test of Christian University of Thailand valid and reliable?

Methodology

1. Research Design

This study employed instrument development design, a research approach focused on creating and validating a tool or test through systematic procedures. This design is appropriate when the primary goal of the research is to develop a measurement instrument—such as a questionnaire, survey, or test—that is both reliable and valid for its intended purpose. Creswell and Guetterman (2019); Devellis (2016). In this research, the final output is the Christian University of Thailand English Proficiency Test (CUT-EPT), which is specifically designed to assess the English proficiency of students before qualifying for course completion. The instrument development design typically involves several phases, including defining the construct, generating items, conducting expert reviews, pilot testing, analyzing item performance, and refining the instrument based on empirical evidence. (Fraenkel, Wallen, and Hyun (2019). Following these stages ensures that the final tool is contextually appropriate and aligned with the institution's academic goal. Given the localized nature and purpose of the CUT-EPT, this design effectively supports the goal of producing a standardized and reliable proficiency test tailored to the needs of the university. This research was anchored on the IRB certificate number ข.04/2567 issued by the human research ethics committee of Christian University of Thailand.

2. Participants

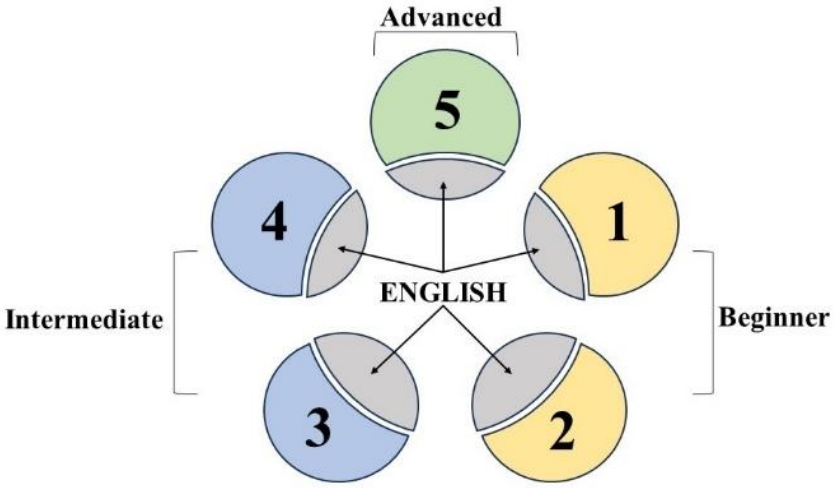
The participants of this study were 217 undergraduate students from various academic programs at Christian University of Thailand, ranging from first-year to fourth-year levels. They were purposively selected based on specific criteria relevant to the study's objectives. To ensure a representative range of English proficiency levels—from beginner to advanced—students were chosen from different year levels and academic disciplines, and only those who had completed or were enrolled in at least one university's English language courses were considered. According to Purnawan (2023), test takers should be composed of beginner to proficient level when it comes to English language skills.

To categorize participants' English proficiency levels, we used the university's English curriculum as a guide. The required English courses include English 1 (listening and speaking), English 2 (reading and writing), English 3 (reading), English 4 (writing), and English 5 (integrated skills), covering these skills. Since all undergraduate students are required to take

these courses as part of their academic program, they served as a natural framework for stratifying participants according to their exposure and skill development in English. Additional demographics such as age, gender, program of study, and year level were collected to provide further context to the findings and ensure a diverse sample.

Figure 3

The Five English Courses of All Courses in Christian University of Thailand



This research categorized the participants according to the level of English course that they are currently taking. Students who are taking English 5 are categorized at the advanced level since all skills in English are being taught to them. In another way, English 3 and 4 are at the intermediate level because the two major skills were discussed and developed with the students in a whole semester, meaning, in this course, mastery and proficiency are a major focus of the course. Lastly, English 1 (Listening and Speaking) and English 2 (Writing and Speaking) are at the beginning level because two skills were discussed and introduced to the students initially as the preparation for the combined skills in English 3, 4, and 5. These English courses were all prerequisite subjects.

3. Research Instrument

The main instrument used in this study is the Christian University of Thailand English Proficiency Test (CUT-EPT), which was developed to assess students' English language abilities through two major components: listening and reading comprehension. The listening sections include a variety of task types, such as photographs, question-response items, conversations, and talks. Each part varies in terms of the number of items and time allocation, but the entire listening test consists of 50 items and is designed to be completed within 30 minutes. The structure of the listening section was intentionally crafted to mirror real-world communication, focusing on the students' overall communicative competence rather than inaccessible test performance. This approach follows the principles outlined by Reynolds

(2021), who emphasized the importance of aligning listening and assessments with actual language use and communicative situations outside the classroom.

The reading comprehension section is composed of incomplete sentences, text completion, and reading passages. Like the listening section, it also includes 50 items, which test takers are expected to complete within approximately 60 minutes. This part of the test evaluates not only grammatical and vocabulary knowledge but also higher-order reading skills necessary for academic success. The inclusion of both listening and reading components was carefully considered by the test development committee to ensure that the CUT-EPT accurately reflects the English proficiency demands placed on the students in academic and real-world contexts. Moreover, the instrument is designed to serve as a valid and reliable tool for determining students' readiness to complete their programs in a globalized learning environment.

The existing references of the English Department were adopted and followed. [1] TOEIC Practice Exams, 4th Edition, Lin Lougheed, Ed. D., Teachers College Columbia University, Barron's Educational Series, Inc. *Library of Congress Control Number: 2018950640, ISBN: 978-1-4380-1182-0,* and [2]Oxford, Preparation Course for the TOEIC Test, New Edition, 2006, *Oxford University Press India, ISBN: 10-0194454355, ISBN: 13-978-0194454355*. Various items and questions were collected, carefully selected, and included on the actual test tool. *Data Collection*

Before collecting the data, the researchers ensure that the Institutional Review Board administers the ethical considerations of the research, and the IRB certification was received by the researchers. Participants were also carefully informed and signed a consent to participate in the research processes. Class-based test administration was done properly and on schedule based on the availability of the participants, the research committee, and the facility as well. Paper-based test papers and answer sheets were requested, granted, and printed by the academic section and properly distributed to the committee and to the participants as well. After the face-to-face test, the answer sheets were checked by the respective academic section, and the result was officially forwarded to the researchers for the succeeding processes, like measuring the reliability and validity of the test.

4. Data Analysis

The index of difficulty for each test item was analyzed using Microsoft Excel. This involved calculating the proportion of students who answered each item correctly. In addition, SPSS was used to validate and support the analysis by generating item-wise descriptive statistics. The goal was to ensure each item's appropriateness for the test takers' proficiency levels.

To calculate the index of discrimination for each item, we divided the students into upper and lower groups based on the total scores. The number of correct answers in each group was analyzed. The discrimination index was computed using Microsoft Excel. SPSS was also used to cross-check group performances for accuracy.

Test of Reliability and Validity of the Test. In order to test the reliability of the test, the split-half method was done by administering the test once, and the results were broken down into halves by "odd-even" division. The scores of the students in the "odd" and "even" items were used in computing the reliability using the Pearson Product Moment of Correlation.
For the teacher-made test, a reliability index of 0.50 and above is acceptable. Reliability was determined through the Pearson correlation coefficient, and the analysis was performed using SPSS to assess the internal consistency of the test items. A high reliability coefficient indicates that the test items consistently measure the same construct. While the content validity was ensured through expert review by a language assessment specialist who evaluated the alignment of each item with the CEFR framework and the course outcomes of the university. In addition, construct validity was supported by conducting item analysis using Excel and SPSS, focusing on the item difficulty and discrimination indices. This analysis ensured that each test item appropriately measured English language proficiency.

Results

1. Stages of Test Development from Council of Europe's (2011) Manual for Language Test Development and Examining

Stage 1: Planning

This stage was started last August 2023 when the English Section Team conducted the first meeting and agreed to evaluate and realign the existing English Proficiency Test of Christian University of Thailand. The table below shows the CUT-EPT specifications.
The data presented below in table 2 were the finalized and agreed test specifications of the team based on the Manual for Language Test Development and Examining and based on the decision of the committee.

According to Irwing (2018), specification of the test should be made beforehand, including the whole process of test development from the specification of a test need and construct definition, through item generation and scale development, to preparing a user manual. It was supported by Khanal (2020), who presents key considerations while constructing, scoring, and analyzing test items that serve one of the major requirements of teaching both at the school and university levels. Specifically, this article provides major issues to be considered in a full cycle of testing, starting from the preparation of test objectives and specification tables to the step of analyzing the effectiveness of each item, the process commonly called item analysis.

In this research, the committee decided on the above-stated test specifications. To further discuss, from the name, which is the CUT-EPT Christian University of Thailand English Proficiency Test, this was aligned on the underlying principles of the university to have its own proficiency examination. Primarily, the committee considered the formulation of the test objectives, which was to mainly assess the proficiency level of the undergraduate students. After identifying the objectives, scores were identified and aligned from the CEFR for cut-off scores corresponding to A1, A2, B1, B2, C1, and C2 of the CEFR, respectively. This pointing system will help the committee with how they are going to categorize the students based on the

test results. The two major skills were focused on listening and reading comprehension, as stated by Elleman (2019).

Table 2
CUT-EPT's Specifications

Specifications	Decision
Name of the Test	CUT-EPT Christian University of Thailand English Proficiency Test
Test Takers	Undergraduate Students
Objectives	To assess the level of the student's English proficiency as a basis for placement and other purposes of the test
Cut-off scores aligned from the CEFR	For cut-off scores corresponding to A1, A2, B1, B2, C1, and C2 of the CEFR, respectively.
Test's Contents	Listening and Reading Comprehension
Test's format and Item #	100 multiple-choice items with four options in each item
Test Duration	1 hour and thirty minutes
Test's Administration	Face-to-face with answer sheet and test paper
Scoring	1 point for the correct answer and zero for the wrong answer
Test Completion date	June 2024

Improving reading scores will require a concerted and collaborative effort by researchers, educators, and policymakers with a focus on long-term solutions. An early and sustained focus on developing background knowledge, vocabulary, inference, and comprehension monitoring skills across development will be necessary to improve comprehension. The test's duration is 1 hour and 30 minutes; the listening test duration is 30 minutes, and reading comprehension is 1 hour. This is in accordance with the study conducted by Cheewasukthaworn (2020), who stated that the 100 items were also believed to fit the test duration of one hour and thirty minutes, which was viewed by the committee as proper because this duration could help minimize the test-takers' anxiety and fatigue from doing the test. Since the university is now considering a full face-to-face class, test administration was in the classroom set up with its answer sheet and test paper, respectively.

Stage 2: Designing

Considering the test specification stated and presented above, the committee decided to divide the proficiency test into two major parts: listening and reading comprehension. This is aligned to the principles of the TOEIC examination, which was mainly the measurement of the English tool of the university before the students get qualified for graduation. Since the committee was also considering the alignment of the test to the CEFR, the scores and cut-offs that were

presented in table 1 were also given importance. That's why the details below were resolved by the committee as the main components of the CUT-EPT.

Table 3

CUT-EPT Components

Skill	Part	Focus	Points	Item Placement	Duration
Listening	I	*Photographs*	5	1-5	3 min & 40 sec
	II	*Question-Response*	15	6-20	7 min & 29 sec
	III	*Conversations*	15	21-35	8 min & 20 sec
	IV	*Talks*	15	36-50	9 min & 11 sec
		Total	50	50 items	30 minutes
Reading	I	*Incomplete Sentences*	10	1-10	60 minutes
	II	*Text Completion*	16	11-26	
	III	*Reading Comprehension*	24	27-50	
		Total	50	50 items	60 minutes
Overall Total			100	100 items	90 minutes

Based on table 3, listening was composed of four major parts: 1 (Photographs), 2 (Question-Response), 3 (Conversations), and 4 (Talks); however, only the photograph's part has the least number of items, which is 5 points, item placement 1 to 5, and the shortest duration, which is 3 minutes and 40 seconds. The second part, which is the question-response, has 15 points in total, placed on items 6 to 20, having a duration of 7 minutes and 29 seconds. The third part, which is Conversation, obtained the same points, which is 15, with its placement 21-35 within 8 minutes and 20 seconds time duration. Lastly, the final part of the listening test, which is Talks, was also composed of 15 points, item placement 36-50, and the longest time duration of 9 minutes and 11 seconds. Generally, the listening test was completed as the committee agreed to have its total of 50 points to be answered and completed within 30 minutes. The importance of including these skills as one of the major skills in learning English was aligned to the principles of Reynolds (2021), who stated and emphasized that while teachers tended to draw on textbook listening and speaking activities to assess those skills, how they graded students focused heavily on the students' communicative competence as listeners and speakers of English rather than on their ability to answer comprehension questions correctly in the classroom assessments. Students identified a mismatch between classroom instruction and assessments and also a mismatch between the English used in assessments and the English used in real-world communication. The committee also believed that having this skill to be learned and mastered by the students will definitely help them to explore and take the chances that they will have in the future, not only here in Thailand, but also in other countries.

On the other hand, the reading comprehension test was composed of three major parts: Part 1, Incomplete Sentences with 10 points and placed on item number 1 to 10; Part 2, Text Completion with 16 points and placed on the item number 11-26; and Part 3, Reading Comprehension with highest points of 24 and placed on the item number 27 to 50. This part of the test should be completed within 60 minutes depending on the ability of the test takers. In

the same consideration of the inclusion of the reading comprehension test of Christian University of Thailand English proficiency test, the committee and this research were believed that reading proficiency has an indirect effect on the employment of strategy use through attitude. The main significance of this research is that the model we have developed has proved to be valid for each year in the sample and for the future of the students in this globalization era.

Stage 3: Tryout

This stage was mainly focused on the actual test administration, as the Manual for Language Test Development and Examining suggests. The tryout test was administered in 4 different classrooms inside the Christian University of Thailand, which was composed of 217 undergraduate students from different courses and programs. The test takers, or participants, were mixed with different levels, as the committee identified beforehand as beginner, intermediate, and advanced. Before taking the test, the participants were informed 2 weeks before and requested to participate in the study. All participants were officially registered and enrolled during the academic year 2567. The committee conducted an orientation before the test was administered wherein the objectives and purpose of the test were communicated to the participants. Processes and procedures were mentioned during the orientation, such as the participants should finish part 1 first before taking the part 2, the rules and regulations for taking the test, the duration of the test, the pointing system, and the guidelines for computing the result of the test aligned to the CEFR descriptors and level.

It was revealed that the duration of the test, which is 1 hour and 30 minutes, is enough and appropriate for the students since most of them finished the test 5 to 10 minutes before the set time duration of the committee. After the test, an item analysis was made by the committee using the statistical tools, and the result revealed that the test was a reliable instrument in determining the English proficiency level of the students; however, the results on the item discrimination still need to be improved.

Stage 4: Informing Stakeholders

The committee started and planned the careful execution and implementation of the test by conducting a series of meetings and publishing and announcing the conduct and availability of the test based on its purpose, which was communicated by posting the announcements and publishing the entire details of the English Proficiency Test on the official website of Christian University of Thailand. Moreover, directing memorandums and official statements from the committee approved by the academic office were released and posted through different means such as the line group, the official Facebook page of Multidisciplinary College and the University, as well as proper and official communications on the Undergraduate Committee of Christian University of Thailand.

2. Validity and Reliability of the English Proficiency Test of Christian University of Thailand

Table 4
Summary of Discrimination of the Test Items

Skills	Total Number Of Items	Summary No. of Students who got the items right		Summary of Proportion		Summary of Index		Item Category	Remarks
		Upper	Lower	PU	PL	DF	DS		
Listening	50	5.47	4.14	0.53	0.05	0.51	0.59	Moderate	Retained
Reading Compre-hension	50	5.38	5.2	0.36	0.04	0.34	0.40	Moderate	Retained

Table 4 presents the summary discrimination of the test with two major components, which are listening and reading comprehension. It was revealed that the 100-item test was good, considering the result of the statistical formula in determining the index of discrimination was by Diederich (1964). Listening skill obtained; summary of students who got the items right: upper with 5.47; lower with 4.14; proportion summary with 0.53 (upper); proportion summary with 0.05 (lower); with 0.51 index of difficulty and 0.59 index of discrimination. The item category fell to a moderate level and can be retained in the test. On the other hand, reading comprehension was obtained: a summary of students who got the items right, upper with 5.38; lower with 5.2; proportion summary with 0.37 (upper); proportion summary with 0.04 (lower); with 0.34 index of difficulty and 0.40 index of discrimination. The item category fell to a moderate level and can be retained in the test, proving that the test is valid.

Table 5
The Reliability of the English Proficiency Test (CUT-EPT)

Indicator	
Reliability of the Half Test	
Number of Cases	217
Sum of XY	6355
Sum of x^2	59372
Sum of y^2	43043
Computed $r\frac{1}{2}\frac{1}{2}$	11.95
Reliability of the Whole Test	
Computed $r\frac{1}{2}\frac{1}{2}$	1.845
Statistical Validity	
Computed $\sqrt{rtt}$	1.35
Test of Significant Reliability	
Computed t-test result	0.90
Tabular t-test result at 0.05	
0.025	
0.01	
0.005	
Decision on Hypothesis	Accepted
Level of Significance	0.005

Table 5 below presents the reliability of the Christian University of Thailand English Proficiency Test. The result of 1.845 denotes that the test is a reliable instrument in measuring the English proficiency level of the undergraduate students at Christian University of Thailand.

Discussion

According to the research objectives and research questions, two topics are discussed in this section: the stages of test development and the validity and reliability of the test.

1. Stages of Test Development from Council of Europe's (2011) Manual for Language Test Development and Examining

The stage 1 results, accompanied by the name of the test, were from the CUT-EPT Christian University of Thailand English Proficiency Test, which will be taken by the college undergraduate students. It was anchored with its objective—to assess the level of the student's English proficiency; the basis for placement and other purposes of the test aligned with the cut-off scores aligned from the CEFR for cut-off scores corresponding to A1, A2, B1, B2, C1, and C2 of the CEFR, respectively. It has two major parts: listening and reading comprehension, with a total of 100 items with 1 corresponding point in each item. The test should be conducted in 1 hour and thirty minutes inside the testing venue or on-site administration.

Stage 2, which covers the test components, presents the skill, part, focus, points, item placement, and duration of the test. Listening was composed of four major parts: 1 (Photographs), 2 (Question-Response), 3 (Conversations), and 4 (Talks), with 5 points, item placement 1 to 5, and the shortest duration, which is 3 minutes and 40 seconds. The second part, which is the question-response, has 15 points in total, placed on the items 6 to 20, having a duration of 7 minutes and 29 seconds. Next is Conversation, which obtained the same points, which is 15, with its item placement 21-35 within 8 minutes and 20 seconds time duration, and Talks was also composed of 15 points, item placement 36-50, and the longest time duration of 9 minutes and 11 seconds. The listening test was completed as the committee agreed to have its total of 50 points to be answered and completed within 30 minutes. The reading comprehension test was composed of three major parts: Part 1, Incomplete Sentences with 10 points and placed on the item numbers 1 to 10; Part 2, Text Completion with 16 points and placed on the item numbers 11 to 26; and Part 3, Reading Comprehension with the highest points of 24 and placed on the item numbers 27 to 50. This part of the test should be completed within 60 minutes depending on the ability of the test takers.

Stage 3, which is the try-out test, was administered in 4 different classrooms inside the Christian University of Thailand, which was composed of 217 undergraduate students from different courses and programs. The test takers, or participants, were mixed with different levels, as the committee identified beforehand as beginner, intermediate, and advanced. Before taking the test, the participants were informed 2 weeks before and requested to participate in the study. All participants were officially registered and enrolled during the academic year 2567.

Lastly, Stage 4: the committee undertook and planned the careful execution and implementation of the test by conducting a series of meetings and publishing and announcing the conduct and availability of the test based on its purpose. This was communicated by posting the announcements and publishing the entire details of the English Proficiency Test on the official website of Christian University of Thailand. Moreover, directing memorandums and official statements from the committee approved by the academic office were released and posted through different means, such as the Line group, the official Facebook page of Multidisciplinary College and the University, as well as proper and official communications on the Undergraduate Committee of Christian University of Thailand.

2. Validity and Reliability of the English Proficiency Test of Christian University of Thailand

The summary discrimination of the test with two major components, which are listening and reading comprehension. It was revealed that the 100-item test was good, considering the result of the statistical formula in determining the index of discrimination was by Diederich (1964). Listening skill obtained; summary of students who got the items right: upper with 5.47; lower with 4.14; proportion summary with 0.53 (upper); proportion summary with 0.05 (lower); with 0.51 index of difficulty and 0.59 index of discrimination. The item category fell to a moderate level and can be retained in the test. On the other hand, reading comprehension was obtained: a summary of students who got the items right, upper with 5.38; lower with 5.2; proportion summary with 0.37 (upper); proportion summary with 0.04 (lower); with 0.34 index of difficulty and 0.40 index of discrimination. The item category fell to a moderate level and can be retained in the test, proving that the test is valid. The reliability of the Christian University of Thailand English Proficiency Test. The result of 1.845 denotes that the test is a reliable instrument in measuring the English proficiency level of the undergraduate students at Christian University of Thailand.

Recommendations

Based on the findings and conclusion of the study, this research recommends that the English Proficiency Test developed by the Christian University of Thailand be officially adopted as a standardized assessment tool for measuring the English proficiency of the undergraduate students. The test demonstrated reliability and validity to make it a strong foundation for institutional language evaluation. To further enhance its effectiveness and alignment with international standards, future research should focus on refining the test items based on CEFR descriptors, conducting related studies to evaluate the consistency of the test results over time, and expanding the sample size to include students from other faculty and academic levels.

Conclusions

The development and validation of the Christian University of Thailand English proficiency test (CUT-EPT) demonstrate a systematic and research-based approach aligned with the CEFR (2011) Manual for Language Test Development and Examining. The test, designed specifically for undergraduate students, underwent rigorous stages from initial planning, detailed test design, piloting and stakeholders' communication. The test structure, which focused on

listening and reading comprehension, reflects the competencies outlined by the CEFR, ensuring relevance, clarity, and applicability in academic and real-world contexts. The item analysis through evaluation of difficulty and discrimination indices, confirmed that the majority of the items were at moderate level and suitable for retention. Moreover, the strong reliability scores indicates that the CUT-EPT is a dependable tool for measuring English proficiency among university students. The process also emphasized the importance of collaboration among test evaluators and stakeholders in achieving clarity and standardization. Overall, the CUT-EPT not only provides a valid and reliable measure of English proficiency but also serves as model for other institutions aiming to design localized, CEFR-aligned language assessment tailored to the learners' contexts and needs.

References

Abd Gani, N. I., Rathakrishnan, M., & Krishnasamy, H. N. (2020). A pilot test for establishing validity and reliability of qualitative interview in the blended learning English proficiency course. *Journal of Critical Reviews*, *7*(5), 140–143. https://doi.org/10.31838/jcr.07.05.23

Cheewasukthaworn, K. (2020). Developing a standardized English proficiency test in alignment with the CEFR. *Journal of Language Teaching and Learning in Thailand*, *63*, 66–92. https://doi.org/10.58837/CHULA.PASAA.63.1.3

Chen, Y. (2024). Rethinking contextualized English assessment: A case of reading and listening in China's university entrance exams. *Language Assessment Quarterly*, *21*(1), 1–21. https://doi.org/10.1080/15434303.2023.2285164

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

EF Education First. (2022). *EF English proficiency index: A ranking of 111 countries and regions by English skills*. https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2022/ef-epi-2022-english.pdf

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (4th ed.). Sage Publications.

Elleman, A. M., & Oslund, E. L. (2019). Reading comprehension research: Implications for practice and policy. *Policy Insights from the Behavioral and Brain Sciences*, *6*(1), 3–11. https://doi.org/10.1177/2372732218816339

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education* (10th ed.). McGraw-Hill Education.

Hendricks, R., & Smith, A. (2024). Designing curriculum-aligned English proficiency tests: Lessons from the ALLTest project in Malaysia. *Asian EFL Journal*, *26*(2), 45–67. https://doi.org/10.55545/aej.v26i2.1053

Irwing, P., & Hughes, D. J. (2018). Test development. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 1–47). Wiley. https://doi.org/10.1002/9781118489772.ch1

Jayanti, D., Husna, N., & Hidayat, D. N. (2019). The validity and reliability analysis of English national final examination for junior high school. *Voices of English Language Education Society*, *3*(2), 127–135. https://doi.org/10.29408/veles.v3i2.1551.g929

Jing, X. (2019). The reliability and validity of language proficiency assessments for English language learners. *Frontier of Higher Education*, *1*(1), 36–42. https://doi.org/10.36012/fhe.v1i1.893

Khanal, P. (2020). Key considerations in test construction, scoring and analysis: A guide to pre-service and in-service teachers. *International Journal of Research Studies in Education*, *9*(5), 15–24. https://doi.org/10.5861/ijrse.2020.5027

Orozco, R. A. Z., & Shin, S. Y. (2019). Developing and validating an English proficiency test. *MEXTESOL Journal*, *43*(3), 1-11.

Oxford University Press. (2006). *Preparation course for the TOEIC test* (New ed.). Oxford University Press.

Papageorgiou, S., Wu, S., So, E., & Wu, J. (2022). Aligning a global English language test to a local scale: Validity, methodology, and implementation. *Language Testing*, *39*(3), 404–426. https://doi.org/10.1177/02655322221076377

Park, J. Y., Harding, L., & Shin, D.-S. (2022). Developing a locally situated academic listening test using unscripted videos: A needs-based approach. *Assessing Writing*, *52*, 100600. https://doi.org/10.1016/j.asw.2022.100600

Purnawan, A., Nurharjanto, A. A., & Ilmi, A. N. (2023). Problems faced by English teacher candidates in developing test kits for assessing students' learning. *Script Journal: Journal of Linguistics and English Teaching*, *8*(2), 215–225. https://doi.org/10.24903/sj.v8i2.1441

Rosaroso, R. C. (2015). Using reliability measures in test validation. *European Scientific Journal*, 11(18).

Setiabudi, A., Mulyadi, M., & Puspita, H. (2019). An analysis of validity and reliability of a teacher-made test. *Journal of English Education and Teaching*, *3*(4), 522–532. https://doi.org/10.33369/jeet.3.4.522-532

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. https://doi.org/10.6294/EaGLE.2015.0102.04

Sridhanyarat, K., Pathong, S., Suranakkharin, T., & Ammaralikit, A. (2021). The development of STEP, the CEFR-based English proficiency test. *English Language Teaching*, *14*(7), 95–106. https://doi.org/10.5539/elt.v14n7p95

Sugianto, A. (2017). Validity and reliability of English summative test for senior high school. *Indonesian EFL Journal: Journal of ELT, Linguistics, and Literature*, *3*(2), 22–38.

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. https://doi.org/10.1002/j.2333-8504.2008.tb02120.x

Lougheed, L. (2018). *TOEIC practice exams* (4th ed.). Barron's Educational Series.

van Splunder, F., Verguts, C., De Moor, T., & De Paepe, S. (2022). The interuniversity test of academic English (ITACE) assessing lecturers' English proficiency in Flanders. *Journal of English-Medium Instruction*, *1*(2), 255–274. https://doi.org/10.1075/jemi.21007.van

Wudthayagorn, J. (2022). An exploration of the English exit examination policy in Thai public universities. *Language Assessment Quarterly*, *19*(2), 107–123. https://doi.org/10.1080/15434303.2021.1937174

Yu, L. T., Chen, M. C., Chiu, C. W., Hsu, C. C., & Yuan, Y. P. (2022). Examining English ability-grouping practices by aligning CEFR levels with university-level General English courses in Taiwan. *Sustainability*, *14*(8), 4629. https://doi.org/10.3390/su14084629

Yu, M. H., Reynolds, B. L., & Ding, C. (2021). Listening and speaking for real-world communication: What teachers do and what students learn from classroom assessments. *SAGE Open*, *11*(2). https://doi.org/10.1177/21582440211009163