

# A Development of English Proficiency Test of Undergraduate English Major Students of University of Phayao

Sanit Yeunsak<sup>1</sup>, Chuanpit Sriwichai<sup>2</sup>, Chanapa Duangfai<sup>3</sup> and

Narisa Paichareon<sup>4</sup>

Received: August 11, 2022  
Revised: September 08, 2022  
Accepted: September 14, 2022

(saint9563@gmail.com)

---

## Abstract

This study aimed to develop English proficiency tests for undergraduate English major students of School of Liberal Arts, University of Phayao based on CEFR C1 level and to evaluate the quality of the developed tests. Pursuing the research objectives, two versions of the tests had been developed and conducted. Each version of the test consisted of 60 items, and it was divided into 4 parts including vocabulary, reading, grammar and writing, and conversation. There were 15 items in each part. The samples of this study were 174 fourth-year English major students of the School of Liberal Arts, University of Phayao in academic year 2021. The content validity of the two tests was verified by the 3 experts in the fields of English language and Education Evaluation and Assessment using Index of Item-Objective Congruence (IOC). Then the tests were tried out with the samples to analyze difficulty and discrimination based on the 27 % technique of Chung Teh-Fan indicating that the criteria for an index of difficulty should be between 0.20 and 0.80 and the index of discrimination should be at least 0.20. Next, the tests were analyzed for reliability using KR-20 formula.

**The research results were as follows;**

---

<sup>1,2,3,4</sup> English Department, School of Liberal Arts, University of Phayao

1. For the first objective of the study; To develop University of Phayao Standard Test, the study resulted in the two versions of C1 Standard Test of English Proficiency which consist of 60 items in four parts; vocabulary, reading, grammar and writing and conversation respectively for the fourth years students in University of Phayao.

2. For the second objective of the study: evaluating the quality of University of Phayao standard test of English proficiency C 1 level, the index of item objective congruence (IOC) which are satisfied with the mean level at over 0.50 expressed that version one of the tests ranging from 0.33 to 1.00, was at 0.94 of the IOC indexes mean, followed by the IOC index means of version two ranging from 0.67 to 1.00 at 0.99. Both versions reflected the high content validity. In terms of the standardized quality of the difficulties and the discrimination of the test, there were 57 qualified items in the first version of the tests, and all items are qualified in the second one. The reliability index of version one of the English proficiency test was at 0.82 and the value of version two was at 0.83 indicating that both tests reached acceptable reliability value.

**Keywords:** English Proficiency Test, Undergraduate Students, English, University of Phayao

## Introduction

An important mean in using English in communication is undeniable in the era of globalization. In Thailand, English is considered as a core subject in all levels of curriculum from primary education to higher education. However, the English language proficiency scores of Thai citizens are still at a low level. The report from the National Institute of Educational Testing Service (NIETS) showed that the average scores of the English subject in the

Ordinary National Educational Test (O-NET) across all levels of basic education have remained low. (NIETS as cited in Wudthayagorn, 2018). Moreover, English proficiency of Thai students did not match the standard required when they took the English proficiency test (Jin, 201).

To enhance the quality of education in Thailand to reach the same level as other developing countries, the Ministry of Education formulated the educational policy in 2016 stating that The Common European Framework of Reference for Languages or CEFR should be introduced for all schools and academic institutions across the country, including those in the level of higher education. Before graduating in each academic level, the students need to surpass the English proficiency standard enforced. It is expected that the English skills of students at least in a level of communication in professional context contributes to the quality of education in Thailand. Hence, all Thai Universities that are responsible in teaching and assessing English proficiency of undergraduate students have realized on this issue. Thai Universities have conducted various English proficiency tests for their undergraduate students in order to match the requirement suggested by the Common European Framework of Reference (CEFR) that has been introduced to implemented in Thailand by The Office of the Basic Education Commission. The important aim of implementing this CEFR policy is to set an achievement benchmark for Thai students, indicating that students graduating from grade 6 should achieve an English proficiency level of at least A1, grade 9 of A2, and grade 12 and vocational college of B1.

University of Phayao conformed fully with the policy enforced by the Ministry of Education aiming to raise the students' English proficiency. The educational curriculum focused especially on the students' communicative skills and their ability to broaden their knowledge written in English has been implemented. Moreover, rules and regulations concerning students' English proficiency has been imposed. According to the university's

policy, all of the graduates need to have the official result of University of Phayao Standard Test of English Proficiency TOEFL IBT TOFEL ITP or IELTS marking in B1 level; except the English major students who need to achieve standard marking C1 or above.

As the students' English proficiency is like the indicator for their graduation and job opportunities, University of Phayao Standard Test of English Proficiency is needed to be researched and standardized. In 2019, the researchers conducted the research proven the standard of University of Phayao Standard Test of English Proficiency in level of B1. The test was implemented to test students' English proficiency effectively according to the result of the research. Following that successful development and evaluation of the B 1 level test, the researchers believe that it is also very necessary to carry out research to develop the test and ensure the quality enhancement of University of Phayao Standard Test of English Proficiency in C1 level. This English Proficiency test is aimed to use for the fourth year English students at University of Phayao.

## **Objectives of Research**

1. To develop University of Phayao Standard Test of English Proficiency in level of C1
2. To evaluate the quality of University of Phayao standard test of English proficiency

## **Research Methodology**

**Research Design:** Quantitative Research

**Sample group:** One hundred and seventy-four out of two hundred fourth-year English major students of the School of Liberal Arts, the University of Phayao in academic year 2021.

## Data Collection Instruments

**Instrument:** Two versions of Standard English Proficiency Test C 1 level developed based on CEFR framework, University of Phayao

### Instruments Design

1. Outline the content and form of University of Phayao Standard Test of English Proficiency by using Common European Framework of Reference level C1 as a reference
2. Develop the test according to the outline
3. Validate the content validity of the test by three experts in English and Measurement and Evaluation
4. Revise the test according to the experts' suggestions
5. Conduct the first pilot trail
6. Analyze the pilot data in terms of difficulty, discrimination and reliability of the test

### Validity and Reliability of Data Collection Instruments

**1. Content validity:** The test is ensured the content validity by three experts in the field of evaluation & assessment and English language by using the index of item-objective congruence (IOC). The items gaining IOC value lower than 0.50 are **revised** according to the experts' suggestions, and the items with the value 0.50 or higher are reserved.

**2. Difficulty index:** Saiyot and Saiyot (2001)'s the index ( $p$ ) of difficulty was used to evaluated the difficulty level ( $p$ ) of the test. The criteria were as follows:

**Table 1** Index of difficulty ( $p$ )

Value	Meaning	Recommendation
< 0.19	Very difficult	Remove or Revise
0.20 - 0.39	Somewhat difficult	Retain
0.40 – 0.60	Moderately difficult	Retain
0.61 – 0.80	Somewhat easy	Retain
0.81 – 1.00	Very Easy	Remove or Revise

**3. Discrimination power (r):** The discrimination index of Saiyot and Saiyot (2001) was used as the criteria to determine discrimination power (r). The criteria were as follows:

**Table 2** Index of discrimination (r)

Value	Meaning	Recommendation
-1.00 – 0.09	Very low discrimination ability	Remove or Revise
0.10 – 0.19	Low discrimination ability	Remove or Revise
0.20 – 0.39	Moderate discrimination ability	Retain
0.40 – 0.59	High discrimination ability	Retain
≥ 0.60	Very high discrimination ability	Retain

The qualified test items which passed both criteria of difficult index and discrimination index were retained, and unqualified items which did not meet both criteria or one of them were removed or should be revised.

**4. Reliability index:** The Kuder-Richardson procedure (KR-20) is utilized to calculate the reliability index in the test because it is the objective test with multiple choices. Each item of the test score as 1 for the correct item and 0 points for the incorrect one. According to Burn, Grove and Sutherland (2017, p. 583), the reliability value of the tester-made test should be 0.70 and over.

**Data collection Methods:** The rules and regulations about English major students' English Proficiency complying with Common European Framework of Reference level C1 and rules and regulations about Thailand Qualification Framework for English major students' English Proficiency are studied as the theoretical concept of the test development. After the test is developed, it is evaluated the content validity by three experts in the fields of English teaching and education evaluation and assessment. Then the test is tried out with the sample. Next the test's scores are estimated difficulty index, discrimination power and reliability value.

### Data Analysis and Interpretation

1. Analyze content validity index by using index of item objective Congruence or IOC
2. Analyze each item of the test by using difficulty index
3. Analyze each item of the test by using discrimination index
4. Analyze reliability index of the test by using Kuder-Richardson 20 (KR-20)

### Results of Research

#### 1. The results of developing University of Phayao Standard Test of English Proficiency in level of C1

University of Phayao Standard Test of English Proficiency in level of C1 was conducted by studying rules and regulations about Thailand Qualification Framework for English major students' English Proficiency and regulations about English major students' English Proficiency complying with Common European Framework of Reference level C1. Then researchers outline the contents and forms of University of Phayao Standard Test of English Proficiency based on Common European Framework of Reference level C1. From the framework of development of University of Phayao Standard Test of English Proficiency in level of C1. The researchers had achieved two versions of the C1 Standard Test of English Proficiency for the fourth years students in University of Phayao. Each version of the C1 Standard Test of English Proficiency consists of the contents as follows:

Part 1: Vocabulary Test	15 items No. 1-15
Part 2: Reading Test	15 items No. 16-30
Part 3: Grammar & Writing Test	15 items No. 31-45
Part 4: Conversation Test	15 items No. 46- 60
Total 60 items in each version	

In sum, the two versions of C1 Standard Test of English Proficiency which consist of 60 items in four parts; vocabulary, reading, grammar and writing and conversation respectively for the fourth years students in University of Phayao have been complete conducted and ready for test taking.

## 2. The results of evaluating the quality of University of Phayao standard test of English proficiency C 1 level

The data of University of Phayao Standard Test of English Proficiency in level of C1 were evaluated under the three analyses. They are; Content Validity Analysis by using index of item objective congruence or IOC, Difficulty Level and Discrimination Analysis and Reliability Analysis by Kuder-Richardson 20 (KR-20). The results of the study were shown below.

### 2.1 The content validity

The content validity of the English proficiency tests was analyzed by three experts in the fields of English language and evaluation and assessment in education by using the index of item objective congruence (IOC). The content validity of the tests was shown in Table 3

**Table 3** The results of content validity analysis

Test Version	The IOC index mean of the expert scores of each item	Valid items ( $\geq .50$ )	Invalid items ( $< .50$ )	The overall IOC index mean of the expert scores
Version 1	0.33 – 1.00	57 items	3 items No.23,32,37	<b>0.94</b>
Version 2	0.67 – 1.00	60 items	<b>none</b>	<b>0.99</b>

From Table 3, it showed that the IOC index means of the test items of version one ranged from 0.33 to 1.00 indicated that there were 57 items passed the criteria, and there were three items gained IOC mean lower than .50. They were items No. 23, 32, and 37 respectively. That meant these items needed to be revised. However, the overall IOC index mean of the English proficiency test version one was at 0.94 indicating that the test was valid.



For the test items of version two, the IOC index means were from 0.67 to 1.00. The results showed that all items of this version passed the criteria. The overall IOC index mean was 0.99 which reflected the high content validity.

However, the experts suggested that some words were used repeatedly in the choices of some items and some vocabularies needed to be revised for the well appropriate to CEFR C1 level test. Also, the contexts of some test questions were too simple for C1 levels, so they should be revised.

## 2.2 The difficulty level and discrimination power

The analysis of difficulty level and discrimination power was done by using the criteria recommended by Saiyot and Saiyot (2001). The acceptable scales of the difficulty index (p) ranging from 0.20 – 0.80 and discrimination index (r) ranging from 0.20- >0.60. Those items that were not matched in the expected range were removed. The results of the difficulty level and discrimination power analysis were presented in Table 4

**Table 4** Difficulty Level and Discrimination Power

Test Version	Difficulty Level (P)	Discrimination Power (r)	Results	
			Retained Items	Removed Items
1	0.16 - 0.86	-0.04 - 0.52	42 items	18 items (No. 9, 12, 16-19, 21, 24-25, 27, 29-30, 34-35, 37, 39, 41, and 42)
2	0.17 - 0.92	-0.05 - 0.55	45 items	15 items (No. 3, 5-6, 13-15, 17, 19, 24, 28-29, 35, 46, and 60)

From Table 4, version one of the English proficiency tests showed that the difficulty index (p) ranged from 0.16 to 0.86, and the discrimination power index (r) were from -0.04 to 0.52. According to the evaluating range of the test, the qualified items that gained the difficulty level from 0.20 to 0.80 and the discrimination power from 0.20 were retained. From the tryout of

the test with the 87 fourth year's students, there were 42 qualified test items, and 18 items did not meet the criteria of difficulty level or discrimination power index. Those items were No. 9, 12, 16-19, 21, 24-25, 27, 29-30, 34-35, 37, 39, 41, and 42. They were removed or revised.

The result of evaluating version two of the English proficiency test revealed that the difficulty index ( $p$ ) was from 0.17 to 0.92, and the discrimination power index ( $r$ ) varied from -0.05 to 0.55. However, the qualified items should obtain the difficulty index from 0.20 to 0.80 and the discrimination level from 0.20. Based on the tryout with 87 fourth year students, 45 test items were qualified, but there were 15 items should be removed or revised. The unqualified items were No. 3, 5-6, 13-15, 17, 19, 24, 28-29, 35, 46, and 60.

### 2.3 The analysis of reliability

The Kuder-Richardson procedure (KR-20) is utilized to calculate the reliability index in the test because it is an objective test with multiple choices. Each item of the test score is 1 for the correct item and 0 points for the incorrect one. According to Burn, Grove, and Sutherland (2017, p. 583), the reliability value of the tester-made test should be 0.70 and over. The reliability of the tests was demonstrated in Table 5

**Table 5** Reliability of the tests

English Proficiency Test	Reliability	Meaning
Version 1	0.82	Acceptable
Version 2	0.83	Acceptable

Table 5 showed that the reliability index of version one of the English proficiency tests was 0.82 and the value of version two was 0.83. This indicated that the two versions of English proficiency tests had reached an acceptable reliability value.

## Conclusion and Discussion

In this present study, two versions of English proficiency tests for undergraduate English major students of the School of Liberal Arts, University of Phayao were developed aligned with the CEFR C1 level. Each version of the test consisted of 60 items. The tests were divided into 4 parts including vocabulary, reading, grammar and writing, and conversation. There were 15 items in each part.

To verify the content validity of the University of Phayao standard test of English Proficiency, the Index of Item-Objective Congruence (IOC) was employed by three experts to ensure that the test items serve their purposes as standardized tests. The IOC index means of the test items of version one ranged from 0.33 to 1.00. Out of 60 items, there were 57 items passed the criteria, and there were three items gained IOC means lower than .50. Hence, these items needed to be revised. The overall IOC index of the English proficiency test version one was 0.94. For the test items of version two, the IOC index means were from 0.67 to 1.00. The results showed that all items of this version passed the criteria. The overall IOC index mean was 0.99 which reflected the high content validity. This indicated that two versions of the tests were valid.

The use of IOC for content validity verification is also found in the development of SWU-SET as a CEFR standardized English test carried out by Athiworakun, Vathanalaoha, Thongprayoon, Rajprasit, and Yaemtui (2018). In this previous study, IOC was utilized to confirm that the test items can assess the content that they intend to assess. According to Turner and Carlson (2003), it is important for test development to affirm that the developed test items are measuring the content they are intended to measure. The evidence of content validity can be provided by the evaluation administered by an independent expert panel. IOC has been helpful for test designers in providing an independent evaluation of the validity of test items before carrying out a pilot test.

Two versions of the University of Phayao Standard Test of English Proficiency were developed in this study are multiple choice questions. Danuwijaya (2018) mentions that the quality of the items of multiple-choice questions should be assessed. To create a good item bank, it is useful to identify the difficulty and easiness of the items and to check how they can discriminate between low and high-proficiency learners. Therefore, the difficulty level and discrimination power were calculated. The difficulty level was used to prove that the test items were not too easy or too difficult. Discrimination power was also used to evidence that tests can distinguish the test takers with high proficiency and low proficiency. Saiyot and Saiyot (2001) suggest that the qualified test items should meet the criteria of both difficulty and discrimination indexes. The acceptable value of difficulty level is from 0.20 - 0.80, and recommended value of discrimination power ranges from 0.20 – 1.00. Based on these criteria, there were 42 items of version 1 of the test were qualified, and there were 45 items of version 2 that met the criteria and were retained.

In the aspect of the verification of the reliability, two sets of the University of Phayao standard test of English Proficiency for English major students were analyzed by using the Kuder-Richardson procedure (KR-20) since they are the objectives test with multiple choices. The reliability index of version one of the English proficiency tests was 0.82 and the value of version two was 0.83. The results pointed out that the tests had high reliability. According to Burn, Grove, and Sutherland (2017), an in-house developed test should gain a reliability index from 0.70 and over. The results of the reliability analysis were in line with the study of Athiworakun et. al. (2018). It was reported that based on KR-20 analysis, the SWU-SET as a CEFR standardized English test gained high reliability at 0.93 and 0.94.

In conclusion, the objectives of the research were completely accomplished because the researchers could develop the tests which

reached the high content validity, met the criteria of difficulty and discrimination indexes, and gained acceptable reliability value. However, there were some test items that should be revised according to the difficulty level and discrimination power.

### **New Body of Knowledge**

To achieve the standard in-house proficiency test, an effective plan must be set based on the given criteria of a particular test strictly. From the study, there were 4 processes that were undergone throughout this.

First, the process of studying the concept of CEFR C1 level in depth and in every aspect of this particular test and study more on the sample of the C1 level test from other institutes from the European Universities. Studying several C1 level tests makes us learning and experiencing. Learning never ends in one life.

Second, the process of co-operation among our colleague in the school. Co-operation started when the tests had been already completed. Many colleagues and experts from English department spent their time to take the C1 test and gave some comments. This supported the perfection of the test. The researchers had learned much from them.

Third, the process of revising was done after the comments had been given as soon as possible. Procrastination was not with us during conducting the two versions of the test.

Finally, the revised versions were tried out again by the researchers, then the two versions of the test were tested by 150 fourth year students.

In sum, conducting in-house proficiency test by following these steps would bring success on this challenge.

## Suggestions for Research

### Suggestion on Policy

1. To develop English proficiency test takes much time and expenditure. It would be more beneficial if there is a good co-operation between or among Thai universities. This would also make the test more validity in all aspects

2. The attitudes and opinions of the examination candidates should be observed and used for revising the tests.

### Suggestion for further study

The English proficiency test on listening and speaking tests should be developed and evaluate as a main part of C1 level so as that the candidates' English skills can be completely validated.

## References

- Athiworakul C. et. al. (2018). SWU-SET as a CEFR standardized English test. *Journal of Language Teaching and Research*, 9(2), 261-267.
- Athiworakul, C., & Wudthayagorn, J. (2018). Mapping Srinakharinwirot University Standardized English Test (SWU-SET) onto the Common European Framework of Reference (CEFR). *Suranaree Journal of Social Science*, 12(2), 69-84.
- Bachman, L. F. (1990). *Fundamental considerations in Language Testing*. Oxford: Oxford University Press.
- Brown, H.D. (2004) *Language Assessment: Principles and Classroom Practices*. White Plains, NY: Pearson.
- Burns, N., Grove, S.K. & Sutherland, S. (2017). *Burn and Grove's the Practice of Nursing Research: Appraisal, synthesis, and generation of evidence*. 8th Edition. Missouri: Elsevier.

- Carr, N.T. (2011). *Designing and Analyzing Language Tests*. Oxford: Oxford University Press.
- Danuwijaya, A. A. (2018). Item analysis of reading comprehension test for post-graduate students. *English Review: Journal of English Education*, 7(1), 29-40.
- Heaton, J.B. (1988). *Writing English Language Test*. (new ed.). Essex: Longman.
- Hughes, A. (2003) *Testing for Language Teachers*. (2nd ed.). Cambridge: Cambridge University Press.
- Jin, Y. (2011). "Fundamental Concerns in High-Stakes Language Testing: The Case of the College English Test". in *Pan-Pacific Association of Applied Linguistics.*, 15(2), 71- 83.
- Lord, F.M. (1952). The Relationship of the Reliability of Multiple-Choice Test to the Distribution of Item Difficulties. *Psychometrika*, 18, 181-194.
- Mcnamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.
- Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Ngadkatok, S. et al. (2018). Quality of Ordinary National Educational Test and Guidelines for Enhancing the Test Quality. *Sukhothai Thammathirat Open University Journal*, 31(2), 110-123.
- Salkind, N.J. (2013) *Test & Measurement for People who (think they) Hate Tests & Measurement* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage Publications.
- Sanglertuthai, C. (2015). Research Instruments. *Sakon Nakhon Graduate Studies Journal*, 12(58), 13-24.
- Saiyot, L., & Saiyot, A. (2001). *Assessment Techniques*. Bangkok: Suweeriyasarn.
- Sinjindawong, S. (2018). Items Analysis Methods. *Sripatum Review of Science and Technology*, 4(1), 21-32.
- Tongnak, S., Sirichai K., & Kerdsuwan, S. (2015). Development of a Competency Test for Student Teachers Based on Teaching Professional Standards Using Polytomously Scored Item. *Journal of Multidisciplinary in Social Science*, 9(1), 169-187.

- Turner, R.C., & Carlson, L. (2003). Indexes of Item-Objective Congruence for Multidimensional Items. *International Journal of Testing*, 3(2), 163-171.
- Wanichdee, A., Sripunworasakul. S., & Thongsorn. N. (2008). *Development of English Test via Electronic Media in a Distance Learning System*. Bangkok: Sukhothai Thammathirat Open University.
- Wudthayagorn, J. (2018). Mapping the CU-TEP to the Common European Framework of Reference (CEFR). *LEARN Journal: Language Education and Acquisition Research Network Journal*, 11(2), 163-180.