

# การพัฒนาและการวิเคราะห์คุณภาพ แบบทดสอบความสามารถภาษา อังกฤษตามกรอบมาตรฐาน CEFR สำหรับนักศึกษาหลักสูตรภาษาอังกฤษ มหาวิทยาลัยราชภัฏเชียงใหม่

รชฎ นุเสน

คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏเชียงใหม่

กนิษฐ์ ดีहनอ

คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏเชียงใหม่

พะเยา นุเสน

คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏเชียงใหม่

สลิลา ศรีรัตนบัลล์

คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏเชียงใหม่

Received : August 3, 2024

Revised : December 18, 2025

Accepted: January 8, 2026

## บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบทดสอบความสามารถทางภาษาอังกฤษของนักศึกษาหลักสูตรภาษาอังกฤษ มหาวิทยาลัยราชภัฏเชียงใหม่ ที่มีคุณภาพ เครื่องมือวิจัยที่ใช้ได้แก่ 1) แบบทดสอบภาษาอังกฤษของหลักสูตร ซึ่งเป็นข้อสอบปรนัยแบบสี่ตัวเลือก จำนวน 100 ข้อ วัดความสามารถทางภาษาในสี่ด้าน คือ ทักษะการฟัง ทักษะการอ่าน ความรู้เกี่ยวกับภาษาที่ใช้ในการสนทนา และความรู้เกี่ยวกับภาษา คำศัพท์ และไวยากรณ์ที่ใช้ในการเขียน ข้อสอบได้รับการออกแบบมาให้สอดคล้องกับคำอธิบายความสามารถทางภาษา

ผู้รับผิดชอบบทความ : รชฎ นุเสน คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏเชียงใหม่ เลขที่ 202  
ตำบลช้างเผือก อำเภอเมือง เชียงใหม่ 50300 อีเมล rachod.nus@cmru.ac.th

ตามกรอบมาตรฐาน CEFR ในระดับ A2, B1, และ B2 2) แบบประเมินความสอดคล้องระหว่างข้อสอบกับ วัตถุประสงค์หลายข้อ 3) แบบประเมินคุณภาพด้านคุณประโยชน์ของแบบทดสอบ และ 4) แบบบันทึก การสัมภาษณ์ผู้ทดลองสอบแบบกึ่งมีโครงสร้าง ผลการวิจัยพบว่า ข้อสอบทุกข้อมีค่าความเที่ยงตรงทางเนื้อหา มากกว่า .75 จากการประเมินของผู้เชี่ยวชาญ และผู้เชี่ยวชาญ “เห็นด้วยอย่างยิ่ง” ว่าแบบทดสอบมีคุณภาพ ด้านคุณประโยชน์ โดยด้านที่ได้คะแนนประเมินสูงที่สุดมีด้วยกันสองด้าน คือ ด้านความเชื่อมั่นและด้าน ผลกระทบที่เกิดขึ้น การสัมภาษณ์ผู้ทดลองสอบจำนวน 11 คนแสดงให้เห็นได้ว่าการจัดสอบโดยภาพรวม ดำเนินไปอย่างมีประสิทธิภาพ ไม่มีปัจจัยใดที่จะส่งผลสำคัญต่อความเชื่อมั่นและความถูกต้องของคะแนน สอบ จากนั้นคณะผู้วิจัยใช้วิธีการคัดเลือกแบบเจาะจง เลือกนักศึกษามาสอบแบบทดสอบจำนวน 100 คน เพื่อเก็บข้อมูลคุณภาพของแบบทดสอบ ผลการวิเคราะห์พบว่า แบบทดสอบมีค่าความเชื่อมั่นตามสูตร KR-20 อยู่ที่ .93 มีค่าความยากง่ายอยู่ที่ .57 มีค่าอำนาจจำแนกอยู่ที่ .46 และมีค่าเบี่ยงเบนมาตรฐานอยู่ที่ 16.64

**คำสำคัญ :** แบบทดสอบความสามารถทางภาษาอังกฤษ / การพัฒนาแบบทดสอบ / การวิเคราะห์ คุณภาพแบบทดสอบ

# Developing and Analyzing a CEFR-Based English Proficiency Test for Students in English Programs at Chiang Mai Rajabhat University

---

Rachod Nusen

Faculty of Humanities and Social Sciences, Chiang Mai Rajabhat University

Kanit Deenor

Faculty of Humanities and Social Sciences, Chiang Mai Rajabhat University

Phayao Nusen

Faculty of Humanities and Social Sciences, Chiang Mai Rajabhat University

Salila Sriratanaban

Faculty of Humanities and Social Sciences, Chiang Mai Rajabhat University

## Abstract

This research aims to develop a quality English proficiency test for students in English programs at Chiang Mai Rajabhat University. The research tools used include: 1) an English proficiency test, comprising 100 four multi-choices items, assessing four areas of language proficiency: listening, reading, conversational language knowledge, and knowledge of language, vocabulary, and grammar used in writing. The test was designed to align with the standard framework of CEFR at levels A2, B1, and B2; 2) an assessment form for indexes of item-objective congruence for multidimensional items; 3) an assessment form for evaluating the test's usefulness; and 4) a semi-structured interview form. The results showed that all test items had a content validity index greater than .75 according to experts' evaluations and the experts "strongly agreed" that the test was of high quality in terms of its usefulness, with the highest scores in reliability and impact.

**Corresponding Author** : Rachod Nusen, Faculty of Humanities and Social Sciences, Chiang Mai Rajabhat University, 202 T.Changpuek, A.Muang, Chiang Mai, 50300, Email : rachod.nus@cmru.ac.th

Interviews with 11 test takers indicated that the overall test administration was effective, with no significant factors affecting the reliability and accuracy of the test scores. The researchers then used purposive sampling to select 100 students to take the test to collect data on the test's quality. The analysis revealed that the test had a reliability coefficient (KR-20) of .93, a difficulty index of .57, a discrimination index of .46, and a standard deviation of 16.64.

**Keywords :** English proficiency test / test development / test analysis

## 1. บทนำ

คณะกรรมการมาตรฐานการอุดมศึกษา (2567) กระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัย และนวัตกรรม ได้ออกประกาศ เรื่อง นโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา เพื่อพัฒนาความสามารถทางภาษาอังกฤษของคนในประเทศ ดังข้อความตอนหนึ่งว่า ให้สถาบันอุดมศึกษาพิจารณาจัดให้นิสิตนักศึกษาทุกคน สอบวัดความรู้และทักษะภาษาอังกฤษก่อนสำเร็จการศึกษา ตามแบบทดสอบที่สถาบันอุดมศึกษาพัฒนาขึ้นหรือแบบทดสอบตามมาตรฐานสากลอื่น ๆ โดยเทียบเคียงผลกับ *Common European Framework of Reference for Languages (CEFR)* เพื่อให้ทราบถึงผลสัมฤทธิ์ด้านทักษะภาษาอังกฤษของนิสิตนักศึกษาแต่ละคน และสถาบันอุดมศึกษาอาจพิจารณานำผลการทดสอบความรู้และทักษะภาษาอังกฤษบันทึกในใบรับรองผลการศึกษา หรือจัดทำเป็นประกาศนียบัตร ทั้งนี้ เป้าหมายผลสัมฤทธิ์ด้านทักษะภาษาอังกฤษ [...] ระดับปริญญาตรี ควรกำหนดเป้าหมายผลสัมฤทธิ์ด้านทักษะภาษาอังกฤษก่อนสำเร็จการศึกษา เทียบเคียงผลกับ CEFR ในระดับเกณฑ์ตั้งแต่ B2 ขึ้นไป

เพื่อให้การดำเนินงานเป็นไปอย่างสอดคล้องกับนโยบายการยกระดับมาตรฐานภาษาอังกฤษ สถาบันอุดมศึกษาในประเทศไทยหลายแห่งจึงพัฒนาแบบทดสอบความสามารถทางภาษาอังกฤษตามกรอบมาตรฐาน CEFR ขึ้นมา เพื่อวัดความสามารถของนักศึกษา อย่างไรก็ตาม ปรากฏว่าแบบทดสอบที่มหาวิทยาลัยต่าง ๆ พัฒนาขึ้นมักมีเป้าหมายเพื่อวัดความสามารถทางภาษาอังกฤษของนักศึกษาที่กำลังจะสำเร็จการศึกษาตามประกาศข้างต้น แบบทดสอบลักษณะนี้จัดเป็นแบบทดสอบรวบยอด (summative test) ซึ่งจัดขึ้นหลังกระบวนการ

การเรียนการสอนเสร็จสิ้นแล้ว ดังนั้น การที่มหาวิทยาลัยต่าง ๆ จะนำผลการสอบไปดำเนินการอย่างใดอย่างหนึ่งเพื่อพัฒนาผู้เข้าสอบแต่ละคนต่อไป จึงเป็นเรื่องยาก และหากจัดให้มีการสอบแบบทดสอบของมหาวิทยาลัยล่วงหน้าหลายครั้ง ก็อาจจะทำให้เกิดการร่วไหลของข้อสอบ

หลักสูตรครุศาสตรบัณฑิต สาขาวิชาภาษาอังกฤษ และหลักสูตรศิลปศาสตรบัณฑิต สาขาวิชาภาษาอังกฤษ และสาขาวิชาภาษาอังกฤษธุรกิจ มหาวิทยาลัยราชภัฏเชียงใหม่ ได้ร่วมกันสร้างแบบทดสอบความสามารถทางภาษาอังกฤษตามกรอบมาตรฐาน CEFR ขึ้น เพื่อใช้วัดความสามารถทางภาษาของนักศึกษาในหลักสูตร (ต่อไปนี้จะเรียกแบบทดสอบที่พัฒนาขึ้นว่า แบบทดสอบภาษาอังกฤษของหลักสูตร) การมีแบบทดสอบเป็นของตนเองจะทำให้แต่ละหลักสูตรสามารถวัดความสามารถของนักศึกษาเมื่อใดก็ได้ตามที่เห็นเหมาะสม สาเหตุอีกประการหนึ่งที่คณะผู้วิจัยเห็นว่าควรมีการสร้างแบบทดสอบขึ้นมาเอง โดยไม่ใช้ข้อสอบมาตรฐานที่สามารถเทียบค่าการกำหนดเกณฑ์มาตรฐาน CEFR เช่น ข้อสอบ IELTS และ TOEFL เนื่องจากการสอบข้อสอบมาตรฐานเหล่านี้มีค่าใช้จ่ายที่สูงมากเมื่อเทียบกับค่าธรรมเนียมการศึกษาของมหาวิทยาลัยราชภัฏเชียงใหม่ การกำหนดให้สอบจะสร้างภาระทางการเงินให้กับนักศึกษา ค่าใช้จ่ายในการสอบข้อสอบมาตรฐานบางสำนักสูงกว่าค่าธรรมเนียมการศึกษาที่ต้องชำระในหนึ่งภาคการศึกษา

## 2. วัตถุประสงค์

เพื่อพัฒนาแบบทดสอบความสามารถทางภาษาอังกฤษของนักศึกษาหลักสูตรภาษาอังกฤษ มหาวิทยาลัยราชภัฏเชียงใหม่ ที่มีคุณภาพ

### 3. วิธีดำเนินการวิจัย

ขั้นตอนการวิจัยเพื่อพัฒนาแบบทดสอบความสามารถภาษาอังกฤษของหลักสูตรปรับเปลี่ยนมาจากข้อเสนอของสหภาพยุโรป (Council of Europe, 2011) ว่าด้วยขั้นตอนการพัฒนาแบบทดสอบความสามารถทางภาษาตามกรอบมาตรฐาน CEFR การวิจัยครั้งนี้เริ่มต้นด้วยการกำหนดคุณลักษณะเฉพาะของแบบทดสอบการแต่งข้อสอบ การประชุมแก้ไขข้อสอบ การตรวจสอบความถูกต้องของภาษา การประเมินค่าความเที่ยงตรงทางเนื้อหาและการประเมินค่าคุณภาพด้านคุณประโยชน์ของแบบทดสอบโดยผู้เชี่ยวชาญ การทดลองสอบการสัมภาษณ์ผู้ทดลองสอบ การปรับวิธีการจัดสอบตามข้อมูลจากการสัมภาษณ์ การสอบ และการวิเคราะห์ข้อมูลเพื่อหาคุณภาพข้อสอบ

#### 3.1 กลุ่มเป้าหมาย

กลุ่มเป้าหมายซึ่งเข้าร่วมการวิจัยแบ่งออกได้เป็นสองกลุ่ม กลุ่มที่หนึ่งเป็นกลุ่มเป้าหมายในช่วงการพัฒนาข้อสอบ ได้แก่ นักศึกษาวิชาเอกภาษาอังกฤษจำนวน 11 คน ซึ่งอาสาเข้าร่วมทดลองทำแบบทดสอบ เพื่อให้ข้อมูลเกี่ยวกับความเหมาะสมของแบบทดสอบและการจัดสอบในภาพรวม

กลุ่มที่สองเป็นกลุ่มเป้าหมายในการตรวจสอบคุณภาพข้อสอบ ได้แก่ นักศึกษาของหลักสูตรครุศาสตรบัณฑิต สาขาวิชาภาษาอังกฤษและหลักสูตรศิลปศาสตรบัณฑิต สาขาวิชาภาษาอังกฤษ และสาขาวิชาภาษาอังกฤษธุรกิจ จำนวน 100 คน

#### 3.2 เครื่องมือวิจัย

3.2.1 แบบทดสอบภาษาอังกฤษของหลักสูตร ซึ่งเป็นข้อสอบปรนัยแบบสี่ตัวเลือก จำนวน 100 ข้อ มีวัตถุประสงค์เพื่อวัดความสามารถทาง

ภาษาในสี่ด้าน คือ ทักษะการฟัง ทักษะการอ่าน ความรู้เกี่ยวกับภาษาที่ใช้ในการสนทนา และความรู้เกี่ยวกับภาษา คำศัพท์ และไวยากรณ์ที่ใช้ในการเขียน ข้อสอบได้รับการออกแบบมาให้สอดคล้องกับคำอธิบายความสามารถทางภาษาตามกรอบมาตรฐาน CEFR ในระดับ A2 B1 และ B2 ผู้พัฒนาแบบทดสอบกำหนดให้แบบทดสอบวัดความสามารถในสามระดับนี้ เพราะตามประสบการณ์ของคณะผู้วิจัยแล้ว ความสามารถในสามระดับนี้ครอบคลุมความสามารถของนักศึกษาหลักสูตรภาษาอังกฤษมหาวิทยาลัยราชภัฏเชียงใหม่

3.2.2 แบบประเมินความสอดคล้องระหว่างข้อสอบกับวัตถุประสงค์หลายข้อ (Index of Item-Objective Congruence for Multidimensional Items) ซึ่งกำหนดให้ผู้เชี่ยวชาญ 5 คน ประเมินความสอดคล้องของข้อสอบจำนวน 100 ข้อ กับวัตถุประสงค์สองประการ คือ ทักษะหรือความรู้ที่ข้อสอบวัด และระดับ CEFR ที่ข้อสอบวัด ผู้เชี่ยวชาญทุกคนเป็นผู้ที่มีประสบการณ์ในการสอนภาษาอังกฤษมาไม่น้อยกว่า 15 ปี และมีประสบการณ์ในการพัฒนาแบบทดสอบความสามารถทางภาษาอังกฤษตามกรอบมาตรฐาน CEFR จากนั้นคณะผู้วิจัยหาค่า IOC ของข้อสอบตามสูตรที่เทิร์นเนอร์และคาร์ลสัน (Turner & Carlson, 2003) เสนอ ดังนี้

$$I'_{ik} = \frac{(N)\mu_k - (N-p)\mu_i}{2N-p}$$

$I'_{ik}$  แทน ดัชนีความสอดคล้องระหว่างข้อสอบ (i) กับวัตถุประสงค์ทั้งหมด (k)  $N$  แทน จำนวน วัตถุประสงค์  $P$  แทน จำนวนวัตถุประสงค์ที่ ข้อสอบข้อนั้นต้องการวัด  $\mu_k$  แทนคะแนนเฉลี่ย ของผู้เชี่ยวชาญในวัตถุประสงค์ที่ต้องการวัด และ  $\mu_i$  แทน คะแนนเฉลี่ยของผู้เชี่ยวชาญใน วัตถุประสงค์ที่ไม่ต้องการวัด

3.2.3 แบบบันทึกการสัมภาษณ์ผู้ทดลอง สอบแบบกึ่งมีโครงสร้าง ซึ่งประกอบด้วยคำถาม 6 ข้อ การสัมภาษณ์นี้มีวัตถุประสงค์เพื่อให้ผู้ทดลอง สอบสะท้อนความเห็นของตนต่อการจัดสอบหกด้าน ได้แก่ ความเหมาะสมของเวลาที่ใช้ในการจัดสอบ คุณภาพของระบบเสียงสำหรับการทดสอบการฟัง คุณภาพของรูปภาพ ความเหมาะสมของขนาด ตัวอักษร คุณภาพการถ่ายสำเนาแบบทดสอบ และ ความเหมาะสมของห้องสอบ แบบบันทึกนี้ผ่านการ ตรวจสอบความเหมาะสมในภาพรวมโดย ผู้เชี่ยวชาญจำนวน 3 คน แล้วก่อน การนำใช้สัมภาษณ์

3.2.4 แบบประเมินคุณภาพด้าน คุณประโยชน์ของแบบทดสอบ (test usefulness evaluation form) แบบประเมินนี้สร้างขึ้นเพื่อ วิเคราะห์คุณภาพของแบบทดสอบตามที่ แบชแมน และพาล์มเมอร์ (Bachman & Palmer, 1996) ได้เสนอว่า แบบทดสอบที่มีคุณภาพประกอบด้วย คุณสมบัติหกประการ คือ ความเชื่อมั่น (reliability) ความเที่ยงตรงทางโครงสร้าง (construct validity) ความเหมือนจริง (authenticity) การปฏิสัมพันธ์ (interactiveness) ผลกระทบที่เกิดขึ้น (impact) และการนำไปใช้ได้จริง (practicability) แบบ ประเมินนี้ผ่านการตรวจสอบความเหมาะสมในภาพ รวมโดยผู้เชี่ยวชาญจำนวน 3 คน แล้วก่อนการนำใช้ เก็บข้อมูล

## 4. ผลการวิจัย

### 4.1. การกำหนดคุณลักษณะเฉพาะของ แบบทดสอบ

คณะผู้พัฒนาแบบทดสอบประชุมร่วมกัน เพื่อกำหนดวัตถุประสงค์ของแบบทดสอบและวิธีการ ทดสอบที่ประชุมระบุวัตถุประสงค์ของแบบทดสอบ ไว้สองวัตถุประสงค์ คือ หนึ่ง เพื่อวัดความสามารถ ทางภาษาอังกฤษของนักศึกษาในหลักสูตร และ สอง เพื่อให้ข้อมูลสะท้อนกลับจากผลการสอบกับ ผู้เกี่ยวข้อง นอกจากนี้ คณะผู้พัฒนาแบบทดสอบ ยังเห็นควรให้พัฒนาแบบทดสอบในลักษณะของ ข้อสอบปรนัยแบบสี่ตัวเลือก จำนวน 100 ข้อ เพื่อวัด ความสามารถทางภาษาตามกรอบมาตรฐาน CEFR ในระดับ A2 B1 และ B2 ทั้งนี้เพื่อให้แบบทดสอบ ภาษาอังกฤษของหลักสูตรมีความสอดคล้องกับ คุณสมบัติเฉพาะของแบบทดสอบภาษาอังกฤษ CMRU-TEP ของมหาวิทยาลัยราชภัฏเชียงใหม่

ในการประชุมครั้งต่อมา คณะผู้พัฒนาแบบ ทดสอบได้ร่วมกันกำหนดรายละเอียดของการ กำหนดคุณลักษณะเฉพาะของแบบทดสอบโดยใช้ แบบฟอร์มเพื่อการบรรยายลักษณะแบบทดสอบของ สหภาพยุโรป (Council of Europe, 2009) และ ฮิวส์ และฮิวส์ (Hughes & Hughes, 2020) เป็นแนวทาง รายละเอียดที่ร่วมกันกำหนดมีหลายเรื่อง เช่น จำนวนของบทอ่านในแบบทดสอบ ความยาวของ บทอ่าน ความยากง่ายของบทอ่าน ประเภทของ บทอ่าน และเนื้อเรื่องของบทอ่าน เป็นต้น

นอกจากนั้น เพื่อให้ข้อสอบมีความเที่ยงตรง (validity) และสามารถเปรียบเทียบกับกรอบ มาตรฐาน CEFR ได้ คณะผู้พัฒนาจึงร่วมกันกำหนด โครงสร้างแบบทดสอบ และตกลงให้ออกแบบ ข้อสอบแต่ละข้อให้วัดทักษะตามคำอธิบายเกณฑ์ มาตรฐาน CEFR ในเอกสาร *Common European*

*Framework of Reference for Languages: Learning, Teaching, Assessment* ของสหภาพยุโรป (Council of Europe, 2020) และคำอธิบายในเอกสาร *EAQUALS Core Inventory for General English* ของ นอร์ธ, โอเตกา และชีแฮน (North, Ortega, & Sheehan, 2011) ซึ่งได้รับการสนับสนุนจากบริติช เคานซิล (British Council)

#### 4.2 การวิเคราะห์และแก้ไขข้อสอบ

คณะผู้พัฒนาแบบทดสอบประชุมร่วมกัน เพื่อพิจารณาความเที่ยงตรงของข้อสอบกับคำอธิบายตามกรอบมาตรฐาน CEFR (ตารางที่ 1-4)

ตารางที่ 1 การวิเคราะห์ข้อสอบการฟังตามคำอธิบายกรอบมาตรฐาน CEFR (Council of Europe, 2020; North, Ortega, & Sheehan, 2011)

ลำดับข้อ	ทักษะที่ใช้	หัวข้อ	ระดับ CEFR
9	Describing habits and routines	Leisure activities	A2
10	Expressing opinions	Books and literature	B1
12	Describing hopes and plans	Education	B2

ตารางที่ 2 การวิเคราะห์ข้อสอบการอ่านตามคำอธิบายกรอบมาตรฐาน CEFR (Council of Europe, 2020; North, Ortega, & Sheehan, 2011)

ลำดับข้อ	ทักษะที่ใช้	หัวข้อ	ระดับ CEFR
32	Identifying the purpose of an email	Job Interview	A2
47	Identifying the incorrect statement	News Report	B1
55	Identifying the author's tone toward the article	Academic Article	B2

ตารางที่ 3 การวิเคราะห์ข้อสอบความรู้เกี่ยวกับภาษาที่ใช้ในการสนทนาตามคำอธิบายกรอบมาตรฐาน CEFR (Council of Europe, 2020; North, Ortega, & Sheehan, 2011)

ลำดับข้อ	ทักษะที่ใช้	หัวข้อ	ระดับ CEFR
56	Describing things	Education	A2
62	Requests	Holidays	B1
76	Expressing agreement	Education	B2

ตารางที่ 4 การวิเคราะห์ข้อสอบความรู้เกี่ยวกับภาษา คำศัพท์ และไวยากรณ์ที่ใช้ในการเขียนตามคำอธิบายกรอบมาตรฐาน CEFR (Council of Europe, 2020; North, Ortega, & Sheehan, 2011)

ลำดับข้อ	ทักษะที่ใช้	หัวข้อ	ระดับ CEFR
81	Using quantifiers in simple structures correctly	Quantifiers	A2
90	Using tenses accurately in familiar contexts	Tenses	B1
100	Using conjunctions in complex grammatical forms	Conjunctions	B2

นอกจากนั้น ผู้พัฒนาข้อสอบยังได้ช่วยกันแก้ไขข้อสอบให้มีความยากง่ายสอดคล้องกับระดับมีคำถามและคำตอบที่ชัดเจน และมีตัวลวงที่เหมาะสม

มีโอกาสลวงให้ผู้เข้าสอบที่มีความสามารถต่ำมาเลือกได้ แบบทดสอบที่ผ่านการแก้ไขมีโครงสร้างดังที่แสดงในตารางที่ 5-8

ตารางที่ 5 โครงสร้างข้อสอบการฟัง (Council of Europe, 2020; North, Ortega, & Sheehan, 2011)

จำนวนข้อ	ระดับ CEFR	จำนวนบทสนทนา	ทักษะที่วัด
30 items	A2 = 9 items B1 = 16 items B2 = 5 items	<p><b>Pictures</b> (items 1-7)</p> <p><b>Question-Response</b> (items 8-13)</p> <p><b>Conversations</b> (items 14-19)</p> <ul style="list-style-type: none"> <li>- Conversation 1 (3 items)</li> <li>- Conversation 2 (3 items)</li> </ul> <p><b>Talks</b> (items 20-25)</p> <ul style="list-style-type: none"> <li>- Talk 1 (3 items)</li> <li>- Talk 2 (3 items)</li> </ul>	<ul style="list-style-type: none"> <li>- Can understand enough to be able to meet needs of a concrete type. (7 items)</li> <li>- Can understand straightforward factual information about common everyday or job-related topics, identifying both general messages and specific details. (12 items)</li> <li>- Can understand the main ideas of propositionally and linguistically complex discourse on both concrete and abstract topics delivered in standard language or a familiar variety, including technical discussions in their field of specialisation. (6 items)</li> </ul>

ตารางที่ 6 โครงสร้างข้อสอบการอ่าน (Council of Europe, 2020; North, Ortega, &amp; Sheehan, 2011)

จำนวนข้อ	ระดับ CEFR	จำนวนบทอ่าน	ทักษะที่วัด
30 items	A2 = 9 items B1 = 16 items B2 = 5 items	<b>Short Passages</b> (items 26-34) - Passage 1 (3 items) - Passage 2 (3 items) - Passage 3 (3 items) <b>Short Reading Passages</b> (items 35-47) - Passage 1 (4 items) - Passage 2 (3 items) - Passage 3 (3 items) - Passage 4 (3 items) <b>Extended Reading Passage</b> (items 48-55)	- Can identify the main ideas or suitable titles of the passage. (5 items) - Can identifying the main ideas and paragraph purposes. (3 items) - Can identify details in the passages. (10 items) - Can recognize implied information. (4 items) - Can identify the purpose of the passage. (2 items) - Can understand vocabulary in context. (4 items) - Can understand pronoun references. (1 item) - Can understanding the writer's tone and attitude. (1 item)

ตารางที่ 7 โครงสร้างข้อสอบความรู้เกี่ยวกับภาษาที่ใช้ในการสนทนา (Council of Europe, 2020; North, Ortega, & Sheehan, 2011)

จำนวนข้อ	ระดับ CEFR	จำนวนบทอ่าน	ทักษะที่วัด
25 items	A2 = 6 items B1 = 14 items B2 = 5 items	<b>Short Conversations</b> (items 56-61) - Dialogue 1 (3 items) - Dialogue 2 (3 items) - Dialogue 3 (3 items) <b>Medium Length Conversations</b> (items 62-75) - Dialogue 1 (4 items) - Dialogue 2 (3 items) - Dialogue 3 (3 items) - Dialogue 4 (4 items) <b>Extended Conversation</b> (items 76-80)	- Know how to interact in short conversations. (6 items) - Know how to communicate on familiar routine and non-routine matters related to their interests and professional field. (14 items) - Know how to use the language fluently and effectively on academic topics. (5 items)

ตารางที่ 8 โครงสร้างข้อสอบความรู้เกี่ยวกับภาษา คำศัพท์ และไวยากรณ์ที่ใช้ในการเขียน (Council of Europe, 2020; North, Ortega, & Sheehan, 2011)

จำนวนข้อ	ระดับ CEFR	จำนวนบทอ่าน	ทักษะที่วัด
20 items	A2 = 5 items B1 = 9 items B2 = 6 items	<b>Error Identification</b> (Items 81 - 92) <b>Cloze Test</b> (Items 93 – 100)	- Quantifiers (1 item) - Gerunds (1 item) - Conditional sentences (2 items) - Tenses (3 items) - Prepositions of time (1 item) - Adverbs (1 item) - Infinitives (1 item) - Agreement of subject and verb (2 items) - Phrasal verbs (1 item) - Modal verbs (1 item) - Word choices (3 items) - Conjunctions (3 items)

#### 4.3 การตรวจค่าความเที่ยงตรงทางเนื้อหา และการประเมินค่าคุณภาพด้านคุณประโยชน์ของแบบทดสอบ

คณะผู้วิจัยนำผลการประเมินค่าความสอดคล้องระหว่างข้อสอบกับวัตถุประสงค์หลายข้อ (Index of Item-Objective Congruence for Multidimensional Items) ของผู้เชี่ยวชาญมาวิเคราะห์ตามสูตรของ เทิร์นเนอร์และคาร์ลสัน (Turner & Carlson, 2003) ซึ่งเสนอว่าข้อสอบที่มีความเที่ยงตรงทางเนื้อหาต้องมีค่าความสอดคล้องอย่างน้อย .75 ในการประเมินครั้งแรก มีข้อสอบบางข้อที่มีค่าความสอดคล้องน้อยกว่า .75 คณะผู้พัฒนาจึงได้ปรับปรุงข้อสอบเดิมหรือเขียนข้อสอบขึ้นมาใหม่ จากนั้นจึงส่งให้ผู้เชี่ยวชาญประเมินอีกครั้ง จนได้ข้อสอบครบ 100 ข้อที่มีค่าความสอดคล้องมากกว่า .75

สำหรับการประเมินคุณภาพด้านคุณประโยชน์ของแบบทดสอบจำนวนหกด้าน คณะผู้วิจัยได้ให้ผู้เชี่ยวชาญทำการประเมินตามรายละเอียด ดังต่อไปนี้

1. ความเชื่อมั่น (reliability) การตรวจแบบทดสอบมีความเป็นแบบภววิสัย (objective) ไม่ว่าจะเป็นผู้ตรวจหรือมีการตรวจซ้ำ คะแนนที่ผู้เข้าสอบแต่ละคนได้ก็จะเหมือนเดิม

2. ความเที่ยงตรงทางโครงสร้าง (construct validity)

2.1 แบบทดสอบสามารถสะท้อนความสามารถในการฟังที่แท้จริงของผู้เข้าสอบได้

2.2 แบบทดสอบสามารถสะท้อนความสามารถในการอ่านที่แท้จริงของผู้เข้าสอบได้

2.3 แบบทดสอบสามารถสะท้อนความรู้ที่แท้จริงของผู้เข้าสอบเกี่ยวกับเรื่องภาษาที่ใช้ในการพูดได้

2.4 แบบทดสอบสามารถสะท้อนความรู้ที่แท้จริงของผู้เข้าสอบเกี่ยวกับเรื่องภาษาคำศัพท์ และไวยากรณ์ที่ใช้ในการเขียนได้

3. ความเหมือนจริง (authenticity)

3.1 บทอ่านและเรื่องที่ฟังในแบบทดสอบใช้ภาษาซึ่งเหมือนกับภาษาที่ใช้กันในสถานการณ์จริงในบริบทนั้น ๆ

3.2 ข้อสอบถามสิ่งสำคัญซึ่งผู้เข้าสอบต้องเข้าใจเมื่ออ่านบทอ่านหรือฟังเรื่องราวนั้น ๆ ในสถานการณ์จริง

4. การปฏิสัมพันธ์ (interactiveness)

4.1 แบบทดสอบกระตุ้นให้ผู้เข้าสอบสนใจทำแบบทดสอบเพราะบทอ่านและเรื่องที่ฟังสัมพันธ์กับระดับความรู้ทางภาษาของผู้เข้าสอบ

4.2 แบบทดสอบกระตุ้นให้ผู้เข้าสอบสนใจทำแบบทดสอบเพราะบทอ่านและเรื่องที่ฟังเชื่อมโยงกับความรู้ภูมิหลังของผู้เข้าสอบ

4.3 แบบทดสอบกระตุ้นให้ผู้เข้าสอบสนใจทำแบบทดสอบเพราะบทอ่านและเรื่องที่ฟังเชื่อมโยงกับความชอบความสนใจของผู้เข้าสอบ

5. ผลกระทบที่เกิดขึ้น (impact)

5.1 การทำแบบทดสอบนี้จะส่งผลดีต่อผู้เข้าสอบในเรื่องการพัฒนาทักษะภาษาอังกฤษ

5.2 การจัดสอบแบบทดสอบนี้จะส่งผลดีต่อการจัดการเรียนการสอนภาษาอังกฤษของหลักสูตร

6. การนำไปใช้ได้จริง (practicability)

6.1 สาขาวิชาภาษาอังกฤษมีบุคลากรมากพอที่จะพัฒนา จัดสอบ และตรวจคะแนนจากการทำแบบทดสอบนี้

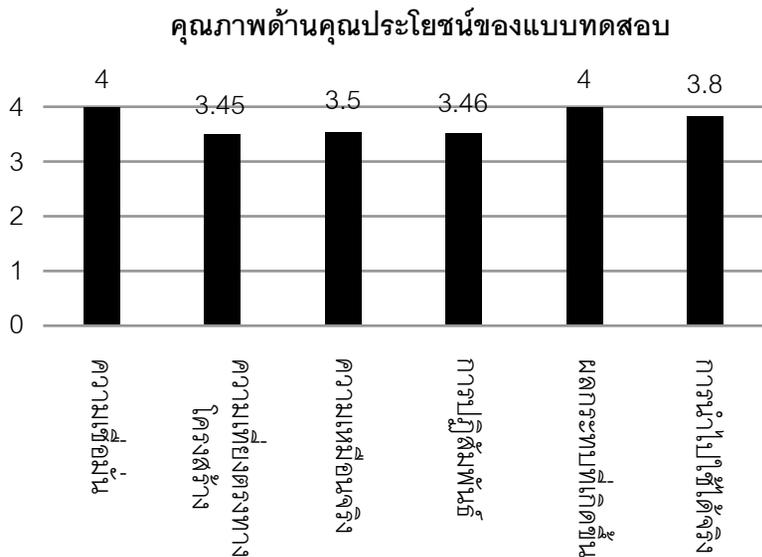
6.2 สาขาวิชาภาษาอังกฤษมีวัสดุอุปกรณ์มากพอที่จะพัฒนา จัดสอบ และตรวจคะแนนจากการทำแบบทดสอบนี้

6.3 คณาจารย์สาขาวิชาภาษาอังกฤษมีเวลามากพอที่จะพัฒนา จัดสอบ และตรวจคะแนนจากการทำแบบทดสอบนี้

คณะผู้วิจัยตีความค่าคุณภาพด้านคุณประโยชน์ของแบบทดสอบจากการประเมินของผู้เชี่ยวชาญโดยแบ่งเป็น 4 ระดับตามที่ ทอดด์ (Todd, 2011) เสนอ ดังนี้

1.00 - 1.75	ไม่เห็นด้วย
1.76 - 2.50	ไม่มีความคิดเห็นไปในทางใดทางหนึ่ง
2.51 - 3.25	เห็นด้วย
3.26 - 4.00	เห็นด้วยอย่างยิ่ง

แผนภูมิที่ 1 แสดงผลการวิเคราะห์ค่าคุณภาพด้านคุณประโยชน์ของแบบทดสอบภาษาอังกฤษของหลักสูตร



ในภาพรวม ผู้เชี่ยวชาญ “เห็นด้วยอย่างยิ่ง” (3.7) ว่าแบบทดสอบความสามารถภาษาอังกฤษของหลักสูตรมีคุณภาพด้านคุณประโยชน์ ด้านที่ได้คะแนนเฉลี่ยสูงที่สุด (4) มีด้วยกันสองด้าน คือ ด้านความเชื่อมั่นและด้านผลกระทบที่เกิดขึ้น ผู้เชี่ยวชาญเห็นว่า เนื่องจากแบบทดสอบมีลักษณะ

เป็นการเลือกคำตอบซึ่งมีคำตอบที่ถูกต้องเพียงข้อเดียว การตรวจแบบทดสอบจึงมีความเป็นแบบภววิสัย (objective) ไม่ว่าใครจะเป็นผู้ตรวจหรือมีการตรวจซ้ำ คะแนนที่ผู้เข้าสอบแต่ละคนได้ก็จะเหมือนเดิม และการสอบแบบทดสอบจะทำให้ผู้เข้าสอบและอาจารย์ผู้รับผิดชอบหลักสูตรได้ข้อมูล

สะท้อนกลับเกี่ยวกับความสามารถทางภาษาอังกฤษของผู้เข้าสอบแต่ละคนตามกรอบมาตรฐาน CEFR รวมถึงทราบว่าผู้เข้าสอบทำคะแนนในส่วนใดได้มาก และส่วนใดได้น้อย ข้อมูลจากการสอบจะทำให้ผู้เข้าสอบและอาจารย์ทราบว่าผู้เข้าสอบมีทักษะใดที่ควรได้รับการพัฒนาต่อไป

คุณภาพด้านคุณประโยชน์ซึ่งได้คะแนนจากการประเมินน้อยที่สุดสองอันดับแรก คือ ด้านความเที่ยงตรงทางโครงสร้าง (3.45) และด้านการปฏิสัมพันธ์ (3.46) สาเหตุที่คุณภาพด้านคุณประโยชน์สองด้านนี้ได้คะแนนน้อยเนื่องจากผู้เชี่ยวชาญเห็นว่าแบบทดสอบเป็นการวัดความรู้เกี่ยวกับภาษาที่ใช้ในการพูดและภาษาที่ใช้ในการเขียน ไม่ใช่การวัดทักษะทั้งสองโดยตรง นอกจากนี้ผู้เชี่ยวชาญยังได้แนะนำให้เพิ่มบทอ่านซึ่งผู้เข้าสอบน่าจะสนใจให้มากขึ้น เช่น บทอ่านเกี่ยวกับเทคโนโลยีสมัยใหม่

#### 4.4 การตรวจสอบความถูกต้องของภาษา

คณะผู้วิจัยนำแบบทดสอบไปให้อาจารย์ชาวแคนาดา ซึ่งมีประสบการณ์การสอนภาษาอังกฤษมาแล้วมากกว่า 10 ปี ตรวจสอบความถูกต้องและความเป็นธรรมชาติของภาษาก่อนการทดลองสอบ

#### 4.5 การทดลองสอบ การสัมภาษณ์ ผู้ทดลองสอบ และการปรับวิธีการจัดสอบตามข้อมูลจากการสัมภาษณ์

คณะผู้วิจัยนำแบบทดสอบไปทดลองสอบกับอาสาสมัครซึ่งเป็นนักศึกษาหลักสูตรภาษาอังกฤษ มหาวิทยาลัยราชภัฏเชียงใหม่ จำนวน 11 คน และสัมภาษณ์ผู้ทดลองสอบด้วยคำถามในแบบบันทึกการสัมภาษณ์แบบกึ่งมีโครงสร้าง ซึ่งมีคำถามหลักจำนวน 6 คำถาม ดังนี้

1. คุณมีความคิดเห็นหรือข้อเสนอแนะอย่างไรบ้างเกี่ยวกับการกำหนดให้ทำแบบทดสอบในระยะเวลาสองชั่วโมง
2. คุณมีความคิดเห็นหรือข้อเสนอแนะอย่างไรบ้างเกี่ยวกับระบบเสียงที่ใช้สำหรับการทดสอบทักษะการฟัง
3. คุณมีความคิดเห็นหรือข้อเสนอแนะอย่างไรบ้างเกี่ยวกับภาพที่ใช้ประกอบข้อสอบการฟัง
4. คุณมีความคิดเห็นหรือข้อเสนอแนะอย่างไรบ้างเกี่ยวกับขนาดตัวอักษรที่ใช้ในแบบทดสอบ
5. คุณมีความคิดเห็นหรือข้อเสนอแนะอย่างไรบ้างเกี่ยวกับการถ่ายสำเนาแบบทดสอบ
6. คุณมีความคิดเห็นหรือข้อเสนอแนะอย่างไรบ้างเกี่ยวกับห้องที่ใช้จัดสอบ

จากการสัมภาษณ์ ผู้ทดลองสอบเห็นว่าเวลาที่ใช้ในการสอบอาจน้อยเกินไป ผู้ทดลองสอบจำนวนหนึ่งบอกว่าตนทำข้อสอบบางส่วนไม่ทัน สำหรับเรื่องระบบเสียงสำหรับข้อสอบการฟังนั้น ผู้ทดลองสอบบอกว่าเสียงดังและชัดเจนดี สำหรับเรื่องคุณภาพของรูปภาพที่ใช้ในการทดสอบการฟัง ผู้ทดลองสอบบอกว่ามีรูปภาพบางรูปที่พื้นหลังค่อนข้างดำ แต่โดยรวมแล้วภาพที่ใช้มีความชัดเจน ไม่เป็นอุปสรรคต่อการทดสอบ สำหรับเรื่องตัวอักษร ผู้ทดลองสอบเห็นว่ามีความเหมาะสม ตัวอักษรมีขนาดไม่เล็กจนเกินไป สำหรับเรื่องคุณภาพของการถ่ายสำเนา ผู้ทดลองสอบเห็นว่าภาพถ่ายสำเนาชัดเจน สามารถอ่านได้ตามปกติ และสำหรับเรื่องห้องที่ใช้จัดสอบ ผู้ทดลองสอบเห็นว่าห้องสอบเก็บเสียงได้ดี อากาศเย็นสบายเพราะมีเครื่องปรับอากาศ เมื่อพิจารณาความเห็นของผู้ทดลองสอบในประเด็นทั้งหมด จะเห็นได้ว่าการจัดสอบโดยภาพรวมดำเนินไปอย่างมีประสิทธิภาพ ไม่เป็นอุปสรรคต่อการทดสอบ ไม่มี

ปัจจัยใดที่จะส่งผลสำคัญต่อความเชื่อมั่นและความถูกต้องของคะแนนสอบ

เนื่องจากในการทดลองสอบ มีผู้ทดลองสอบที่ทำแบบทดสอบไม่เสร็จภายในสองชั่วโมงจำนวนสองคน เพื่อจัดการกับปัญหาการทำแบบทดสอบไม่เสร็จตามเวลา ผู้วิจัยจึงได้กำหนดแนวทางการปฏิบัติในการคุมสอบเพิ่มขึ้น คือ ข้อที่หนึ่ง ให้ผู้คุมสอบบอกให้ผู้เข้าสอบเขียนข้อมูลส่วนตัวของตนเองบนกระดาษคำตอบให้เรียบร้อยก่อนแจกแบบทดสอบและเริ่มจับเวลาสอบ และข้อที่สอง ให้ผู้คุมสอบบอกเตือนเวลาเมื่อเหลือเวลาในการสอบ 30 นาที และบอกเตือนเวลาอีกครั้งเมื่อเหลือเวลาในการสอบ 5 นาที

#### 4.6 การสอบ

คณะผู้วิจัยใช้วิธีการคัดเลือกแบบเจาะจง (purposive sampling) เลือกนักศึกษามาสอบจำนวน 100 คน โดยเลือกแบบคละหลักสูตร คละชั้นปี และคละความสามารถตามจำนวนที่เชื่อว่าสะท้อนความสามารถของนักศึกษาหลักสูตรภาษาอังกฤษ มหาวิทยาลัยราชภัฏเชียงใหม่ในภาพรวมได้ ผู้วิจัยเลือกนักศึกษาที่ได้คะแนนเฉลี่ยรายวิชาเอกภาษาอังกฤษ 3.50 หรือมากกว่ามาจำนวน 25 คน เลือกนักศึกษาที่ได้คะแนนเฉลี่ยรายวิชาเอกภาษาอังกฤษ 2.75-3.49 มาจำนวน 50 คน และเลือกนักศึกษาที่ได้คะแนนเฉลี่ยรายวิชาเอกภาษาอังกฤษ 2.74 หรือน้อยกว่ามาจำนวน 25 คน

มีงานวิจัยที่นำผลการสอบครั้งนี้ไปทำการกำหนดมาตรฐาน (standard setting) ด้วยวิธี Yes/No Angoff โดยกำหนดให้คะแนนจุดตัดระดับ A2 อยู่ที่ 26 คะแนน ระดับ B1 อยู่ที่ 64 คะแนน และระดับ B2 อยู่ที่ 84 คะแนน ผลการวิจัยครั้งนั้นแสดงให้เห็นว่าการจัดกลุ่มผู้เข้าสอบตามเกณฑ์มาตรฐาน CEFR จากคะแนนการสอบและการจัดกลุ่ม

ผู้เข้าสอบตามประสบการณ์ของอาจารย์ที่เคยสอน นักศึกษาผู้เข้าสอบมีค่าสหสัมพันธ์ของเพียร์สัน (Pearson's correlation coefficient) อยู่ที่ .63 (ความสัมพันธ์ค่อนข้างสูง) (ดูรายละเอียดเพิ่มเติมที่ รัชฎา นุเสณ, จุติญาณี เพิ่มทันจิตต์, ปิยะณัฐ จันตะคาต, ปิยะพันธ์ กันทิสา และลลิตา วิบูลวิชร, 2568)

นอกจากนั้น การกำหนดมาตรฐานยังแสดงให้เห็นว่านักศึกษาผู้เข้าสอบส่วนใหญ่ (ร้อยละ 62) มีความสามารถทางภาษาอังกฤษอยู่ในระดับ A2 ซึ่งต่ำกว่าความคาดหวังของคณะกรรมการมาตรฐานการอุดมศึกษา และผู้เข้าสอบได้คะแนนเฉลี่ยในแบบทดสอบส่วนความรู้เกี่ยวกับภาษาที่ใช้ในการเขียนน้อยที่สุด คณะผู้วิจัยได้แจ้งให้อาจารย์ผู้รับผิดชอบหลักสูตรภาษาอังกฤษทราบข้อมูลสะท้อนกลับที่ได้จากการสอบแล้วเพื่อให้ผู้เกี่ยวข้องหาทางจัดการเรียนการสอนและโครงการที่ช่วยพัฒนาความสามารถทางภาษาอังกฤษของนักศึกษาอย่างเหมาะสมต่อไป

#### 4.7 การวิเคราะห์ข้อมูลเพื่อหาคุณภาพข้อสอบ

การวิเคราะห์ผลการสอบแสดงให้เห็นว่าแบบทดสอบทั้งฉบับมีค่าความเชื่อมั่น (reliability) ตามสูตร KR-20 อยู่ที่ .93 สูงกว่าที่แมกโคเวนและแมกโคเวน (McCowan & McCowan, 1999) เสนอว่าค่าความเชื่อมั่นขั้นต่ำที่แบบทดสอบควรจะมีคือ .80 แบบทดสอบมีค่าความยากง่าย (p) อยู่ที่ .57 ซึ่งสมนึก ภัททิยธนี (2565) เสนอว่าเป็นค่าความยากง่ายที่ดีมาก ข้อสอบมีความยากง่ายพอเหมาะ ค่าอำนาจจำแนกของแบบทดสอบ (r) คือ .46 สูงกว่าที่แมกโคเวนและแมกโคเวน (McCowan & McCowan, 1999) เสนอว่าข้อสอบที่ดีควรมีค่าอำนาจจำแนกอย่างน้อย .20 คะแนนสูงสุดจากการสอบครั้งนี้คือ 92 คะแนน คะแนนต่ำสุดคือ 21

คะแนน แบบทดสอบมีค่าเบี่ยงเบนมาตรฐาน (SD) อยู่ที่ 16.64 และค่าความคลาดเคลื่อนของการวัด (SEM) อยู่ที่ 4.20

## 5. การอภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อให้ได้แบบทดสอบความสามารถทางภาษาอังกฤษของนักศึกษาหลักสูตรภาษาอังกฤษ มหาวิทยาลัยราชภัฏเชียงใหม่ ที่มีคุณภาพ การวิจัยเริ่มต้นจากการกำหนดคุณลักษณะเฉพาะของแบบทดสอบ จากนั้นผู้พัฒนาข้อสอบได้แต่งข้อสอบ ประชุมแก้ไขข้อสอบ และเชิญอาจารย์ชาวแคนาดามาตรวจสอบความถูกต้องของภาษา ในขั้นตอนต่อไป คณะผู้วิจัยได้ส่งแบบทดสอบไปให้ผู้เชี่ยวชาญตรวจสอบเพื่อหาค่าความเที่ยงตรงทางเนื้อหาและการประเมินค่าคุณภาพด้านคุณประโยชน์ของแบบทดสอบข้อสอบทุกข้อในแบบทดสอบมีค่า IOC ตามสูตรของเทิร์นเนอร์และคาร์ลสัน (Turner & Carlson, 2003) มากกว่า .75 และแบบทดสอบมีคุณภาพด้านคุณประโยชน์ในภาพรวมที่ 3.7 ซึ่งหมายความว่าผู้เชี่ยวชาญเห็นด้วยอย่างยิ่งว่าแบบทดสอบมีคุณภาพด้านคุณประโยชน์ หลังจากนั้นคณะผู้วิจัยนำแบบทดสอบไปให้นักศึกษาจำนวน 11 คนทดลองสอบ เพื่อให้แน่ใจว่าการจัดสอบมีประสิทธิภาพ ไม่มีเป็นอุปสรรคต่อความเชื่อมั่นและความเที่ยงตรงของการทดสอบ จากการสัมภาษณ์ ผู้ทดลองสอบมีความเห็นว่าการจัดสอบเป็นอย่างไรมีประสิทธิภาพ ในขั้นตอนต่อมา คณะผู้วิจัยได้ให้นักศึกษาจำนวน 100 คนทดสอบแบบทดสอบที่พัฒนาขึ้นและนำข้อมูลมาวิเคราะห์ข้อมูลเพื่อหาคุณภาพข้อสอบ ผลการวิเคราะห์พบว่า แบบทดสอบมีค่าความเชื่อมั่นตาม

สูตร KR-20 อยู่ที่ .93 มีค่าความยากง่ายอยู่ที่ .57 มีค่าอำนาจจำแนกอยู่ที่ .46 และมีค่าเบี่ยงเบนมาตรฐานอยู่ที่ 16.64

ขั้นตอนการพัฒนาแบบทดสอบของงานวิจัยนี้ดำเนินไปโดยใช้ข้อเสนอของสหภาพยุโรปว่าด้วยการพัฒนาแบบทดสอบตามเกณฑ์มาตรฐาน CEFR (Council of Europe, 2011) เป็นแนวทางสอดคล้องกับกาญจนา ชีวาสุขถาวร (Cheewasukthawon, 2022) ซึ่งใช้ข้อเสนอดังกล่าวเช่นกันเป็นแนวทางในการพัฒนาแบบทดสอบความสามารถทางภาษาอังกฤษของมหาวิทยาลัยแห่งหนึ่ง

แบบทดสอบภาษาอังกฤษของหลักสูตรเป็นแบบทดสอบปรนัยสี่ตัวเลือกจำนวน 100 ข้อ ซึ่งกาญจนา ชีวาสุขถาวร (Cheewasukthawon, 2022) เห็นว่ามีความเหมาะสม ข้อสอบมีจำนวนไม่มากเกินไปจนอาจส่งผลให้ผู้เข้าสอบเหนื่อยล้าหรือเกิดความกังวล แบบทดสอบภาษาอังกฤษของหลักสูตรวัดความสามารถทางการฟังและการอ่านโดยตรง และวัดทักษะการพูดและการเขียนโดยอ้อม ซึ่งเป็นวิธีการทดสอบเดียวกันกับที่ เกียรตินาวิน ศรีธัญรัตน์, ศุภกาญจน์ ผาทอง, ทศพล สุรศักดิ์รินทร์ และอมรรัตน์ อัมราลิขิต (Sridhanyarat et al., 2021) ใช้ในการทดสอบความสามารถทางภาษาอังกฤษของมหาวิทยาลัยแห่งหนึ่ง การวัดทักษะในการผลิตภาษาโดยทางอ้อมมักมีเหตุผลมาจากคุณประโยชน์ของแบบทดสอบในด้านการนำไปใช้ได้จริง (practicability) เนื่องจากการวัดความสามารถทางการพูดและการเขียนโดยตรงต้องใช้เวลาและงบประมาณในฝึกอบรมผู้ประเมินและการตรวจข้อสอบมาก

แบบทดสอบวัดความสามารถทางภาษานี้ออกแบบมาสำหรับวัดความสามารถทางภาษาอังกฤษของนักศึกษาสาขาวิชาภาษาอังกฤษของมหาวิทยาลัยราชภัฏเชียงใหม่โดยเฉพาะ ซึ่งตามประสบการณ์ของคณะผู้วิจัย นักศึกษามีความสามารถทางภาษาตามกรอบมาตรฐาน CEFR ในระดับ A2 B1 และ B2 ต่างจากแบบทดสอบอื่นที่ออกแบบมาเพื่อวัดความสามารถของนักศึกษาทั้งมหาวิทยาลัย ซึ่งผู้เข้าสอบย่อมมีความสามารถที่ต่างกันมากจึงต้องวัดความสามารถในระดับ A1 และระดับ C1 ด้วย (Balleras, Yippikun, Castro, & Prasertsak, 2025; Waluyo, Zahabi, & Ruangsung, 2024; Cheewasukthawon, 2022) ซึ่งสำหรับกลุ่มเป้าหมายของแบบทดสอบนี้เป็นระดับที่ต่ำและสูงเกินความจำเป็น

แบบทดสอบภาษาอังกฤษของหลักสูตรได้รับการตรวจสอบคุณภาพโดยการวิเคราะห์ความเที่ยงตรงทางเนื้อหาโดยการคำนวณค่า IOC เช่นเดียวกับงานวิจัยที่ผ่านมาจำนวนหนึ่ง (Cheewasukthawon, 2022, จันทิมา ชูวานนท์, 2561) งานวิจัยนี้ได้ใช้สูตรการคำนวณของเทิร์นเนอร์และคาร์ลสัน (Turner & Carlson, 2003) และข้อเสนอของทั้งสองซึ่งระบุว่าความเที่ยงตรงทางเนื้อหาต้องมีค่าความสอดคล้องอย่างน้อย .75 ค่าความสอดคล้อง .75 เป็นค่าเดียวกับที่เกี่ยวข้องในวิน ศรีธัญรัตน์, ศุภกาญจน์ ผาทอง, ทศพล สุรนครศรีรินทร์ และอมรรัตน์ อัมราลิขิต (Sridhanyarat et al., 2021) ใช้สาเหตุที่งานวิจัยนี้เลือกใช้สูตรการคำนวณของเทิร์นเนอร์และคาร์ลสันเนื่องจากต้องการวัดความเที่ยงตรงทั้งทางเนื้อหาและระดับมาตรฐาน CEFR ที่ข้อสอบแต่ละข้อวัด

นอกจากนั้น แบบทดสอบภาษาอังกฤษของหลักสูตรยังได้รับการตรวจสอบคุณภาพผ่านการหาค่าความเชื่อมั่นตามสูตร KR-20 ค่าความยากง่าย และค่าอำนาจจำแนก เช่นเดียวกับงานวิจัยที่ผ่านมาหลายเรื่อง (Cheewasukthawon, 2022; จตุพงศ์ โมรา และคณะ, 2564; Sridhanyarat et al., 2021; จันทิมา ชูวานนท์, 2561)

สำหรับการประเมินคุณภาพด้านคุณประโยชน์ของแบบทดสอบเป็นการประเมินคุณภาพทางด้านของแบบทดสอบตามที่ แบชแมน และพาล์มเมอร์ (Bachman & Palmer, 1996) เสนอ อังชรินทร์ทองปาน และคณะ (2564) ได้ใช้กรอบความคิดของแบชแมน และพาล์มเมอร์ เช่นกันในการตรวจสอบคุณภาพของแบบทดสอบความสามารถทางภาษาอังกฤษ อย่างไรก็ตาม ในขณะที่ยังคงคุณภาพด้านคุณประโยชน์ในงานวิจัยนี้ได้ผลการประเมินสูงที่สุดคือ ค่าความเชื่อมั่น และด้านที่มีผลการประเมินต่ำที่สุดคือ ความเที่ยงตรงทางโครงสร้างแบบทดสอบของอังชรินทร์และคณะมีค่าคุณภาพด้านคุณประโยชน์ที่ตรงกันข้าม โดยด้านที่ได้ผลการประเมินสูงที่สุดคือ ความเที่ยงตรงทางโครงสร้างและด้านที่ได้ผลการประเมินต่ำที่สุดคือ ค่าความเชื่อมั่น เหตุผลที่การประเมินออกมาแตกต่างกันนี้น่าจะมาจากสาเหตุที่แบบทดสอบของอังชรินทร์และคณะเป็นแบบทดสอบซึ่งวัดทักษะการเขียนโดยตรง ผู้เข้าสอบต้องเขียนคำตอบตามคำสั่งและจำนวนคำที่กำหนด ดังนั้นจึงมีความเที่ยงตรงทางโครงสร้างมากกว่างานวิจัยนี้ แต่การให้คะแนนข้อสอบเขียนก็มีโอกาสที่การให้คะแนนจะต่างกันไป ซึ่งอาจส่งผลต่อความเชื่อมั่น

## 6. ข้อเสนอแนะ

หากจะมีการพัฒนาแบบทดสอบวัดความสามารถทางภาษาอังกฤษและการวิจัยในเรื่องนี้อีกในอนาคต ผู้วิจัยควรเพิ่มเนื้อหาที่มีบริบทระดับท้องถิ่น (local context) ในทุกทักษะ เพื่อให้ข้อสอบมีความใกล้เคียงกับบริบทการใช้ภาษาอังกฤษในชีวิตจริงของผู้เข้าสอบให้มากขึ้น และควรเพิ่มการทดสอบความสามารถทางการพูดและการเขียนโดยตรง เพื่อ

ให้แบบทดสอบมีลักษณะเป็นการวัดสมรรถนะทางภาษา (performance based) โดยตรงมากขึ้น อย่างไรก็ตาม การจัดสอบการเขียนและการพูดโดยตรงจะมีความท้าทายในหลายเรื่องซึ่งผู้พัฒนาแบบทดสอบต้องนำมาพิจารณา ไม่ว่าจะเป็นเวลาในการฝึกอบรมผู้ประเมิน ความเชื่อมั่นต่อการให้คะแนนของผู้ประเมิน และงบประมาณที่จะต้องใช้งามากขึ้น

## References

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- 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 English Language and Linguistics*, 6(2), 169-189.
- Cheewasukthawon, K. (2022). Developing a standardized English proficiency test in alignment with the CEFR. *PASAA*, 63, 66-92.
- Chuwarnond, C. (2018). Development of English proficiency test for DRU students. *Academic Journal for the Humanities and Social Sciences, Dhonburi Rajabhat University*, 1(1), 9-23. (in Thai)
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment, companion volume*. Retrieved from <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>
- Council of Europe. (2011). *Manual for language test development and examining*. Retrieved from <https://www.coe.int/en/web/common-european-framework-reference-languages/developing-tests-examining>
- Council of Europe. (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR)*. Retrieved from <https://www.coe.int/en/web/common-european-framework-reference-languages/relating-examinations-to-the-cefr>
- Hughes, A. & Hughes, J. (2020). *Testing for language teachers*. Cambridge University Press.
- McCowan, R. J., & McCowan, S. C. (1999). *Item analysis for criterion-referenced tests*. Center for Development of Human Services.

- Mora, J., Torut, B., Torut, S., & Nilakorn, P. (2021). Development of a university English proficiency test. *Journal of Suvarnabhumi Institute of Technology*, 7(2), 368-386. (in Thai)
- North, B., Ortega, A. & Sheehan, S. (2011). *EAQUALS core inventory for general English*. Retrieved from [https://www.eaquals.org/wpcontent/uploads/EAQUALS\\_British\\_Council\\_Core\\_Curriculum\\_April2011.pdf](https://www.eaquals.org/wpcontent/uploads/EAQUALS_British_Council_Core_Curriculum_April2011.pdf)
- Nusen, R., Permthanjit, J., Chantakhat, P., Kantisa, P., & Wiboonwachara, L. (2025). Yes/No Angoff standard setting on an English proficiency test: Setting cut-off scores and analyzing standard setting validation. *Journal of Social Sciences in Measurement, Evaluation, Statistics and Research*, 6(1), 71-90. (in Thai)
- Pattiyanthani, S. (2022). *Educational measurement*. Prasan Publishing. (in Thai)
- Sridhanyarat, K., Panthong, S., Suranakkharin, T., & Ammaralikit, A. (2021). The development of STEP, the CEFR-based English proficiency test. *English Language Teaching*, 14(7), 95-106.
- The Commission on Higher Education Standards, Ministry of Higher Education, Science, Research and Innovations. (2024). *Announcement from the Higher Education Standards Committee on the policy for upgrading English standard in Higher Education B.E. 2567*. Retrieved from <https://www.ops.go.th/th/ches-downloads/edu-standard/item/9625-2567> (in Thai)
- Thongpan, A., Pantawee, T., Banbang, C., Wongnak, D., & Choolserm, S. (2021). Development of English proficiency tests for Roi Et Rajabhat University's graduate students. *Journal of Educational Measurement, Mahasarakham University*, 27(2), 336-352. (in Thai)
- Todd, R. W. (2011). Analyzing and interpreting rating scale data from questionnaires. *rEFLECTIONS*, 14, 69–77.
- Turner, R. & Carlson, L. A. (2003). Indexes of item-objective congruence for multidimensional items. *International Journal of Testing*, 3(2), 163-172.
- Waluyo, B., Zahabi, A., & Ruangsung, L. (2024). Language assessment at a Thai university: A CEFR-based test of English proficiency development. *rEFLECTIONS*, 31(1), 25-47.