

การกำหนดค่ามาตรฐานแบบทดสอบความสามารถทางภาษาอังกฤษด้วยวิธี Yes/No Angoff
: การหาคะแนนจุดตัดและการวิเคราะห์ความถูกต้องของการกำหนดมาตรฐาน
Yes/No Angoff Standard Setting on an English Proficiency Test: Setting
Cut-Off Scores and Analyzing Standard Setting Validation

รชณ นูเสน^{1*}, จุติยาณี เพ็มหันจิตต์², ปิยะณัฐ จันทะคาต³, ปิยะพันธ์ กันทิสา⁴ และ ลลิดา วิบูลวัชร⁵
Rachod Nusen^{1*}, Jutiyanee Permethanjit², Piyanat Chantakhat³, Piyapan Kantisa⁴,
and Lalida Wiboonwachara⁵

¹PhD (English), ผศ. ดร. คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏเชียงใหม่

Corresponding author E-mail: rachod_nus@cmru.ac.th

²ศษ.ม. (การสอนภาษาอังกฤษ) อาจารย์ คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏเชียงใหม่

E-mail: jutiyanee@cmru.ac.th

³M.A. (Education), อาจารย์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏเชียงใหม่

E-mail: piyanat_cha@cmru.ac.th

⁴ปร.ด. (การสอนภาษาอังกฤษ) อาจารย์ คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏเชียงใหม่

E-mail: jay_avenue@hotmail.com

⁵ปร.ด. (ภาษาอังกฤษเพื่อการพัฒนวิชาชีพ) อาจารย์ คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏเชียงใหม่

E-mail: lalida_wib@cmru.ac.th

Received: July 23, 2024 / Revised: July 16, 2024 / Accepted: April 4, 2025

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) กำหนดคะแนนจุดตัดตามกรอบมาตรฐาน CEFR ระดับ A2 B1 และ B2 โดยวิธี Yes/No Angoff ให้กับแบบทดสอบความสามารถทางภาษาอังกฤษที่ได้รับการพัฒนาโดยคณาจารย์ สาขาวิชาภาษาอังกฤษ มหาวิทยาลัยราชภัฏเชียงใหม่ และ 2) วิเคราะห์ความถูกต้องของการกำหนดมาตรฐาน และคะแนนจุดตัดที่กำหนด ผู้ประเมินที่เข้าร่วมการกำหนดมาตรฐานครั้งนี้มีทั้งสิ้น 10 คน เครื่องมือที่ใช้ในการวิจัย ได้แก่ แบบทดสอบความสามารถทางภาษาอังกฤษของหลักสูตร (คะแนนเต็มเท่ากับ 100 คะแนน) แบบทดสอบความเข้าใจในเกณฑ์มาตรฐาน CEFR แบบสำรวจความคิดเห็นของผู้ประเมินต่อการอบรมก่อนการกำหนดมาตรฐาน แบบประเมินข้อสอบโดยวิธี Yes/No Angoff และแบบสำรวจความคิดเห็นของผู้ประเมินต่อกิจกรรมการกำหนดมาตรฐาน ผลการวิจัยมีดังนี้ 1) คะแนนจุดตัดตามกรอบมาตรฐาน CEFR ของแบบทดสอบความสามารถทางภาษาอังกฤษของหลักสูตรในระดับ A2 อยู่ที่ 26 คะแนน ระดับ B1 อยู่ที่ 64 คะแนน และระดับ B2 อยู่ที่ 84 คะแนน 2) ความถูกต้องภายในของกระบวนการกำหนดมาตรฐานและคะแนนจุดตัดที่กำหนดได้รับการตรวจสอบผ่านการหาค่าสหสัมพันธ์ภายในชั้น (ICC) ของการประเมิน (A2 = .92, B1 = .92 และ B2 = .81) สำหรับความถูกต้องภายนอกได้รับการตรวจสอบผ่านวิธีการสามวิธี

สำหรับวิธีที่หนึ่งเป็นการอภิปรายความเหมาะสมของคะแนนจุดโดยตัวแทนอาจารย์ผู้รับผิดชอบหลักสูตร วิธีที่สองคือการเปรียบเทียบการจัดกลุ่มผู้เข้าสอบตามเกณฑ์มาตรฐาน CEFR ของอาจารย์ผู้มีประสบการณ์เคยสอนผู้เข้าสอบกับการจัดกลุ่มโดยการกำหนดคะแนนจุดตัด (ค่าสหสัมพันธ์ของเพียร์สัน = .63) และวิธีที่สามคือการประเมินความเที่ยงตรงของกระบวนการโดยการถามความเห็นของผู้ประเมินเกี่ยวกับกระบวนการกำหนดมาตรฐานที่ตนได้เข้าร่วม พบว่าผู้ประเมินเห็นด้วยอย่างยิ่ง (คะแนนเฉลี่ย = 3.78) กับกระบวนการกำหนดมาตรฐานในภาพรวม

คำสำคัญ : แบบทดสอบความสามารถทางภาษาอังกฤษ, การกำหนดมาตรฐานแบบ Yes/No Angoff, ความถูกต้องของการกำหนดมาตรฐาน

Abstract

This research aimed to determine cut-off scores based on the CEFR framework levels A2, B1, and B2 using the Yes/No Angoff method for an English proficiency test developed by the English Department at Chiang Mai Rajabhat University, and to analyze the validation of the standard setting and the determined cut-off scores. There were ten judges in this study. The research instruments included an English proficiency test with a total score of 100, a CEFR standard comprehension test, a survey of judges' opinions on the pre-standard setting training, the Yes/No Angoff evaluation form, and a survey of judges' opinions on the standard setting process. The research findings indicated that the cut-off scores for the aforementioned English proficiency test, based on the CEFR framework, were set at 26 points for level A2, 64 points for level B1, and 84 points for level B2. The internal validation of the standard setting process and the assigned cut-off scores was verified through the Intraclass Correlation Coefficient (ICC) calculations, showing results of .92 for A2, .92 for B1, and .81 for B2. External validation, on the other hand, was examined using three methods: the first method involved focus group discussions on the reasonableness of the cut-off scores with representatives of curriculum administrators and lecturers, the second method compared the grouping of test takers according to the CEFR levels by an experienced lecturer with the grouping based on the determined cut-off scores (Pearson correlation = .63), and the third method assessed the procedural validation by asking judges for their opinions on the standard setting process they participated in. It was found that the judges strongly agreed (mean = 3.78) with the overall standard setting process.

Keywords: English Proficiency Test, Yes/No Angoff Standard Setting, Standard Setting Validation

บทนำ

ความสามารถในการใช้ภาษาอังกฤษเป็นสิ่งที่ช่วยเพิ่มศักยภาพของบุคคลทั้งทางวิชาการและวิชาชีพ ดังนั้นหน่วยงานจำนวนมาก ทั้งภาครัฐและเอกชน จึงมักกำหนดให้ความสามารถทางภาษาอังกฤษเป็นเกณฑ์หนึ่งในการตัดสินใจในเรื่องต่างๆ เช่น การรับพนักงานใหม่เข้าทำงาน หรือการคัดเลือกบุคคลเข้าศึกษาต่อ เกณฑ์การวัดความสามารถทางภาษาเกณฑ์หนึ่งที่หน่วยงานในประเทศไทยนิยมใช้ได้แก่ Common European Framework of Reference for Languages (CEFR) ซึ่งพัฒนาขึ้นโดยสหภาพยุโรป กรอบมาตรฐานดังกล่าวแบ่งผู้ใช้ภาษาออกเป็น 3 กลุ่ม คือ กลุ่ม basic user (A1 และ A2) กลุ่ม independent user (B1 และ B2) และกลุ่ม proficient user (C1 และ C2)

สำหรับประเทศไทยนั้น ใน พ.ศ. 2557 สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ได้กำหนดแนวทางให้สถาบันการศึกษาในประเทศไทยใช้ CEFR เป็นกรอบมาตรฐานในการวัดความสามารถทางภาษาอังกฤษของผู้เรียน โดยตั้งเกณฑ์ให้ผู้สำเร็จการศึกษาระดับประถมศึกษา มีความสามารถทางภาษาอังกฤษอย่างน้อยในระดับ A1 ส่วนผู้สำเร็จการศึกษาระดับมัธยมศึกษาตอนต้นมีความสามารถทางภาษาอังกฤษอย่างน้อยในระดับ A2 และผู้สำเร็จการศึกษาระดับมัธยมศึกษาตอนปลายมีความสามารถทางภาษาอังกฤษอย่างน้อยในระดับ B1 (Wudthayagorn, 2018) ต่อมาคณะกรรมการมาตรฐานการอุดมศึกษา (2567) กระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม ก็ได้ออกประกาศเรื่อง นโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา ความว่าให้สถาบันอุดมศึกษาพิจารณาจัดให้นิสิตนักศึกษาทุกคน สอบวัดความรู้และทักษะภาษาอังกฤษก่อนสำเร็จการศึกษา ตามแบบทดสอบที่สถาบันอุดมศึกษาพัฒนาขึ้นหรือแบบทดสอบตามมาตรฐานสากลอื่น ๆ โดยเทียบเคียงผลกับ Common European Framework of Reference for Languages (CEFR) เพื่อให้ทราบถึงผลสัมฤทธิ์ ด้านทักษะภาษาอังกฤษของนิสิตนักศึกษาแต่ละคน และสถาบันอุดมศึกษาอาจพิจารณานำผลการทดสอบความรู้และทักษะภาษาอังกฤษบันทึกในใบรับรองผลการศึกษา หรือจัดทำเป็นประกาศนียบัตร ทั้งนี้ เป้าหมายผลสัมฤทธิ์ด้านทักษะภาษาอังกฤษ ระดับปริญญาตรี ควรกำหนดเป้าหมายผลสัมฤทธิ์ด้านทักษะภาษาอังกฤษก่อนสำเร็จการศึกษา เทียบเคียงผลกับ CEFR ในระดับเกณฑ์ตั้งแต่ B2 ขึ้นไป

อย่างไรก็ดี สหภาพยุโรปได้พัฒนารอบมาตรฐาน CEFR ขึ้นมาเพื่อนำไปใช้จัดลำดับความสามารถในการใช้ภาษาทั้งหลายที่ผู้คนใช้กันในสหภาพยุโรป มิได้เจาะจงเพียงการจัดลำดับความสามารถทางภาษาอังกฤษเท่านั้น ยิ่งไปกว่านั้น สหภาพยุโรปยังมิได้พัฒนาแบบทดสอบและเกณฑ์การให้คะแนนเพื่อวัดความสามารถทางภาษาอังกฤษของคนแต่ละคน ดังนั้นสถาบันการศึกษาหลายแห่งในประเทศไทยจึงได้ดำเนินการพัฒนาแบบทดสอบความสามารถทางภาษาอังกฤษตามกรอบมาตรฐาน CEFR ขึ้นมา

หลักสูตรครุศาสตรบัณฑิต (สาขาวิชาภาษาอังกฤษ) และหลักสูตรศิลปศาสตรบัณฑิต (สาขาวิชาภาษาอังกฤษ และสาขาวิชาภาษาอังกฤษธุรกิจ) มหาวิทยาลัยราชภัฏเชียงใหม่ ได้เล็งเห็นถึงความสำคัญในการสร้างแบบทดสอบความสามารถทางภาษาอังกฤษตามกรอบมาตรฐาน CEFR เพื่อใช้วัดความสามารถทางภาษาของนักศึกษาในหลักสูตร (ต่อไปนี้จะเรียกว่า แบบทดสอบภาษาอังกฤษของหลักสูตร) ข้อมูลที่ได้จากแบบทดสอบจะช่วยในการปรับปรุงหลักสูตรและกำหนดแนวทาง การเรียนการสอนให้สอดคล้องกับความต้องการ

ต้องการของผู้เรียน เพื่อพัฒนาทักษะทางภาษาให้เป็นตามเกณฑ์มาตรฐาน CEFR แบบทดสอบที่พัฒนาขึ้นเป็นแบบทดสอบแบบปรนัยสี่ตัวเลือก จำนวน 100 ข้อ คะแนนเต็ม 100 คะแนน โดยแบ่งเป็น 4 ส่วน คือ ข้อสอบการฟัง ข้อสอบบทสนทนา (เติมบทสนทนาให้สมบูรณ์จากประโยคตัวเลือกที่ให้มา) ข้อสอบการอ่าน และข้อสอบการเขียน (เป็นข้อสอบ error identification และข้อสอบ cloze test) แบบทดสอบชุดนี้มุ่งวัดความสามารถและความรู้ทางภาษาตามกรอบมาตรฐาน CEFR ในระดับ A2 B1 และ B2 คุณสมบัติเฉพาะของแบบทดสอบนำมาจากคุณสมบัติเฉพาะที่ศูนย์ภาษา มหาวิทยาลัยราชภัฏเชียงใหม่ กำหนดขึ้นมาใหม่ใน พ.ศ. 2566 เพื่อใช้เป็นกรอบในการปรับปรุงแบบทดสอบความสามารถทางภาษาอังกฤษของมหาวิทยาลัยราชภัฏเชียงใหม่ (CMRU-TEP) แบบทดสอบภาษาอังกฤษของหลักสูตรได้ผ่านการประเมินคุณภาพโดยมีการตรวจสอบค่าความเที่ยงตรงตามเนื้อหา (content validity) ของข้อสอบโดยการพิจารณาของผู้เชี่ยวชาญ มีการทดลองแบบทดสอบกับผู้เข้าสอบเพื่อหาค่าความเชื่อมั่น (KR-20) ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ค่าความยากง่าย และค่าอำนาจจำแนก

ขั้นตอนต่อไปในการพัฒนาแบบทดสอบความสามารถทางภาษาอังกฤษตามกรอบมาตรฐาน CEFR ก่อนนำไปใช้จริงก็คือการกำหนดมาตรฐาน (standard setting) และการหาคะแนนจุดตัด (cut-off score) ให้กับแบบทดสอบ ไชเชก (Cizek, 2012) อธิบายว่าการกำหนดมาตรฐาน คือ การดำเนินการตามกระบวนการหรือระบบที่กำหนดไว้อย่างมีเหตุผล เพื่อกำหนดตัวเลขที่จะใช้ในการแยกแยะระดับหรือผลการทดสอบ หรือกล่าวอีกทางหนึ่งได้ว่าการกำหนดมาตรฐานเป็นการให้ความหมายกับคะแนนตัวเลขจากการทำแบบทดสอบซึ่งช่วยในการแบ่งผู้เข้าสอบออกเป็นกลุ่มต่างๆ สำหรับงานวิจัยนี้ คณะผู้วิจัยได้ดำเนินการกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดสามจุดตัดในการแบ่งผู้เข้าสอบออกเป็นผู้ที่มีความสามารถตาม กรอบมาตรฐาน CEFR ในระดับ A2 B1 และ B2 หรือสูงกว่า

วิธีการกำหนดมาตรฐานทำได้หลายวิธี เช่น วิธี Angoff ซึ่งเป็นการหาคะแนนจุดตัดโดยการประเมินร้อยละของผู้เข้าสอบที่จะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้อง วิธี Bookmark ซึ่งหาคะแนนจุดตัดโดยการเรียงระดับข้อสอบจากข้อง่ายไปข้อยาก วิธี Ebel ซึ่งหาคะแนนจุดตัดโดยการประเมินร้อยละของผู้เข้าสอบที่มีโอกาสสอบผ่าน และวิธี Nedelsky ซึ่งกำหนดคะแนนจุดตัดจากจำนวนตัวลวงที่ผู้เข้าสอบอาจเลือกตอบ (Tiyawongsuwan et al., 2013) วิธีการกำหนดมาตรฐานวิธีหนึ่งที่สถาบันการศึกษาในประเทศไทยได้นำมาใช้ คือ วิธีที่เรียกกันว่า Yes/No Angoff ซึ่งนำเสนอโดย อิมพารา และเพลก (Impara & Plake, 1997) จิรดา วุฒยมกร (Wudthayagorn, 2018) ใช้วิธีกำหนดมาตรฐานแบบ Yes/No Angoff เพื่อหาคะแนนจุดตัดของแบบทดสอบ CU-TEP ตามมาตรฐาน CEFR ในระดับ A2 B1 B2 และ C1 นอกจากนี้ ชื่นจิตต์ อธิวรกุล และ จิรดา วุฒยมกร (Athiworakun & Wudthayagorn, 2018) ได้ใช้วิธีดังกล่าวกำหนดคะแนนจุดตัดของผู้ที่สอบแบบทดสอบ SWU-SET ตามกรอบมาตรฐาน CEFR ในระดับ A2 B1 และ B2 และปัทมรัศม์ นาคนิษฐนนต์ (Nakanitanon, 2021) ก็ได้ใช้วิธี Yes/No Angoff หาคะแนนจุดตัดของแบบทดสอบ CMRU-TEP ตามกรอบมาตรฐาน CEFR ในระดับ A1 A2 B1 B2 และ C1 นอกจากนั้นยังมีผู้เชี่ยวชาญใช้วิธีนี้ในการเทียบมาตรฐาน CEFR กับข้อสอบมาตรฐาน เช่น ทาเนนเบอัม และแบรอน (Tannenbaum & Baron, 2015b) ทำการเปรียบเทียบคะแนนการสอบ TOEIC กับเกณฑ์มาตรฐานภาษาอังกฤษของประเทศเวียดนามโดยวิธี

Yes/No Angoff หลังจากนั้นนักวิจัยทั้งสองยังได้ใช้วิธีการเดิมในการเปรียบเทียบคะแนนจากแบบทดสอบ TOEFL Junior Comprehensive Test ในส่วนของข้อสอบการฟังและข้อสอบการอ่านกับกรอบมาตรฐาน CEFR (Tannenbaum & Baron, 2015a)

ความนิยมในการใช้วิธีกำหนดมาตรฐานแบบ Yes/No Angoff มาจากจุดแข็งของวิธีการนี้ที่ง่ายต่อการประเมินเมื่อเทียบกับวิธีการอื่น (Tiyawongsuwan et al., 2013; Cizek & Bunch, 2007) และเหมาะสมกับการกำหนดคะแนนจุดตัดหลายจุด เนื่องจากหากผู้ประเมินประเมินว่าผู้เข้าสอบที่มีความสามารถต่ำสุดในระดับ A2 จะสามารถตอบข้อสอบข้อใดได้ถูกต้อง ก็หมายความว่าผู้เข้าสอบที่มีความสามารถต่ำสุดในระดับ B1 และสูงกว่านั้นก็ต้องตอบข้อสอบข้อนั้นได้ถูกต้องเช่นกัน โดยที่ผู้ประเมินไม่ต้องพิจารณาไปที่ระดับวิธีกำหนดมาตรฐานแบบ Yes/No Angoff จึงมีประสิทธิภาพในแง่ของ ความคุ้มค่าทางเวลา (Tannenbaum & Baron, 2015b) อย่างไรก็ตาม สุปัทม์ สุกมลสันต์ (2564) ได้พบทวนงานวิจัยในประเทศไทยที่รายงานผลการกำหนดมาตรฐานให้กับแบบทดสอบความสามารถทางภาษาอังกฤษแล้วพบว่า งานวิจัยที่ผ่านมามีไม่รายงานค่าความเที่ยงตรงของการประเมิน ทั้งที่ค่าความเที่ยงตรงของการประเมินจะเป็นตัวบ่งชี้ที่สำคัญว่าคะแนนจุดตัดที่ได้สามารถแบ่งกลุ่มผู้เข้าสอบได้อย่างเที่ยงตรง

ดังนั้น คณะผู้วิจัยจึงทำการศึกษาระบวนการกำหนดมาตรฐานให้กับแบบทดสอบภาษาอังกฤษของหลักสูตร และวิเคราะห์ความถูกต้องของการกำหนดมาตรฐาน (validation) ทั้งความเชื่อมั่นและความเที่ยงตรงของการประเมิน เพื่อให้ได้คะแนนจุดตัดที่สามารถนำไปใช้จัดกลุ่มความสามารถทางภาษาอังกฤษตามกรอบมาตรฐาน CEFR ของผู้เข้าสอบได้อย่างเที่ยงตรง ซึ่งจะเป็นประโยชน์ต่อผู้เข้าสอบและอาจารย์ผู้สอนในการเลือกวิธีการพัฒนาความสามารถทางภาษาอังกฤษที่ตรงกับความต้องการของผู้เข้าสอบต่อไป อีกทั้งงานวิจัยนี้ยังเสนอแนวทางการตรวจสอบความถูกต้องของการกำหนดมาตรฐานที่งานวิจัยในอนาคตสามารถนำไปประยุกต์ใช้ได้

วัตถุประสงค์การวิจัย

1. เพื่อกำหนดคะแนนจุดตัดตามกรอบมาตรฐาน CEFR ระดับ A2 B1 และ B2 โดยวิธี Yes/No Angoff ให้กับแบบทดสอบความสามารถทางภาษาอังกฤษที่ได้รับการพัฒนาโดยหลักสูตรครุศาสตรบัณฑิต (สาขาวิชาภาษาอังกฤษ) และหลักสูตรศิลปศาสตรบัณฑิต (สาขาวิชาภาษาอังกฤษ และสาขาวิชาภาษาอังกฤษธุรกิจ) มหาวิทยาลัยราชภัฏเชียงใหม่

2. เพื่อวิเคราะห์ความถูกต้องของการกำหนดมาตรฐานและคะแนนจุดตัดที่กำหนด

กรอบแนวคิด



ภาพที่ 1 กรอบแนวคิดการวิจัย

วิธีการวิจัย

งานวิจัยนี้เป็นการวิจัยเชิงปริมาณเพื่อกำหนดคะแนนจุดตัดจำนวนสามจุดให้กับแบบทดสอบภาษาอังกฤษของหลักสูตรตามกรอบมาตรฐาน CEFR ระดับ A2 B1 และ B2 และเพื่อวิเคราะห์ความถูกต้องของกระบวนการกำหนดมาตรฐานและคะแนนจุดตัดที่ได้

เครื่องมือวิจัย

เครื่องมือวิจัยที่ใช้ในการเก็บข้อมูล ได้แก่

1. แบบทดสอบภาษาอังกฤษของหลักสูตร ซึ่งเป็นข้อสอบปรนัยแบบสี่ตัวเลือก จำนวน 100 ข้อ แบบทดสอบออกแบบมาเพื่อวัดทักษะการฟัง ทักษะการอ่าน ความรู้เกี่ยวกับภาษาที่ใช้ในการสนทนา และความรู้เกี่ยวกับภาษา คำศัพท์ และไวยากรณ์ที่ใช้ในการเขียนของผู้เข้าสอบ ข้อสอบได้รับการออกแบบมาให้สอดคล้องกับคำอธิบายความสามารถทางภาษาตามกรอบมาตรฐาน CEFR ในระดับ A2 B1 และ B2 ข้อสอบทุกข้อมีค่าความเที่ยงตรงตามเนื้อหา (content validity) ตามสูตร index of item-objective congruence for multidimensional items ของเทอร์เนอร์ และคาร์ลสัน (Turner & Carlson, 2003) มากกว่า .75 แบบทดสอบมีค่าความเชื่อมั่น KR-20 เท่ากับ .93 ค่าเฉลี่ยเท่ากับ 57 ค่าเบี่ยงเบนมาตรฐานเท่ากับ 16.64 ค่าความยากง่ายเฉลี่ยทั้งฉบับเท่ากับ .57 และค่าอำนาจจำแนกเฉลี่ยทั้งฉบับเท่ากับ .46 จากการนำแบบทดสอบไปทดลองกับนักศึกษาสาขาวิชาภาษาอังกฤษ มหาวิทยาลัยราชภัฏเชียงใหม่ จำนวน 100 คน

2. แบบทดสอบความเข้าใจในเกณฑ์มาตรฐาน CEFR ซึ่งกำหนดให้ผู้ประเมินทุกคนทำหลังจากกิจกรรมการอบรมก่อนการกำหนดมาตรฐาน แบบทดสอบเป็นแบบปรนัยสี่ตัวเลือกจำนวน 30 ข้อ โจทย์คำถามเป็นข้อความจากคำอธิบายความสามารถตามกรอบมาตรฐาน CEFR (performance descriptors) คำศัพท์ บทอ่าน และข้อสอบในแบบทดสอบ ผู้ประเมินต้องตอบว่าข้อความที่ยกมาเป็นคำอธิบายความสามารถเป็นคำศัพท์ เป็นบทอ่าน หรือเป็นข้อสอบที่วัดความสามารถทางภาษาอังกฤษตามเกณฑ์มาตรฐาน CEFR ในระดับใด การกำหนดเนื้อหาแบบทดสอบความเข้าใจในเกณฑ์มาตรฐาน CEFR ตามที่กล่าวมาเป็นไปตามข้อเสนอแนะในคู่มือ Relating Language Examinations to the Common European Framework of

Reference for Languages (Council of Europe, 2009) ซึ่งแนะนำให้ทำการทดสอบผู้ประเมินก่อนการกำหนดมาตรฐาน เพื่อตรวจสอบความเชี่ยวชาญของผู้ประเมินเกี่ยวกับกรอบมาตรฐาน CEFR ผู้เชี่ยวชาญจำนวน 3 คน จากหลากหลายสถาบันได้ตรวจสอบความเหมาะสมในภาพรวมแบบทดสอบนี้ และแนะนำให้ปรับภาษาในแบบทดสอบบางส่วนเพื่อความชัดเจน คณะผู้วิจัยจึงแก้ไขแบบทดสอบตามคำแนะนำก่อนการนำไปใช้

3. แบบสำรวจความคิดเห็นของผู้ประเมินต่อการอบรมก่อนการกำหนดมาตรฐาน ซึ่งอยู่ในลักษณะของมาตราส่วนประมาณค่าแบบลิเคิร์ตสเกลแบบสี่ระดับ (four-point Likert scale) โดย 1 หมายถึง ไม่เห็นด้วยอย่างมาก 2 หมายถึง ไม่เห็นด้วย 3 หมายถึง เห็นด้วย และ 4 หมายถึง เห็นด้วยอย่างมาก แบบสำรวจนี้เป็นแบบสำรวจที่มหาวิทยาลัยมิชิแกน (University of Michigan, 2014) ใช้ในการประเมินกิจกรรมก่อนการกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดให้กับแบบทดสอบความสามารถทางภาษาอังกฤษที่มหาวิทยาลัยมิชิแกนพัฒนาขึ้นมา คณะผู้วิจัยได้ทำการแปลและให้ผู้เชี่ยวชาญจำนวน 3 คนจากหลากหลายสถาบันตรวจสอบความเหมาะสมในภาพรวมของแบบสำรวจและการแปล ผู้เชี่ยวชาญแนะนำให้ปรับคำแปลบางส่วนเพื่อให้ความหมายตรงกับต้นฉบับภาษาอังกฤษมากยิ่งขึ้น คณะผู้วิจัยได้แก้ไขภาษาตามคำแนะนำแล้วก่อนการนำไปใช้

4. แบบประเมินข้อสอบโดยวิธี Yes/No Angoff ซึ่งผู้ประเมินใช้ในการหาคะแนนจุดตัด โดยการกรอก 1 ลงไปในแบบประเมิน หากประเมินว่าผู้เข้าสอบที่มีความสามารถต่ำสุดในระดับมาตรฐาน CEFR ระดับใดระดับหนึ่งจะตอบข้อสอบข้อใดได้ถูกต้อง (Yes) และกรอก 0 ลงในแบบประเมิน หากประเมินว่าผู้เข้าสอบจะตอบไม่ถูก (No) ผู้เชี่ยวชาญจำนวน 3 คนจากหลากหลายสถาบันได้ตรวจสอบความเหมาะสมในภาพรวมของแบบประเมินนี้ และให้คำแนะนำว่าควรเขียนอธิบายวิธีการประเมินแบบ Yes/No Angoff ในแบบประเมินด้วย เพื่อให้ผู้ประเมินสามารถตรวจสอบได้ระหว่างการประเมิน คณะผู้วิจัยจึงเพิ่มข้อมูลดังกล่าวก่อนการนำไปใช้

5. แบบสำรวจความคิดเห็นของผู้ประเมินต่อกิจกรรมการกำหนดมาตรฐาน ซึ่งอยู่ในลักษณะของมาตราส่วนประมาณค่าแบบลิเคิร์ตสเกลแบบสี่ระดับ (four-point Likert scale) โดย 1 หมายถึง ไม่เห็นด้วยอย่างมาก 2 หมายถึง ไม่เห็นด้วย 3 หมายถึง เห็นด้วย และ 4 หมายถึง เห็นด้วยอย่างมาก แบบสำรวจนี้เป็นแบบสำรวจซึ่งมหาวิทยาลัยมิชิแกน (University of Michigan, 2014) ใช้ในการสำรวจความคิดเห็นของผู้ประเมินต่อการกำหนดมาตรฐานให้กับแบบทดสอบความสามารถทางภาษาอังกฤษของมหาวิทยาลัย คณะผู้วิจัยได้แปลแบบสำรวจนี้และให้ผู้เชี่ยวชาญจำนวน 3 คนจากหลากหลายสถาบันตรวจสอบความเหมาะสมในภาพรวมของแบบสำรวจ จากนั้นคณะผู้วิจัยจึงปรับแก้ภาษาตามคำแนะนำของผู้เชี่ยวชาญก่อนการนำไปใช้

การเก็บรวบรวมข้อมูล

หลังจากการนำแบบทดสอบภาษาอังกฤษของหลักสูตรไปทดลองใช้และวิเคราะห์คุณภาพข้อสอบแล้ว คณะผู้วิจัยได้ดำเนินการคัดเลือกผู้ที่ทำการประเมินในกิจกรรมการกำหนดเกณฑ์มาตรฐานมาจำนวน 10 คน ซึ่งเป็นจำนวนที่โฟเวลล์ ฟิวเทรล และแมกกลอกลิน (Fowell, Fewtrell & McLaughlin, 2008) เสนอว่าพอเพียงต่อการสร้างความเชื่อมั่นในการประเมิน ผู้ประเมินทุกคนมีประสบการณ์การสอนภาษาอังกฤษใน

ระดับอุดมศึกษามากกว่า 5 ปี และผู้ประเมินจำนวนสามคนจากทั้งหมดมีประสบการณ์การสอนภาษาอังกฤษในระดับมัธยมศึกษา ผู้ประเมินทุกคนเคยพัฒนาข้อสอบวัดความสามารถทางภาษาอังกฤษตามกรอบมาตรฐาน CEFR และเคยเข้าร่วมการอบรมเกี่ยวกับกรอบมาตรฐาน CEFR

ขั้นตอนการกำหนดมาตรฐานที่งานวิจัยนี้ใช้เป็นไปตามข้อเสนอของเพลก และไซเซก (Plake & Cizek, 2012) ซึ่งกำหนดขั้นตอนการกำหนดมาตรฐานแบบ Angoff ว่ามีดังต่อไปนี้

1. การอบรมผู้ประเมิน เพื่อซักซ้อมความเข้าใจเรื่องวัตถุประสงค์ วิธีการกำหนดมาตรฐาน และทดลองประเมินข้อสอบ

2. การประเมินรอบที่หนึ่ง

3. การให้ข้อมูลสะท้อนกลับกับผู้ประเมิน ซึ่งแบ่งได้เป็นสองประเภท ข้อมูลประเภทแรกเรียกว่าข้อมูลสะท้อนกลับตามเกณฑ์ปกติ (normative feedback) เช่น คะแนนจุดตัดของผู้ประเมินแต่ละคน คะแนนจุดตัดของผู้ประเมินทั้งกลุ่ม ค่าเฉลี่ย ค่ามัธยฐาน ค่าเบี่ยงเบนมาตรฐาน คะแนนจุดตัดสูงสุด และคะแนนจุดตัดต่ำสุด ข้อมูลประเภทที่สองคือข้อมูลสะท้อนกลับตามความจริง (reality feedback) เช่น สัดส่วนของผู้เข้าสอบที่ตอบข้อสอบถูกต้อง (p-values) การให้ข้อมูลสะท้อนกลับมีวัตถุประสงค์เพื่อให้ผู้ประเมินมีโอกาสเปรียบเทียบการประเมินของตนกับการประเมินของผู้ประเมินคนอื่นและผลการสอบ

4. การอภิปรายข้อมูลสะท้อนกลับก่อนการประเมินรอบที่สอง การอภิปรายนี้มีวัตถุประสงค์เพื่อให้ผู้ประเมินได้แลกเปลี่ยนความเห็นเกี่ยวกับการประเมินของตน ซึ่งจะทำให้การประเมินเป็นไปในทิศทางเดียวกันมากยิ่งขึ้น

5. การประเมินรอบที่สอง

6. การให้ข้อมูลสะท้อนกลับตามเกณฑ์ปกติ (normative feedback) และข้อมูลสะท้อนกลับเชิงผลกระทบ (impact feedback) จากการประเมินรอบที่สองกับผู้ประเมิน

7. การอภิปรายข้อมูลสะท้อนกลับก่อนการประเมินรอบที่สาม

8. การประเมินรอบที่สาม

9. การคำนวณคะแนนจุดตัด

10. การประเมินกระบวนการกำหนดมาตรฐานที่ได้ดำเนินการมา

กิจกรรมกำหนดมาตรฐานของงานวิจัยนี้ใช้เวลาทั้งหมดสามวัน สำหรับวันแรก ผู้ประเมินทั้ง 10 คน ต้องทำแบบทดสอบภาษาอังกฤษของหลักสูตร เพื่อให้ทราบระดับความยากง่ายของข้อสอบที่ตนจะประเมิน และเข้าร่วมการอบรมและอภิปราย เพื่อสร้างความเข้าใจเกี่ยวกับวัตถุประสงค์ของการกำหนดมาตรฐาน วิธีการกำหนดมาตรฐานแบบ Yes/No Angoff ลักษณะของผู้เข้าสอบที่มีความสามารถต่ำที่สุดตามกรอบมาตรฐาน CEFR แต่ละระดับ และกรอบมาตรฐาน CEFR ที่ใช้ในการประเมินข้อสอบ นอกจากนั้น เพื่อให้มั่นใจว่าผู้ประเมินเข้าใจเกณฑ์มาตรฐาน CEFR และวิธีการกำหนดมาตรฐาน คณะผู้วิจัยได้ให้ผู้ประเมินทุกคนทำแบบทดสอบความเข้าใจในเกณฑ์มาตรฐาน CEFR และตอบแบบสำรวจความคิดเห็นของผู้ประเมินต่อการอบรมก่อนการกำหนดมาตรฐาน คณะผู้วิจัยใช้เวลาประมาณ 6 ชั่วโมงในการจัดกิจกรรมวันแรก

ในวันที่สอง ผู้ประเมินทำการประเมินแบบทดสอบภาษาอังกฤษของหลักสูตรเพื่อหาคะแนนจุดตัดรอบที่หนึ่งและรอบที่สอง และในวันสุดท้าย ผู้ประเมินทำการประเมินรอบที่สาม จากนั้นจึงตอบแบบสำรวจความคิดเห็นของผู้ประเมินต่อกิจกรรมการกำหนดมาตรฐาน

ในขั้นตอนต่อไป คณะผู้วิจัยได้นำข้อมูลที่ได้จากการจัดกิจกรรมการกำหนดมาตรฐานมาวิเคราะห์หาค่าความถูกต้อง

การวิเคราะห์ข้อมูล

คณะผู้วิจัยใช้ค่าเฉลี่ยของการประเมินในกำหนดคะแนนจุดตัดตามกรอบมาตรฐาน CEFR ให้กับแบบทดสอบภาษาอังกฤษของหลักสูตร และใช้ค่าสหสัมพันธ์ภายในชั้นของการประเมิน ค่าสหสัมพันธ์ของเพียร์สันระหว่างการจัดกลุ่มผู้เข้าสอบตามคะแนนจุดตัดที่ได้กับการจัดกลุ่มผู้เข้าสอบโดยอาจารย์ผู้สอน และค่าเฉลี่ยของแบบสำรวจความคิดเห็นของผู้ประเมินเกี่ยวกับกระบวนการกำหนดมาตรฐาน เพื่อวิเคราะห์ความถูกต้องของการกำหนดมาตรฐานและคะแนนจุดตัดที่กำหนด

ผลการวิจัย

ในวันแรกของการกำหนดเกณฑ์มาตรฐาน คณะผู้วิจัยได้จัดให้ผู้ประเมินทำแบบทดสอบภาษาอังกฤษของหลักสูตร และจัดอบรมวิธีการกำหนดเกณฑ์มาตรฐาน CEFR ของแบบทดสอบโดยวิธี Yes/No Angoff จากนั้นผู้ประเมินทุกคนต้องทำแบบทดสอบความเข้าใจในเกณฑ์มาตรฐาน CEFR จำนวน 30 ข้อ (รายละเอียดตามตารางที่ 1) การอบรมและการทดสอบเป็นส่วนหนึ่งของการเก็บหลักฐานความเที่ยงตรงเชิงกระบวนการ (procedural validity) ที่แสดงให้เห็นว่าผู้ประเมินเข้าใจกระบวนการกำหนดเกณฑ์มาตรฐานที่ตนเข้าร่วม คะแนนเฉลี่ยของผู้ประเมินจากการทำแบบทดสอบความเข้าใจในเกณฑ์มาตรฐาน CEFR คือ 21.6 คะแนนจาก 30 คะแนน ซึ่งหมายความว่า ผู้ประเมินสามารถจัดระดับให้แก่คำอธิบายความสามารถตามเกณฑ์มาตรฐาน CEFR (performance descriptors) คำศัพท์ บทอ่าน และข้อสอบตัวอย่างได้ถูกต้องประมาณร้อยละ 71.92 นอกจากนี้ คณะผู้วิจัยยังได้หาค่าสหสัมพันธ์แบบจัดอันดับ (Spearman's rank correlation) ระหว่างระดับ CEFR ของคำอธิบายความสามารถ คำศัพท์ บทอ่าน และข้อสอบตัวอย่างกับการกำหนดระดับ CEFR ของผู้ประเมิน โดยการแปลงระดับ CEFR ให้เป็นตัวเลข ($A1 = 1$, $A2 = 2$, $B1 = 3$ และ $B2 = 4$) พบว่ามีค่าสหสัมพันธ์โดยเฉลี่ยอยู่ที่ .85 ซึ่งหมายความว่ามีความสัมพันธ์กันสูงมาก (อนูวัติ คุณแก้ว, 2566) แสดงว่าผู้ประเมินมีความรู้ความเข้าใจเกี่ยวกับเกณฑ์มาตรฐาน CEFR นอกจากนี้คณะผู้วิจัยยังได้เปิดโอกาสให้ผู้ประเมินที่ตอบข้อสอบถูกแสดงเหตุผลของตนในการตอบข้อสอบแต่ละข้อ เพื่อเสริมสร้างความเข้าใจให้กับผู้ประเมินที่ตอบผิด

ตารางที่ 1 ผลการตอบแบบทดสอบความเข้าใจในเกณฑ์มาตรฐาน CEFR (ผป = ผู้ประเมิน)

ข้อมูล	ผป1	ผป2	ผป3	ผป4	ผป5	ผป6	ผป7	ผป8	ผป9	ผป10	เฉลี่ย
จำนวน	26	22	21	20	19	24	22	26	17	19	21.6
ข้อสอบที่ ตอบถูก											
Correlation	.95	.89	.88	.67	.81	.93	.85	.95	.84	.81	.85
Mean	2.73	2.73	2.47	2.50	2.60	2.73	2.70	2.67	2.50	2.60	2.62

หลังจากนั้น คณะผู้วิจัยได้นำผลการทำแบบทดสอบความเข้าใจในเกณฑ์มาตรฐาน CEFR มาวิเคราะห์เพื่อตรวจสอบความสอดคล้อง (consistency) ของการประเมินภายในกลุ่มผู้ประเมิน ค่าสหสัมพันธ์ภายในชั้น (ICC) ของการตอบแบบทดสอบอยู่ที่ .97 แสดงว่าความเชื่อมั่นของผู้ประเมินมีความสอดคล้องกันในระดับดีมาก (Koo & Li, 2016) ค่าสัมประสิทธิ์คอนคอร์ดานส์ของเคนดัลล์ (Kendall's W) ของการตอบแบบทดสอบ คือ 0.49 แสดงว่าคำตอบของผู้ประเมินมีความสอดคล้องปานกลาง (Landis & Koch, 1977) ค่าสัมประสิทธิ์ครอนบาค (Cronbach's alpha) ของคำตอบ คือ .97 แสดงว่า คำตอบของผู้ประเมินทั้งหมดมีความสอดคล้องภายในดีมาก (Gliem & Gliem, 2003)

นอกจากนั้น ผู้ประเมินยังได้ตอบแบบสำรวจความคิดเห็นของผู้ประเมินต่อการอบรมก่อนการกำหนดมาตรฐาน เพื่อตรวจสอบความเที่ยงตรงเชิงกระบวนการ (ตารางที่ 2) ผลการสำรวจความคิดเห็นของผู้ประเมินต่อการอบรมมีคะแนนเฉลี่ยรวมอยู่ที่ 3.91 ซึ่งหมายถึง “เห็นด้วยอย่างยิ่ง” (Todd, 2011) กับกระบวนการในการจัดกิจกรรมเตรียมความพร้อม ข้อที่ได้รับคะแนนสูงสุดมีด้วยกันทั้งหมด 4 ข้อ ได้แก่ ข้อ 1) เอกสารก่อนการกำหนดมาตรฐานช่วยให้ฉันเข้าใจกรอบมาตรฐาน CEFR ข้อ 4) การพูดคุยเกี่ยวกับเอกสารก่อนการกำหนดมาตรฐานช่วยตอบข้อสงสัยของฉัน ข้อ 5) กิจกรรมเพื่อสร้างความเข้าใจเกี่ยวกับกรอบมาตรฐาน CEFR ช่วย使我เข้าใจกรอบมาตรฐาน CEFR และ ข้อ 8) ฉันมีเวลาเพียงพอในการเข้าร่วมพูดคุยแสดงความคิดเห็นในระหว่างกิจกรรมก่อนการกำหนดมาตรฐาน ซึ่งมีคะแนนเฉลี่ย 4.00 ทุกข้อ คะแนนที่ได้แสดงให้เห็นว่าเอกสารและกิจกรรมที่ให้ข้อมูลมีความชัดเจน ช่วยเสริมสร้างความเข้าใจของผู้ประเมินเกี่ยวกับกรอบมาตรฐาน CEFR อีกทั้งยังมีเวลาให้กับผู้ประเมินอย่างเพียงพอในการแสดงความคิดเห็น ทำให้ผู้ประเมินรู้สึกได้รับการสนับสนุนและเข้าใจ กรอบมาตรฐาน CEFR มากขึ้น ส่วนข้อที่ได้รับคะแนนต่ำที่สุดได้แก่ ข้อ 2) ตัวอย่างจากแบบทดสอบภาษาอังกฤษของหลักสูตรช่วยให้ฉันเข้าใจโครงสร้างและระดับของแบบทดสอบ ซึ่งมีคะแนนเฉลี่ย 3.70 ผู้ประเมินบางคนอาจเห็นว่าตัวอย่างที่ยกมามีจำนวนไม่มากพอ อย่างไรก็ตาม คะแนนเฉลี่ยที่ได้ก็ยังคงอยู่ในเกณฑ์ “เห็นด้วยอย่างยิ่ง”

ตารางที่ 2 ผลการตอบแบบสำรวจความคิดเห็นของผู้ประเมินต่อการอบรมก่อนการกำหนดมาตรฐาน

คำถาม	คะแนนเฉลี่ย	SD	ความหมาย
1. เอกสารก่อนการกำหนดมาตรฐานช่วยให้ฉันเข้าใจกรอบมาตรฐาน CEFR	4.00	.0	เห็นด้วยอย่างยิ่ง
2. ตัวอย่างจากแบบทดสอบภาษาอังกฤษของหลักสูตรช่วยให้ฉันเข้าใจโครงสร้างและระดับของแบบทดสอบ	3.70	.48	เห็นด้วยอย่างยิ่ง
3. การนำเสนอกิจกรรมการกำหนดมาตรฐานช่วยให้ฉันเข้าใจกระบวนการกำหนดมาตรฐาน	3.90	.31	เห็นด้วยอย่างยิ่ง
4. การพูดคุยเกี่ยวกับเอกสารก่อนการกำหนดมาตรฐานช่วยตอบข้อสงสัยของฉัน	4.00	.0	เห็นด้วยอย่างยิ่ง
5. กิจกรรมเพื่อสร้างความเข้าใจเกี่ยวกับกรอบมาตรฐาน CEFR ช่วยให้ฉันเข้าใจกรอบมาตรฐาน CEFR	4.00	.0	เห็นด้วยอย่างยิ่ง
6. ข้อสอบที่ใช้ในการฝึกประเมินช่วยให้ฉันเข้าใจกระบวนการประเมิน	3.80	.42	เห็นด้วยอย่างยิ่ง
7. ฉันมีเวลาเพียงพอในการทำแบบฝึกหัดและกิจกรรมก่อนการกำหนดมาตรฐาน	3.90	.31	เห็นด้วยอย่างยิ่ง
8. ฉันมีเวลาเพียงพอในการเข้าร่วมพูดคุยแสดงความคิดเห็นในระหว่างกิจกรรมก่อนการกำหนดมาตรฐาน	4.00	.0	เห็นด้วยอย่างยิ่ง
รวม	3.91	0.19	เห็นด้วยอย่างยิ่ง

ในวันที่สองของกิจกรรมการกำหนดมาตรฐาน ผู้ประเมินได้เริ่มประเมินข้อสอบจำนวน 100 ข้อ คะแนนเต็ม 100 คะแนน ในการประเมินรอบที่หนึ่ง คะแนนจุดตัดในระดับ A2 อยู่ที่ 23 คะแนน (SD 7.27) คะแนนจุดตัดในระดับ B1 อยู่ที่ 74 คะแนน (SD 11.82) และคะแนนจุดตัดในระดับ B2 อยู่ที่ 91 คะแนน (SD 10.72) (รายละเอียดตามตารางที่ 3)

ตารางที่ 3 ผลการกำหนดมาตรฐานรอบที่หนึ่ง

ลำดับผู้ประเมิน	A2	B1	B2
ผู้ประเมินคนที่ 1	30	66	88
ผู้ประเมินคนที่ 2	21	74	90
ผู้ประเมินคนที่ 3	34	83	97
ผู้ประเมินคนที่ 4	27	74	92
ผู้ประเมินคนที่ 5	24	69	91
ผู้ประเมินคนที่ 6	23	77	93
ผู้ประเมินคนที่ 7	19	64	75
ผู้ประเมินคนที่ 8	7	38	61
ผู้ประเมินคนที่ 9	15	76	97
ผู้ประเมินคนที่ 10	25	76	94
Mean	23	70	88
SD	7.27	11.82	10.72
Min	7	38	61
Max	34	83	97

จากนั้นคณะผู้วิจัยได้สะท้อนข้อมูลจากการประเมินรอบแรกให้แก่ผู้ประเมินทราบ รวมถึงแจ้งค่าสัดส่วนของผู้เข้าสอบที่ตอบข้อสอบแต่ละข้อได้ถูกต้อง (p-values) ในขั้นตอนนี้ คณะผู้วิจัยได้เปิดโอกาสให้ผู้ประเมินอภิปรายแสดงเหตุผลของการประเมินของตน

ในการประเมินรอบที่สอง (รายละเอียดตามตารางที่ 4) คะแนนจุดตัดในระดับ A2 อยู่ที่ 25 คะแนน คะแนนจุดตัดในระดับ B1 อยู่ที่ 63 คะแนน และคะแนนจุดตัดในระดับ B2 อยู่ที่ 86 คะแนน ค่า SD ของการประเมินในระดับ A2 ลดลงมาอยู่ที่ 6.71 และค่า SD ในระดับ B1 ลดลงมาอยู่ที่ 10.85 แสดงให้เห็นว่าข้อมูลสารสนเทศและการอภิปรายทำให้ผู้ประเมินมีมุมมองที่สอดคล้องกันมากขึ้น อย่างไรก็ตาม ค่า SD ของการประเมินคะแนนจุดตัดระดับ B2 เพิ่มขึ้นเล็กน้อยจาก 10.72 เป็น 10.85

ตารางที่ 4 ผลการกำหนดมาตรฐานรอบที่สอง

ลำดับผู้ประเมิน	A2	B1	B2
ผู้ประเมินคนที่ 1	31	63	90
ผู้ประเมินคนที่ 2	17	71	93
ผู้ประเมินคนที่ 3	27	70	93
ผู้ประเมินคนที่ 4	15	35	56
ผู้ประเมินคนที่ 5	23	65	85
ผู้ประเมินคนที่ 6	21	56	89
ผู้ประเมินคนที่ 7	32	62	83
ผู้ประเมินคนที่ 8	37	62	83
ผู้ประเมินคนที่ 9	20	72	94
ผู้ประเมินคนที่ 10	28	75	94
Mean	25	63	86
SD	6.71	10.85	10.82
Min	15	35	56
Max	37	75	94

ในวันที่สามของการกำหนดมาตรฐาน ผู้ประเมินได้ร่วมกันอภิปรายผลการประเมินรอบที่สอง และคณะผู้วิจัยได้แสดงให้เห็นผลกระทบหากนำคะแนนจุดตัดในการประเมินรอบที่สองไปใช้ในการจัดระดับความสามารถทางภาษาอังกฤษตามกรอบมาตรฐาน CEFR ของนักศึกษาที่ทดลองทำแบบทดสอบ หลังจากนั้นจึงทำการประเมินรอบที่สาม

ในการประเมินรอบสาม (รายละเอียดตามตารางที่ 5) คะแนนจุดตัดในระดับ A2 อยู่ที่ 24 คะแนน คะแนนจุดตัดในระดับ B1 อยู่ที่ 64 คะแนน และคะแนนจุดตัดในระดับ B2 อยู่ที่ 88 คะแนน ค่า SD ของการประเมินจุดตัดสำหรับระดับ A2 เพิ่มขึ้นเล็กน้อยจาก 6.71 เป็น 7.79 ส่วนค่า SD ในระดับ B1 ลดลงมาอยู่ที่ 10.08 และค่า SD ในระดับ B2 ลดลงมาอยู่ที่ 8.37

ตารางที่ 5 ผลการกำหนดมาตรฐานรอบที่สาม

ลำดับผู้ประเมิน	A2	B1	B2
ผู้ประเมินคนที่ 1	30	62	90
ผู้ประเมินคนที่ 2	9	77	95
ผู้ประเมินคนที่ 3	24	74	94
ผู้ประเมินคนที่ 4	15	43	66
ผู้ประเมินคนที่ 5	24	65	88
ผู้ประเมินคนที่ 6	21	55	90
ผู้ประเมินคนที่ 7	32	59	84
ผู้ประเมินคนที่ 8	34	62	86
ผู้ประเมินคนที่ 9	20	69	88
ผู้ประเมินคนที่ 10	28	72	94
Mean	24	64	88
SD	7.79	10.08	8.37
Min	2.46	3.19	2.65
Max	9	43	66

หลังจากนั้น คณะผู้วิจัยได้นำผลการกำหนดเกณฑ์มาตรฐานมาตรวจสอบความถูกต้องของการกำหนดมาตรฐานทั้งภายนอกและภายใน (external validation and internal validation) ตามข้อเสนอแนะในคู่มือ Relating Language Examinations to the Common European Framework of Reference for Languages (Council of Europe, 2009) คณะผู้วิจัยตรวจสอบค่าความถูกต้องภายในโดยการวิเคราะห์ค่าความสอดคล้องของการประเมิน จากการวิเคราะห์ค่าสหสัมพันธ์ภายในชั้น (ICC) ของการประเมินคะแนนจุดตัดระดับ A2 คือ .92 ระดับ B1 คือ .92 และระดับ B2 คือ .81 แสดงว่าการประเมินคะแนนจุดตัดระดับ A2 และ B1 มีความสอดคล้องกันในระดับดีมาก และการประเมินคะแนนจุดตัดระดับ B2 มีความสอดคล้องกันในระดับดี (Koo & Li, 2016) ค่าสัมประสิทธิ์คอนคอร์ดานส์ของเคนดัลล์ (Kendall's W) ของการประเมินคะแนนจุดตัดระดับ A2 คือ .73 ระดับ B1 คือ .94 และระดับ B2 คือ .89 แสดงว่าการประเมินคะแนนจุดตัดระดับ A2 มีความสอดคล้องกันในระดับดี และการประเมินคะแนนจุดตัดระดับ B1 และ B2 มีความสอดคล้องกันในระดับดีมาก (Landis & Koch, 1977) ค่าสัมประสิทธิ์ครอนบาค (Cronbach's alpha) ของการประเมินคะแนนจุดตัดระดับ A2 คือ .92 ระดับ B1 คือ .92 และระดับ B2 คือ .81 แสดงว่าการประเมินคะแนนจุดตัดระดับ A2 และ B1 มีความสอดคล้องกันในระดับดีมาก และการประเมินคะแนนจุดตัดระดับ B2 มีความสอดคล้องกันในระดับดี (Gliem & Gliem, 2003)

สำหรับความถูกต้องภายนอกนั้น คณะผู้วิจัยตรวจสอบโดยวิธีการ ดังนี้

1. การอภิปรายความสมเหตุสมผลของคะแนนจุดตัดในการแบ่งผู้เข้าสอบ คณะผู้วิจัยเชิญตัวแทนอาจารย์ผู้รับผิดชอบหลักสูตรครุศาสตรบัณฑิต (สาขาวิชาภาษาอังกฤษ) และหลักสูตรศิลปศาสตรบัณฑิต (สาขาภาษาอังกฤษ และสาขาวิชาภาษาอังกฤษธุรกิจ) จำนวนหกคนมาพิจารณาการจัดระดับความสามารถทางภาษาอังกฤษของผู้เข้าสอบตามเกณฑ์มาตรฐาน CEFR ตามคะแนนจุดตัดจากการประเมินรอบที่สาม

อาจารย์ซึ่งเป็นตัวแทนต่างเคยสอนวิชาภาษาอังกฤษและประเมินความสามารถทางภาษาอังกฤษของนักศึกษา ซึ่งเป็นผู้ทดลองสอบแบบทดสอบมาก่อน อาจารย์ผู้สอนมีความเห็นว่าคะแนนจุดตัดของระดับ A2 น่าจะต่ำเกินไป และคะแนนจุดตัดของระดับ B2 น่าจะสูงเกินไป อาจารย์ผู้สอนเห็นว่า หากพิจารณาความเป็นได้ของการเดาคำตอบถูก การกำหนดคะแนนจุดตัดที่ 24 คะแนน จากข้อสอบจำนวน 100 ข้อ อาจจะต่ำเกินไป และการกำหนดคะแนนจุดตัดระดับ B2 ที่ 88 คะแนน ก็อาจสูงเกินไป เพราะมีผู้เข้าสอบบางคนที่อาจารย์ผู้สอนมั่นใจว่ามีทักษะภาษาอังกฤษในระดับ B2 หรือสูงกว่า ซึ่งได้รับการจัดให้อยู่ในระดับ B1 จากนั้นคณะผู้วิจัยจึงประชุมร่วมกับ ผู้ประเมินและตกลงปรับคะแนนจุดตัดตามความเห็นของอาจารย์ผู้สอน โดยกำหนดให้ปรับตามค่าความคลาดเคลื่อนมาตรฐาน (SEM) ค่า SEM ของการประเมินคะแนนจุดตัดระดับ A2 คือ 2.2 และค่า SEM การประเมินคะแนนจุดตัดระดับ B2 คือ 3.65 คณะผู้วิจัยจึงปรับคะแนนจุดตัดระดับ A2 จาก 24 คะแนน เป็น 26 คะแนน และปรับคะแนนจุดตัดระดับ B2 จาก 88 คะแนน เป็น 84 คะแนน ดังนั้นคะแนนจุดตัดที่งานวิจัยนี้เสนอสำหรับระดับ A2 คือ 26 คะแนน ระดับ B1 คือ 64 คะแนน และระดับ B2 คือ 84 คะแนน วิธีการปรับคะแนนของงานวิจัยนี้เป็นไปตามข้อเสนอของนักวิชาการหลายคน (Tannenbaum & Baron, 2015a; Cizek & Bunch, 2007) ซึ่งมองว่า คะแนนจุดตัดที่ได้จากการกำหนดเกณฑ์มาตรฐานไม่ตายตัวและไม่มีคะแนนใดที่อาจถือได้ว่าเป็นคะแนนที่ “ถูกต้อง” ดังนั้นผู้มีส่วนได้ส่วนเสียจึงควรพิจารณาปรับเปลี่ยนคะแนนจุดตัดตามความเหมาะสมของการนำไปใช้ในแต่ละกรณี ไชเชก และบันช์ (Cizek & Bunch, 2007) แนะนำว่า เพื่อให้การปรับคะแนนจุดตัดเป็นไปอย่างมีหลักการ ผู้มีส่วนได้ส่วนเสียสามารถปรับคะแนนจุดตัดให้สูงขึ้นหรือต่ำลงตามค่า SEM ซึ่งช่วยแสดงความเป็นไปได้ใน การคลาดเคลื่อนของการประเมินคะแนนจุดตัดของผู้ประเมิน

2. การเปรียบเทียบการกำหนดระดับความสามารถของผู้เข้าสอบตามประสบการณ์ของอาจารย์ผู้สอนกับการกำหนดระดับความสามารถของผู้เข้าสอบจากผลการสอบตามการกำหนดเกณฑ์มาตรฐาน เนื่องจากตัวแทนอาจารย์ผู้รับผิดชอบหลักสูตรต่างไม่เคยสอนนักศึกษาซึ่งไม่ได้อยู่ในหลักสูตรของตน สำหรับการเปรียบเทียบในส่วนนี้ คณะผู้วิจัยจึงทำ การเปรียบเทียบเฉพาะกลุ่มนักศึกษาหลักสูตรครุศาสตรบัณฑิต สาขาวิชาภาษาอังกฤษ ซึ่งเป็นกลุ่มที่มีจำนวนมากที่สุด คือ 39 คน และขอความอนุเคราะห์อาจารย์หนึ่งคนที่เคยสอนนักศึกษาทั้ง 39 คน ประเมินความสามารถของนักศึกษาตาม เกณฑ์มาตรฐาน CEFR จากประสบการณ์การสอนและการประเมินผลที่ผ่านมาของตน ผลการคำนวณค่าสหสัมพันธ์ของ เพียร์สัน (Pearson's correlation coefficient) ระหว่างการกำหนดระดับความสามารถตามประสบการณ์ของผู้สอน และการกำหนดระดับความสามารถของผู้เข้าสอบจากผลการสอบ คือ .63 หมายความว่ามีความสัมพันธ์ค่อนข้างสูง (อนูวัติ คุณแก้ว, 2566) เป็นที่น่าสังเกตว่าค่าสหสัมพันธ์ของงานวิจัยนี้สูงกว่าค่าสหสัมพันธ์ของงานวิจัยที่ผ่านมา (University of Michigan, 2014) ซึ่งมีค่าสหสัมพันธ์ระหว่างการประเมินของอาจารย์ผู้สอนและการจัดกลุ่มตามคะแนนจุดตัดของการกำหนดมาตรฐานอยู่ที่ .51

3. การเก็บข้อมูลความเที่ยงตรงเชิงกระบวนการเพิ่มเติม โดยการให้ผู้ประเมินทำแบบสำรวจความคิดเห็นของ ผู้ประเมินต่อกิจกรรมการกำหนดมาตรฐาน (รายละเอียดตามตารางที่ 6) ผลการสำรวจความคิดเห็นต่อกิจกรรมการกำหนดมาตรฐานมีคะแนนเฉลี่ยรวมอยู่ที่ 3.78 ซึ่งหมายถึง “เห็นด้วยอย่างยิ่ง” (Todd, 2011) แสดงถึงความพึงพอใจในกิจกรรมของผู้ประเมินที่ค่อนข้างสูง ข้อที่ได้คะแนนสูงสุดมีทั้งหมด 4 ข้อ ได้แก่ ข้อ 1.

กิจกรรมเพื่อสร้างความเข้าใจเกี่ยวกับกรอบมาตรฐาน CEFR ช่วยให้ฉันเข้าใจกรอบมาตรฐาน CEFR ข้อ 2. ข้อสอบที่ใช้ในการฝึกประเมินช่วยให้ฉันเข้าใจกระบวนการประเมิน ข้อ 4. ฉันเข้าใจเรื่องที่มีการอภิปรายในระหว่างการประเมิน และข้อ 6. มีเวลาเพียงพอให้ฉันเข้าร่วมพูดคุยแสดงความคิดเห็น โดยมีคะแนนเฉลี่ยเท่ากับ 4.00 ทุกข้อ แสดงให้เห็นว่าการให้ข้อมูลและการฝึกประเมินมีความชัดเจน รวมถึงการเปิดโอกาสให้แสดงความคิดเห็นทำให้ผู้ประเมินเข้าใจกระบวนการในการประเมินและมั่นใจมากขึ้น ในขณะที่ข้อที่ได้คะแนนต่ำที่สุดได้แก่ ข้อ 7. ฉันมีความมั่นใจในการตัดสินใจในระหว่างการประเมิน ซึ่งมีคะแนนเฉลี่ยอยู่ที่ 3.30 สอดคล้องกับการศึกษาของมหาวิทยาลัยมิชิแกน (University of Michigan, 2014) ซึ่งผู้ประเมินให้คะแนนข้อนี้น้อยที่สุด (3.09) ผู้ประเมินแสดงความเห็นว่าการแยกแยะเกณฑ์มาตรฐาน CEFR ในระดับที่ใกล้เคียงกันเป็นเรื่องที่ยาก ผู้ประเมินเองควรจะต้องศึกษาเพิ่มเติมและเพิ่มพูนประสบการณ์ในการประเมินเพื่อให้มีความมั่นใจมากขึ้น ข้อที่ได้คะแนนต่ำเป็นอันดับสองคือ ข้อ 8. ฉันพอใจกับคะแนนจุดตัดรอบที่สาม เนื่องจากผู้ประเมินบางคนเห็นว่าคะแนนจุดตัดในระดับ A2 ก่อนการปรับตามค่าความคลาดเคลื่อนมาตรฐาน (SEM) ต่ำเกินไป

ตารางที่ 6 ผลการตอบแบบสำรวจความคิดเห็นของผู้ประเมินต่อกิจกรรมการกำหนดมาตรฐาน

คำถาม	คะแนนเฉลี่ย	SD	ความหมาย
1. กิจกรรมเพื่อสร้างความเข้าใจเกี่ยวกับกรอบมาตรฐาน CEFR ช่วยให้ฉันเข้าใจกรอบมาตรฐาน CEFR	4.00	.0	เห็นด้วยอย่างยิ่ง
2. ข้อสอบที่ใช้ในการฝึกประเมินช่วยให้ฉันเข้าใจกระบวนการประเมิน	4.00	.0	เห็นด้วยอย่างยิ่ง
3. ฉันเข้าใจคำสั่งในการประเมินแต่ละรอบ	3.60	.52	เห็นด้วยอย่างยิ่ง
4. ฉันเข้าใจเรื่องที่มีการอภิปรายในระหว่างการประเมิน	4.00	.0	เห็นด้วยอย่างยิ่ง
5. ฉันมีเวลาเพียงพอในการประเมินข้อสอบ	3.90	.32	เห็นด้วยอย่างยิ่ง
6. มีเวลาเพียงพอให้ฉันเข้าร่วมพูดคุยแสดงความคิดเห็น	4.00	.0	เห็นด้วยอย่างยิ่ง
7. ฉันมีความมั่นใจในการตัดสินใจในระหว่างการประเมิน	3.30	.48	เห็นด้วยอย่างยิ่ง
8. ฉันพอใจกับคะแนนจุดตัดรอบที่สาม	3.50	.42	เห็นด้วยอย่างยิ่ง
รวม	3.78	.28	เห็นด้วยอย่างยิ่ง

สรุปและอภิปรายผล

ผลการวิจัยสามารถสรุปและอภิปรายตามวัตถุประสงค์ได้ดังนี้

1. การกำหนดคะแนนจุดตัดตามกรอบมาตรฐาน CEFR ของแบบทดสอบภาษาอังกฤษของหลักสูตร โดยวิธี Yes/No Angoff ซึ่งกำหนดให้คะแนนจุดตัดระดับ A2 อยู่ที่ 26 คะแนน ระดับ B1 อยู่ที่ 64 คะแนน และระดับ B2 อยู่ที่ 84 คะแนน สามารถอภิปรายได้ดังนี้

ปัจจัยสำคัญประการแรกที่ทำให้คะแนนจุดตัดอยู่ในระดับดังกล่าว อาจเนื่องมาจากปัจจัยด้านกระบวนการออกแบบและพัฒนาแบบทดสอบ กล่าวคือ แบบทดสอบฉบับนี้ได้ดำเนินการออกแบบและพัฒนาตามหลักการสากลด้านการวัดและประเมินผล (Hughes, 2003) มีการกำหนดคุณลักษณะของแบบทดสอบที่ชัดเจน มีการกำหนดสัดส่วนโครงสร้างของ การวัดผลประเมินผลตามแต่ละทักษะภาษาอังกฤษ

และมีการกำหนดอัตราสัดส่วนน้ำหนักของจำนวนข้อสอบในแต่ละระดับขั้นความรู้ทางภาษาอังกฤษ อีกทั้งแบบทดสอบยังได้รับการออกแบบมาให้สอดคล้องกับคำอธิบายความสามารถทางภาษาตามกรอบมาตรฐาน CEFR ในระดับ A2 B1 และ B2 (25%, 50% และ 25% ของจำนวนข้อสอบ) แบบทดสอบมีค่าความเที่ยงตรงตามเนื้อหา ค่าความเชื่อมั่น ค่าความยากง่าย และค่าอำนาจจำแนกในระดับที่ดี จึงส่งผลให้คะแนนจุดตัดที่ได้ใกล้เคียงกับโครงสร้างแบบทดสอบซึ่งทางคณะผู้จัดทำกำหนดไว้ นอกจากนี้โครงสร้างของแบบทดสอบยังส่งผลให้ระยะห่างของคะแนนจุดตัดในระดับต่างๆ มีความต่างกัน กล่าวคือคะแนนจุดตัดในระดับ A2 กับระดับ B1 ห่างกัน 38 คะแนน ในขณะที่คะแนนจุดตัดในระดับ B1 กับระดับ B2 ห่างกัน 20 คะแนน สอดคล้องกับผลการวิจัยที่ผ่านมา (Nakanitanon, 2021; Wudthayagorn, 2018) ซึ่งทำการกำหนดมาตรฐานกับแบบทดสอบที่มีคุณสมบัติเฉพาะที่แตกต่างกันไป ทำให้ระยะห่างของคะแนนจุดตัดระหว่างความสามารถตามกรอบมาตรฐาน CEFR ในแต่ละระดับมีระยะห่างที่ไม่เท่ากัน

หากพิจารณาถึงรายละเอียดในระดับกระบวนการแล้ว ขั้นตอนกระบวนการกำหนดมาตรฐานที่งานวิจัยนี้ใช้เป็นไปตามข้อเสนอของเพลก และไซเซก (Plake & Cizek, 2012) กล่าวคือ มีการอบรมผู้ประเมินเพื่อซักซ้อมความเข้าใจเรื่องวัตถุประสงค์ วิธีการกำหนดมาตรฐาน และทดลองประเมินข้อสอบ ผู้ประเมินสามารถอภิปรายแสดงเหตุผลของการประเมินของตนเองได้ เช่น สามารถอภิปรายความสมเหตุสมผลของคะแนนจุดตัดในการแบ่งผู้เข้าสอบ อภิปรายผลกระทบหากนำคะแนนจุดตัดไปใช้ในการจัดระดับความสามารถทางภาษาอังกฤษตามกรอบมาตรฐาน CEFR ของนักศึกษาที่ทดลองทำข้อสอบ สามารถใช้ข้อมูลสารสนเทศช่วยในการประเมิน อาจกล่าวได้ว่ากระบวนการอภิปรายย่อยเหล่านี้ ทำให้ผู้ประเมินมีมุมมองที่สอดคล้องกันมากขึ้น และส่งผลให้ค่า SD ในภาพรวมลดลงในการหาคะแนนจุดตัดในแต่ละรอบ ซึ่งสอดคล้องกับงานวิจัยที่ผ่านมา (Nakanitanon, 2021; University of Michigan, 2014) ซึ่งค่า SD ค่อยๆ ลดลงในการหาคะแนนจุดตัดในแต่ละรอบผ่านการอภิปรายผลการประเมิน

สำหรับในแง่ของการใช้ประโยชน์นั้น ไซเซก (Cizek, 2012) อธิบายไว้ว่าการกำหนดมาตรฐาน คือการกำหนดตัวเลขที่จะใช้ในการแยกแยะผลการทดสอบ คะแนนจุดตัดที่ได้จึงช่วยในการให้ความหมายกับผลการสอบ จิรดา วุฒยากร (Wudthayagorn, 2018) กล่าวเพิ่มเติมในประเด็นนี้ว่า การกำหนดคะแนนจุดตัดช่วยในการอธิบายระดับความสามารถของ ผู้เข้าสอบ เช่น หากผู้เข้าสอบได้รับการจัดระดับในระดับ B1 หมายความว่า บุคคลผู้นั้นจะสามารถอ่านและฟังเพื่อจับใจความสำคัญของเรื่องซึ่งตนคุ้นเคยในบริบทที่ทำงาน โรงเรียน และการเดินทางได้ ข้อมูลที่ได้จากงานวิจัยนี้จึงช่วยให้ทราบระดับความสามารถของผู้เข้าสอบอย่างละเอียดมากขึ้น อันจะเป็นผลดีต่อการตัดสินใจจัดการเรียนการสอนให้เหมาะสมกับผู้เรียนต่อไป

2. การตรวจสอบค่าความถูกต้องภายในของกระบวนการกำหนดมาตรฐานผ่านการหาค่าสหสัมพันธ์ภายในชั้น (ICC) ของการประเมิน และการตรวจสอบค่าความถูกต้องภายนอกผ่านการอภิปรายความเหมาะสมของคะแนนจุดโดยตัวแทนอาจารย์ผู้รับผิดชอบหลักสูตร การเปรียบเทียบการจัดกลุ่มผู้เข้าสอบตามเกณฑ์มาตรฐาน CEFR ของอาจารย์ผู้สอนและการจัดกลุ่มโดยคะแนนจุดตัด และการประเมินความเที่ยงตรงของกระบวนการโดยการถามความเห็นของผู้ประเมินต่อกระบวนการกำหนดมาตรฐานมีประเด็นที่สามารถอภิปรายได้ดังนี้

ค่าสหสัมพันธ์ภายในชั้น (ICC) ของการประเมินของงานวิจัยนี้อยู่ในเกณฑ์ดี และดีมาก (คะแนนจุดตัดในระดับ A2 และ B1 มีค่าความสอดคล้องกัน = .92 และการประเมินสำหรับระดับ B2 มีค่าความสอดคล้องกัน = .81) ปัจจัยสำคัญประการหนึ่งที่ทำให้ได้ผลเช่นนี้อาจเนื่องมาจากปัจจัยด้านคุณสมบัติของผู้ประเมิน กล่าวคือ ความถูกต้องแม่นยำของการประเมินจากผู้ประเมินในงานวิจัยนี้ส่งผลตรงต่อค่าความถูกต้องของกระบวนการกำหนดมาตรฐานและคะแนนจุดตัดที่กำหนด สอดคล้องกับที่ฮิวส์ (Hughes, 2003) กล่าวว่า หากผู้ประเมินได้รับการฝึกฝน อบรม ก่อนการประเมินแล้วจะส่งผลให้ การประเมินมีค่าความเชื่อมั่นที่ดี กระบวนการที่ช่วยให้ผู้ประเมินสามารถประเมินได้สอดคล้องกันอาจเป็นผลมาจากกิจกรรมก่อนการประเมินที่ผู้ประเมินต้องอบรมและอภิปรายเพื่อสร้างความเข้าใจเกี่ยวกับวัตถุประสงค์ของการกำหนดมาตรฐานแบบ Yes/No Angoff ลักษณะของผู้เข้าสอบที่มีความสามารถต่ำที่สุดตามกรอบมาตรฐาน CEFR แต่ละระดับ และกรอบมาตรฐาน CEFR ที่ใช้ในการประเมินข้อสอบ นอกจากนั้นผู้ประเมินทุกคนยังต้องผ่านการทำแบบทดสอบความเข้าใจในเกณฑ์เหล่านี้ก่อนทำการประเมินด้วย คุณสมบัติอีกหนึ่งประการที่ส่งผลต่อความถูกต้องของการประเมินคือผู้ประเมินทุกคนมีประสบการณ์ การสอนภาษาอังกฤษในระดับมัธยมศึกษาและ/หรือระดับอุดมศึกษามากกว่า 5 ปี และเคยมีประสบการณ์ในการพัฒนาข้อสอบวัดความสามารถทางภาษาอังกฤษตามกรอบมาตรฐาน CEFR มาแล้ว อาจกล่าวได้ว่าการอบรมและการทดสอบเหล่านี้เป็นส่วนหนึ่ง ที่ก่อให้เกิดความถูกต้องเชิงกระบวนการ

ความถูกต้องภายนอกของกระบวนการกำหนดมาตรฐานและคะแนนจุดตัดที่กำหนดได้รับการตรวจสอบผ่านวิธีการสามวิธี สำหรับวิธีที่หนึ่งเป็นการอภิปรายความเหมาะสมของคะแนนจุดโดยตัวแทนอาจารย์ผู้รับผิดชอบหลักสูตร ซึ่งใกล้ชิดและเคยสอนนักศึกษาซึ่งเป็นผู้เข้าสอบ อาจารย์ผู้รับผิดชอบหลักสูตร เห็นว่าคะแนนจุดตัดระดับ A2 ควรจะสูงกว่าที่กำหนด และคะแนนจุดตัดระดับ B2 ควรจะต่ำกว่าที่กำหนด ผู้วิจัยจึงได้ปรับคะแนนตามความเห็นของอาจารย์ผู้รับผิดชอบหลักสูตร โดยปรับตามค่าความคลาดเคลื่อนมาตรฐาน (SEM) ของการประเมิน วิธีที่สองในตรวจสอบความถูกต้องภายนอกคือ การเปรียบเทียบการจัดกลุ่มผู้เข้าสอบตามเกณฑ์มาตรฐาน CEFR ของอาจารย์ตามประสบการณ์ที่เคยสอนนักศึกษาผู้เข้าสอบกับการจัดกลุ่มโดยการกำหนดคะแนนจุดตัด ผลการคำนวณค่าสหสัมพันธ์ของเพียร์สัน (Pearson's correlation coefficient) ของการจัดกลุ่มโดยอาจารย์ผู้สอนและการจัดกลุ่มโดยคะแนนจุดตัดมีความสัมพันธ์ค่อนข้างสูง (.63) และวิธีที่สามคือ การประเมินความเที่ยงตรงของกระบวนการโดยการถามความเห็นของผู้ประเมินเกี่ยวกับกระบวนการกำหนดมาตรฐานที่ตนได้เข้าร่วม พบว่าผู้ประเมินเห็นด้วยอย่างยิ่ง (3.78) กับกระบวนการกำหนดมาตรฐานในภาพรวม การตรวจสอบค่าความถูกต้องภายในผ่านการวิเคราะห์ค่าสหสัมพันธ์ภายในชั้น (ICC) การตรวจสอบความถูกต้องภายนอกผ่านการประเมินของอาจารย์ผู้สอน และการตรวจสอบความถูกต้องของกระบวนการผ่านการสำรวจความคิดเห็นของผู้ประเมินเป็นวิธีการเดียวกันกับที่มหาวิทยาลัยมิชิแกน (University of Michigan, 2014) ใช้ในการตรวจสอบความถูกต้องของการกำหนดมาตรฐานให้กับแบบทดสอบความสามารถทางภาษาอังกฤษของตน

ข้อเสนอแนะ

คณะผู้วิจัยมีข้อเสนอแนะในการจัดกิจกรรมการกำหนดมาตรฐานด้วยวิธี Yes/No Angoff ดังนี้

1. การกำหนดมาตรฐานให้กับแบบทดสอบความสามารถทางภาษาอังกฤษตามเกณฑ์มาตรฐาน CEFR ด้วยวิธี Yes/No Angoff เป็นกระบวนการที่มีความซับซ้อน และมีขั้นตอนที่ต้องดำเนินการหลายขั้นตอน ดังนั้นผู้ประเมินควรได้รับการอบรม หรือฝึกปฏิบัติจนเกิดความชำนาญก่อนการหาคะแนนจุดตัด หากผู้ประเมินขาดความชำนาญอาจส่งผลกระทบโดยตรงต่อความเที่ยงตรงและความเชื่อมั่นของการกำหนดมาตรฐาน งานวิจัยนี้ใช้เวลาอบรมผู้ประเมินประมาณหกชั่วโมงด้วยเหตุผลเรื่องเวลาและงบประมาณที่จำกัด งานวิจัยในอนาคตควรพิจารณาเพิ่มเวลาในการอบรมผู้ประเมินก่อนการกำหนดมาตรฐาน การอบรมที่เพียงพอและมีประสิทธิภาพนอกจากจะทำให้ประเมินได้อย่างแม่นยำ ยังจะช่วยให้สามารถคัดเลือกผู้ประเมินได้ง่าย และหลากหลายยิ่งขึ้น เช่น คัดเลือกผู้ประเมินจากหน่วยงานเอกชนซึ่งมีความชำนาญในการใช้ภาษาอังกฤษแต่ไม่เคยมีประสบการณ์ในการพัฒนาข้อสอบตามเกณฑ์มาตรฐาน CEFR ยิ่งไปกว่านั้น เนื่องจากงานวิจัยนี้ใช้ผู้ประเมินจำนวน 10 คน ซึ่งเป็นจำนวนขั้นต่ำตามที่งานวิจัยอื่นเสนอไว้ (Fowell, Fewtrell & McLaughlin, 2008) ทำให้ต้องนำการประเมินของผู้ประเมินทุกคนมาคำนวณหาคะแนนจุดตัด หากผู้วิจัยในอนาคตเลือกผู้ประเมินมาได้มากกว่า 10 คน ผู้วิจัยสามารถยกเว้นไม่นำผลการประเมินของผู้ประเมินบางคนมาคำนวณ เช่น ผลการประเมินของผู้ประเมินที่ไม่สามารถเข้าร่วมกิจกรรมทั้งหมดได้ หรือผลการประเมินของผู้ประเมินที่ทำแบบทดสอบความเข้าใจในเกณฑ์มาตรฐาน CEFR ได้คะแนนน้อย การยกเว้นดังกล่าวจะทำให้คะแนนจุดตัดที่ได้มีความน่าเชื่อถือมากขึ้น นอกจากนี้ ผู้วิจัยยังสามารถยกเว้นไม่นำผลการประเมินที่สูงหรือต่ำกว่าค่าเฉลี่ยของกลุ่มอย่างมากมาคำนวณ ซึ่งจะช่วยให้การประเมินโดยรวมมีค่า SD ที่น้อยลง

2. การกำหนดค่ามาตรฐานแบบทดสอบความสามารถทางภาษาอังกฤษมีความละเอียดซับซ้อนและใช้เวลามาก อาจมีปัจจัยแทรกแซงที่ทำให้การดำเนินการไม่เป็นไปตามแผนที่กำหนดไว้ ดังนั้น คณะทำงานควรมีความยืดหยุ่นในการดำเนินกิจกรรมตามบริบทที่แตกต่างกันไปของแต่ละสถาบันการศึกษา เช่น หากผู้ประเมินไม่สะดวกเข้าร่วมในทุกกิจกรรมใน สถานที่ตั้งเป็นเวลาสามวัน อาจจัดกิจกรรมบางส่วนเป็นการประชุมรูปแบบออนไลน์

3. ผู้พัฒนาข้อสอบและผู้วิจัยในอนาคตสามารถนำวิธีการตรวจสอบความถูกต้องของการกำหนดมาตรฐานในงานวิจัยนี้ไปใช้ เพื่อสร้างความเชื่อมั่นให้กับผู้เข้าสอบที่จะได้รับผลกระทบจากการกำหนดมาตรฐานของตน อย่างไรก็ตาม วิธีการตรวจสอบความถูกต้องที่งานวิจัยนี้นำมาใช้เป็นเพียงส่วนหนึ่งเท่านั้น งานวิจัยในอนาคตสามารถทำการตรวจสอบความถูกต้องโดยวิธีการอื่นเพิ่มเติมได้ เช่น การกำหนดมาตรฐานให้กับแบบทดสอบเดิมด้วยวิธีการที่ต่างไป เช่น วิธี Bookmark หรือวิธี Nedelsky นอกจากนั้นยังสามารถทำการตรวจสอบความถูกต้องได้ด้วยการประเมินแบบทดสอบเดิมโดยผู้ประเมินอีกคณะหนึ่ง เพื่อนำผลการประเมินของสองคณะมาเปรียบเทียบกัน

รายการอ้างอิง

- คณะกรรมการมาตรฐานการอุดมศึกษา, กระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม. (2567). *ประกาศคณะกรรมการมาตรฐานการอุดมศึกษา เรื่อง นโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา พ.ศ. 2567*. <https://www.ops.go.th/th/ches-downloads/edu-standard/item/9625-2567>
- สุพัฒน์ สุขมลสันต์. (2564). บทวิจารณ์เกี่ยวกับการวิจัยเพื่อกำหนดเกณฑ์ระดับความสามารถของแบบทดสอบด้วย Modified Angoff Method ตามเกณฑ์ของ CEFR และข้อเสนอแนะ: กรณีศึกษาบทความวิจัย 2 เรื่อง. *วารสารภาษาปริทัศน์*, 36, 73–92.
- อนุวัติ คุณแก้ว. (2566). *การทดสอบ การวัดผลและประเมินผล การศึกษาแนวใหม่*. สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- Athiworakun, 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.
- Cizek, G. J. (2012). An introduction to contemporary standard setting: Concepts, characteristics, and contexts. In G. J. Cizek (ed.), *Setting performance standards: Foundations, methods, and innovations* (pp. 3-14). Routledge.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standard on tests*. Sage Publications.
- Council of Europe. (2009). *Relating language examinations to the common European framework of reference for languages: Learning, teaching, assessment (CEFR)*. <https://www.coe.int/en/web/common-european-framework-reference-languages/relating-examinations-to-the-cefr>
- Fowell, S. L., Fewtrell, R., & P. J. McLaughlin. (2008). Estimating the minimum number of judges required for test-centered standard setting on written assessments. Do discussion and iteration have an influence? *Adv in Health Sci Edu*, 13, 11-24.
- Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's Alpha reliability coefficient for Likert-type scales. *Paper presented at Practice Conference in Adult, Continuing, and Community Education*, Columbus.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press.
- Impara, J. C., & Plake, B. S. (1997). Standard setting: An alternative approach, *Journal of Educational Measurement*, 25(2), 353-366.
- Koo, T. K., & Li, M. Y. (2016). A Guideline of selecting and reporting intraclass correlation coefficients for reliability research, *Journal of Chiropractic Medicine*, 15(2), 155–163.

- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-74.
- Nakanitanon, P. (2021). Linking an English proficiency test to CEFR: Setting valid cut scores. *The New English Teacher*, 15(1), 53-64.
- Plake, B., & Cizek, G. J. (2012). Variations on a theme: The modified Angoff, extended Angoff, and yes/no standard setting methods. In G. J. Cizek (ed.), *Setting performance standards: Foundations, methods, and innovations* (pp. 181-200). Routledge.
- Tannenbaum, R. J., & Baron, P. A. (2015a). Mapping scores from the TOEFL Junior Comprehensive Test onto the Common European Framework of Reference (CEFR). *Educational Testing Service*.
- Tannenbaum, R. J., & Baron, P. A. (2015b). *Mapping TOEIC scores to the Vietnamese National Standard: A study to recommend English language requirements for admissions into and graduation from Vietnamese universities*. Educational Testing Service.
- Tiyawongsuwan, S., Lawthong, N., & Kanjanawasee, S. (2013). Standard setting for educational evaluation: Concepts, methods and applications. *Kasetsart Journal of Social Sciences*, 34(2), 225-240.
- 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.
- Wudthayagorn, J. (2018). Mapping the CU-TEP to the common European framework of reference (CEFR). *LEARN: Language Education and Acquisition Research Network Journal*, 11(2), 163-180.
- University of Michigan. (2014). *Linking the Common European Framework of Reference and the MET writing test*. <https://michiganassessment.org>