

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ ผู้วิจัยได้ทบทวนเอกสารและงานวิจัยที่เกี่ยวข้อง ตลอดจนบทความทางวิชาการทั้งในประเทศและต่างประเทศ แบ่งออกเป็น 6 ตอน ได้แก่ 2.1 กรอบอ้างอิงความสามารถทางภาษา 2.2 การสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราช 2.3 แนวคิดเกี่ยวกับการประเมินความสอดคล้องของเครื่องมือวัดหรือเครื่องมือประเมิน 2.4 มาตรฐานของการทดสอบทางการศึกษาและจิตวิทยา 2.5 การกำหนดมาตรฐานเพื่อกำหนดคะแนนจุดตัด และ 2.6 กรอบแนวคิดในการวิจัย รายละเอียดแต่ละตอนมีดังนี้

2.1 กรอบอ้างอิงความสามารถทางภาษา

เนื้อหาในตอนนี้ แบ่งออกเป็น 2 หัวข้อ ได้แก่ 2.1.1 กรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรป และ 2.1.2 กรอบ CEFR กับการจัดการเรียนการสอนภาษาอังกฤษของประเทศไทย รายละเอียดแต่ละหัวข้อมีดังนี้

2.1.1 กรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรป

กรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรป (The Common European Framework of Reference for Languages) หรือกรอบ CEFR เป็นมาตรฐานการประเมินความสามารถทางภาษาที่สหภาพยุโรป จัดทำขึ้นในสหภาพยุโรป ในปี ค.ศ. 2001 (ปีแห่งภาษายุโรป) หลังจากผ่านกระบวนการร่าง การนำร่อง และการให้คำแนะนำอย่างครอบคลุม กรอบ CEFR ได้ดำเนินการตามหลักการศึกษาร่วมกันของสภาแห่งสหภาพยุโรป มีวัตถุประสงค์เพื่อเป็นแนวทางในการเรียนรู้ (learning) การสอน (teaching) และการประเมิน (assessment) (Council of Europe, 2020)

กรอบ CEFR ได้รับการยอมรับอย่างกว้างขวางว่าเป็นมาตรฐานในการจัดลำดับความสามารถทางภาษาของแต่ละบุคคล ผู้ใช้สามารถนำกรอบ CEFR ไปใช้อธิบายความสามารถของผู้ใช้ภาษาในระดับต่าง ๆ ได้มากขึ้น โดยคำอธิบายคุณลักษณะ (descriptors) ของกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปในภาพรวม หรือ Common Reference Levels: global scale สามารถแบ่งออกเป็น 3 กลุ่มหลัก ได้แก่ ผู้ใช้ภาษาขั้นพื้นฐาน (basic user) ผู้ใช้ภาษาขั้นอิสระ (independent user) และผู้ใช้ภาษาขั้นคล่องแคล่ว (proficient user) ประกอบด้วย 6 ระดับ ได้แก่

A1 A2 B1 B2 C1 และ C2 แต่ละระดับมีคำอธิบายคุณลักษณะที่ผู้เรียนสามารถทำได้ รายละเอียดแสดงดังตารางที่ 1 และภาพที่ 1 (Council of Europe, 2001)

ตารางที่ 1 กรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปในภาพรวม

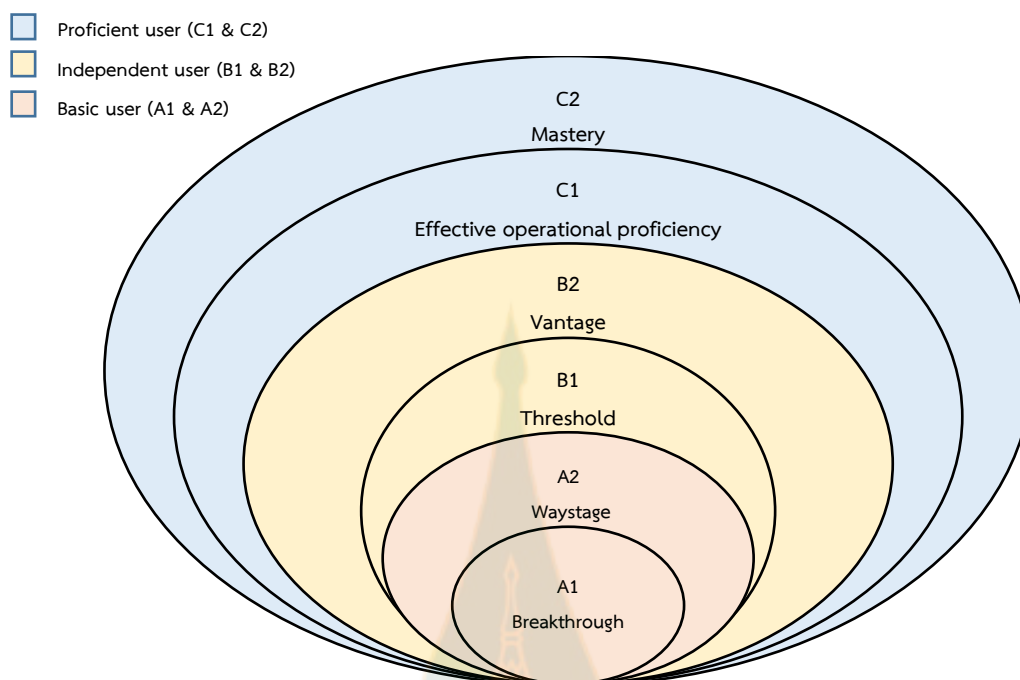
กลุ่ม	ระดับ	คำอธิบาย
ผู้ใช้ภาษา ขั้นพื้นฐาน (Basic user)	ระดับ A1 Breakthrough	<p>สามารถเข้าใจและใช้สำนวนที่คุ้นเคยในชีวิตประจำวันและวลีพื้นฐานที่เน้นความต้องการที่เป็นรูปธรรม สามารถแนะนำตัวเองและผู้อื่น และสามารถถามและตอบคำถามเกี่ยวกับรายละเอียดส่วนตัว เช่น เขาหรือเธออาศัยอยู่ที่ไหน เขาหรือเธอรู้จักใครบ้าง เขาหรือเธอมีสิ่งของใดบ้าง สามารถโต้ตอบง่ายๆ หากคู่สนทนาพูดช้าและชัดเจน และพร้อมที่จะโต้ตอบเมื่อได้รับความช่วยเหลือจากคู่สนทนา</p> <p>Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.</p>
	ระดับ A2 Waystage	<p>สามารถเข้าใจประโยคและสำนวนที่ใช้บ่อยซึ่งเกี่ยวข้องกับเรื่องใกล้ ๆ ตัว เช่น ข้อมูลส่วนตัวและครอบครัว การซื้อของ ภูมิศาสตร์ท้องถิ่น การจ้างงาน สามารถสื่อสารในเรื่องง่าย ๆ และเป็นกิจกรรมที่ต้องมีการแลกเปลี่ยนข้อมูลง่าย ๆ ตรงไปตรงมา เกี่ยวกับสิ่งที่คุ้นและทำเป็นประจำ สามารถบรรยายโดยใช้คำง่าย ๆ ที่เกี่ยวกับข้อมูลพื้นฐานของตนเอง สภาพแวดล้อม และเรื่องอื่น ๆ ที่ต้องการเร่งด่วน</p> <p>Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.</p>

ตารางที่ 1 กรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปในภาพรวม (ต่อ)

กลุ่ม	ระดับ	คำอธิบาย
ผู้ใช้ภาษา ขั้นอิสระ (Independent user)	ระดับ B1 Threshold	<p>สามารถเข้าใจประเด็นสำคัญของเรื่องทั่ว ๆ ไปได้ชัดเจนในเรื่องที่คุ้นเคย ซึ่งมักพบเห็นอยู่เป็นประจำในการทำงาน โรงเรียน การใช้เวลารว่าง ฯลฯ สามารถจัดการกับสถานการณ์ต่าง ๆ ที่อาจเกิดขึ้นระหว่างการเดินทางท่องเที่ยวไปในสถานที่ที่ใช้ภาษาอังกฤษได้ สามารถพูดหรือเขียนข้อความที่ง่าย ๆ ติดต่อกันเกี่ยวกับเรื่องที่คุ้นเคยหรือเรื่องที่ตัวเองสนใจ สามารถบรรยายประสบการณ์และเหตุการณ์ ความฝัน ความหวัง ความทะเยอทะยาน พร้อมทั้งให้เหตุผลสั้น ๆ ตามความคิดเห็นและแผนการได้</p> <p>Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.</p>
	ระดับ B2 Vantage	<p>สามารถเข้าใจประเด็นสำคัญของเรื่องที่ซับซ้อนทั้งเรื่องที่เป็นรูปธรรมและนามธรรม รวมถึงสามารถอภิปรายเชิงเทคนิคในเรื่องที่ตนเองเชี่ยวชาญได้ สามารถสื่อสารได้อย่างคล่องแคล่วและเป็นธรรมชาติ ทำให้สามารถสื่อสารกับเจ้าของภาษาได้อย่างราบรื่น ไม่ติดขัด สามารถพูดหรือเขียนข้อความที่มีรายละเอียดชัดเจนในหัวข้อที่หลากหลาย และสามารถอธิบายมุมมองเกี่ยวกับประเด็นที่กำหนดให้ พร้อมทั้งให้ข้อดีและข้อเสียเพื่อเป็นทางเลือกที่หลากหลาย ๆ</p> <p>Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.</p>

ตารางที่ 1 กรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปในภาพรวม (ต่อ)

กลุ่ม	ระดับ	คำอธิบาย
ผู้ใช้ภาษาขั้น คล่องแคล่ว (Proficient user)	ระดับ C1 Effective operational proficiency	<p>สามารถเข้าใจข้อความยาว ๆ ในหัวข้อที่หลากหลาย และเข้าใจความหมายแฝงได้ สามารถแสดงความคิดเห็นหรือความรู้สึกของตนเองได้อย่างคล่องแคล่วและเป็นธรรมชาติ โดยไม่ต้องหยุดค้นหาคำศัพท์ สามารถใช้ภาษาได้อย่างยืดหยุ่นและมีประสิทธิภาพเพื่อวัตถุประสงค์ทางสังคม วิชาการ และวิชาชีพ สามารถพูดหรือเขียนข้อความที่ซับซ้อนได้อย่างชัดเจน ถูกต้องตามโครงสร้างไวยากรณ์ และสามารถเชื่อมโยงประโยคได้ถูกต้อง</p> <p>Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.</p>
	ระดับ C2 Mastery	<p>สามารถเข้าใจทุกสิ่งทุกอย่างที่ได้ยินหรืออ่านได้อย่างง่ายดาย สามารถสรุปข้อมูลจากแหล่งข้อมูลที่พูดและเขียนที่หลากหลายได้ และสามารถสร้างประเด็นโต้แย้งและนำเสนอได้อย่างต่อเนื่อง สามารถใช้ภาษาได้อย่างเป็นธรรมชาติ คล่องแคล่ว และแม่นยำ และแยกแยะความหมายที่ใกล้เคียงกันในสถานการณ์ที่ซับซ้อนได้</p> <p>Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.</p>



ภาพที่ 1 ระดับความสามารถในการใช้ภาษาอังกฤษตามกรอบ CEFR

แม้ว่ากรอบ CEFR ได้รับการยอมรับและนำไปใช้อย่างแพร่หลาย แต่การประยุกต์ในบริบทนอกยุโรปยังเผชิญกับความท้าทาย เช่น ความไม่สอดคล้องของหลักสูตร ความจำเป็นในการปรับให้เหมาะสมกับวัฒนธรรม และข้อจำกัดด้านผู้เชี่ยวชาญในท้องถิ่น อีกทั้งความเป็นนามธรรมเชิงแนวคิดและการระบุรายละเอียดที่ไม่ชัดเจนของกรอบ CEFR อาจทำให้การศึกษาความสอดคล้องระหว่างแบบทดสอบ STOU-EPT กับกรอบ CEFR มีความซับซ้อน (Alderson, 2005; Hulstijn, 2007) หากดำเนินงานโดยยึดแนวปฏิบัติของสภาแห่งสหภาพยุโรปอย่างเข้มงวดและคำนึงถึงบริบท เช่น Manual for Language Test Development and Examining การศึกษาความสอดคล้องตามกรอบ CEFR สามารถเพิ่มความโปร่งใส และสนับสนุนการตัดสินใจที่มีความสำคัญสูงได้ (high-stakes) เช่น การรับเข้าศึกษา การสำเร็จการศึกษา (Council of Europe, 2001, 2018, 2020; Figueras, 2012)

การวิจัยนี้มุ่งศึกษาความสอดคล้องระหว่างแบบทดสอบ STOU-EPT กับกรอบ CEFR โดยแบบทดสอบ ประกอบด้วย 3 ทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน แต่ละทักษะแบ่งเป็น 3 ระดับ ได้แก่ ระดับ A2 B1 และ B2 เพื่อให้กระบวนการประเมินความสอดคล้องเป็นระบบและสามารถอ้างอิงกรอบแนวคิดที่ได้รับการยอมรับในระดับสากล ผู้วิจัยจึงได้นำคำอธิบายคุณลักษณะ (descriptors) ตามกรอบ CEFR มาประยุกต์โดยอ้างอิงจาก EAQUALS Bank of Descriptors (EAQUALS, 2008) ซึ่งให้คำอธิบายคุณลักษณะของผู้เรียนในแต่ละระดับของกรอบ CEFR อย่างเป็นลำดับขั้น

EAQUALS Bank of Descriptors ครอบคลุมหลายทักษะ ได้แก่ การฟัง การอ่าน การพูด และการเขียน และครอบคลุมระดับความสามารถตั้งแต่ A1 A2 B1 B2 C1 และ C2 ถูกนำมาใช้เป็นกรอบแนวทางในการพัฒนาแบบประเมินความสอดคล้องระหว่างข้อสอบ STOU-EPT กับกรอบ CEFR อย่างไรก็ตาม เนื่องจาก EAQUALS Bank of Descriptors ไม่ได้ให้คำอธิบายคุณลักษณะด้านโครงสร้างภาษาโดยตรง ผู้วิจัยจึงเลือกใช้เฉพาะคำอธิบายคุณลักษณะที่เกี่ยวข้องกับทักษะการฟัง และการอ่าน ซึ่งสามารถเชื่อมโยงกับกรอบ CEFR ได้อย่างชัดเจน และแต่ละทักษะจำแนกเป็น 3 ระดับความสามารถ ได้แก่ ระดับ A2 B1 และ B2 ทั้งนี้เพื่อให้ตรงกับโครงสร้างของแบบทดสอบ STOU-EPT รายละเอียดแสดงดังตารางที่ 2

ตารางที่ 2 คำอธิบายคุณลักษณะตามกรอบ CEFR จำแนกตามทักษะการฟัง และการอ่าน

ทักษะ (skill)	คำอธิบายคุณลักษณะ (descriptors)
การฟัง	
ระดับ A2	<p>สามารถเข้าใจข้อมูลและคำถามง่าย ๆ เกี่ยวกับครอบครัว บุคคล ที่อยู่อาศัย สถานที่ทำงาน และงานอดิเรก สามารถเข้าใจสิ่งที่ผู้พูดพูดในเรื่องง่าย ๆ ที่ใช้ในชีวิตประจำวัน ถ้าผู้พูดพูดอย่างช้า ๆ และชัดเจน สามารถเข้าใจการสนทนาสั้น ๆ เกี่ยวกับครอบครัว งานอดิเรก และการใช้ชีวิตประจำวัน หากผู้สนทนาพูดอย่างช้า ๆ และชัดเจน สามารถเข้าใจเมื่อมีการเปลี่ยนหัวข้อในรายงานข่าวทีวี และเข้าใจข้อมูลหลัก สามารถเข้าใจข้อความง่าย ๆ ชัดเจน และสั้นกระชับ ที่สนามบินสถานีรถไฟ ฯลฯ เช่น รถไฟไปลอนดอนออกเวลา 4:30 สามารถเข้าใจข้อมูลสำคัญในประกาศ ถ้าผู้พูดพูดอย่างชัดเจน เช่น การรายงานสภาพอากาศ ฯลฯ</p> <p>I can understand simple information and questions about family, people, homes, work and hobbies. I can understand what people say to me in simple, everyday conversation, if they speak clearly and slowly and give me help. I can understand short conversations about family, hobbies and daily life, provided that people speak slowly and clearly. I can follow changes of topic in TV news reports and understand the main information. I can understand short, clear and simple messages at the airport, railway station etc. For example: "The train to London leaves at 4:30". I can understand the main information in announcements if people talk very clearly. For example: weather reports, etc.</p>
ระดับ B1	<p>สามารถเข้าใจประเด็นสำคัญ ๆ ในการพูดแบบมาตรฐาน ซึ่งเกี่ยวข้องกับประเด็นที่คุ้นเคยในชีวิตประจำวัน เมื่อผู้พูดมีการกล่าวซ้ำหรืออธิบายเพิ่มเติมเพื่อให้เกิดความกระจ่างในบางครั้ง สามารถเข้าใจเรื่องที่เกี่ยวข้องกับการสนทนาในชีวิตประจำวัน แต่บางครั้งอาจต้องการความช่วยเหลือเพื่อความชัดเจนในรายละเอียดเฉพาะ สามารถเข้าใจประเด็นสำคัญของเรื่องที่พูดคุยเกี่ยวกับหัวข้อที่คุ้นเคยในสถานการณ์ประจำวัน เมื่อผู้พูดพูดอย่างชัดเจน แต่บางครั้งอาจต้องการความช่วยเหลือในการทำความเข้าใจรายละเอียดปลีกย่อยต่างๆ สามารถเข้าใจการพูดคุยสั้น ๆ</p>

ตารางที่ 2 คำอธิบายคุณลักษณะตามกรอบ CEFR จำแนกตามทักษะการฟัง และการอ่าน (ต่อ)

ทักษะ (skill)	คำอธิบายคุณลักษณะ (descriptors)
การฟัง	
ระดับ B1	<p>และตรงไปตรงมาที่ผู้พูดพูดอย่างชัดเจน ในเรื่องที่คุณเคย สามารถเข้าใจประเด็นสำคัญในรายการทีวี ในเรื่องที่คุณเคย เมื่อมีการนำเสนออย่างช้า ๆ และชัดเจน สามารถเข้าใจข้อมูลทางเทคนิคง่าย ๆ เช่น คู่มือการใช้งานสำหรับอุปกรณ์ประเภทที่คุณเคย</p> <p>I can understand the main points of clear standard speech on familiar, everyday subjects, provided there is an opportunity to get repetition or clarification sometimes. I can understand what is said to me in everyday conversations, but I sometimes need help in clarifying particular details. I can understand the main points of discussion on familiar topics in everyday situations when people speak clearly, but I sometimes need help in understanding details. I can follow clearly spoken, straightforward short talks on familiar topics. I can understand the main points in TV programmes on familiar topics when the delivery is relatively slow and clear. I can understand simple technical information, such as operating instructions for familiar types of equipment.</p>
ระดับ B2	<p>สามารถเข้าใจแนวคิดสำคัญของการพูดที่ซับซ้อน ในเรื่องที่เป็นรูปธรรมและนามธรรมที่สื่อสารด้วยภาษามาตรฐาน รวมถึงการอภิปรายทางเทคนิคที่ใช้ศัพท์เฉพาะในศาสตร์ที่ผู้ฟังมีความเชี่ยวชาญ สามารถเข้าใจรายละเอียดในเรื่องที่พูดด้วยภาษามาตรฐาน สามารถเข้าใจเรื่องที่พูดซึ่งเป็นเรื่องไกล ๆ ตัวที่อาจต้องใช้ความพยายามบ้าง แต่อาจยากที่จะเข้าใจการพูดคุยระหว่างผู้พูดหลายคนที่ไม่ได้ปรับเปลี่ยนภาษาของตนให้ง่ายลง สามารถเข้าใจสาระสำคัญในการบรรยาย การพูดคุย รายงาน และรูปแบบอื่น ๆ ของการนำเสนอทางวิชาการหรือวิชาชีพที่ซับซ้อนในศาสตร์ที่เกี่ยวข้องกับผู้ฟัง สามารถเข้าใจละครโทรทัศน์และภาพยนตร์เมื่อสื่อสารด้วยภาษาถิ่นที่ได้มาตรฐาน สามารถเข้าใจข่าวทีวี เหตุการณ์ปัจจุบัน สารคดี บทสัมภาษณ์ รายการทอล์คโชว์ ฯลฯ สามารถเข้าใจประกาศ และข้อความที่ผู้พูดพูดในภาษาถิ่นที่ได้มาตรฐานด้วยความเร็วปกติ</p> <p>I can understand the main ideas of complex speech on concrete and abstract topics delivered in a standard dialect, including technical discussions in my field of specialisation. I can understand in detail what is said to me in standard spoken language. I can with some effort catch much of what is said around me, but may find it difficult to understand a discussion between several speakers who do not modify their language in any way. I can follow the essentials of lectures, talks and reports and other forms of complex academic or professional presentation in my field. I can follow TV drama and the majority of films in standard dialect. I can understand TV news, current affairs, documentaries, interviews, talk shows, etc. I can understand announcements and messages spoken in standard dialect at normal speed.</p>

ตารางที่ 2 คำอธิบายคุณลักษณะตามกรอบ CEFR จำแนกตามทักษะการฟัง และการอ่าน (ต่อ)

ทักษะ (skill)	คำอธิบายคุณลักษณะ (descriptors)
การอ่าน	
ระดับ A2	<p>สามารถเข้าใจข้อความง่าย ๆ สั้น ๆ ที่มีคำศัพท์ที่คุ้นเคย รวมถึงคำศัพท์ที่มาจากภาษาต่างประเทศ สามารถหาข้อมูลสำคัญที่ปรากฏในโฆษณา แผ่นพับ หน้าเว็บ แคตตาล็อก ตารางเวลา เป็นต้น สามารถเข้าใจสาระสำคัญในรายการข่าวง่าย ๆ และสั้น ๆ ถ้ามีความรู้บางอย่างเกี่ยวกับเรื่องนั้น เช่น ข่าวเกี่ยวกับกีฬา หรือบุคคลที่มีชื่อเสียง สามารถเข้าใจวิธีการใช้หรือคำสั่งที่ชัดเจน เช่น วิธีใช้ โทรศัพท์ เครื่องกดเงินสด หรือเครื่องทำเครื่องดื่ม สามารถเข้าใจสาระสำคัญในเรื่องราวง่าย ๆ สั้น ๆ ในชีวิตประจำวันโดยเฉพาะอย่างยิ่งหากมีการนำเสนอด้วยภาพ สามารถเข้าใจข้อความง่าย ๆ สั้น ๆ จากเพื่อน เช่น อีเมล แชทบนเว็บ โปสการ์ด หรือจดหมายสั้น ๆ</p> <p>I can understand short, simple texts containing familiar vocabulary including international words. I can find the most important information in advertisements, information leaflets, webpages, catalogues, timetables etc. I can understand the main points in short, simple news items and descriptions if I already know something about the subject. For example: news about sport or famous people. I can understand clear instructions. For example: how to use a telephone, a cash machine or a drinks machine. I can understand the main points in short, simple, everyday stories, especially if there is visual support. I can understand short simple messages from friends. For example: e-mails, web chats, postcards or short letters.</p>
ระดับ B1	<p>สามารถเข้าใจสาระสำคัญในข้อความ หรือตัวบทที่เป็นข้อเท็จจริง ตรงไปตรงมา ในเรื่องที่เป็นความสนใจส่วนบุคคลหรืออาชีพได้ดี พอที่จะพูดถึงในภายหลังได้ สามารถหาและทำความเข้าใจข้อมูลที่ต้องการในโบรชัวร์ แผ่นพับ และข้อความสั้น ๆ อื่น ๆ ที่เกี่ยวข้องกับความสนใจของบุคคล สามารถเข้าใจสาระสำคัญในบทความสั้น ๆ ที่อยู่บนหนังสือพิมพ์ และนิตยสาร โดยเป็นเรื่องเกี่ยวกับปัจจุบันและเรื่องที่คุ้นเคย สามารถทำตาม เข้าใจคำสั่งที่ง่าย ๆ เช่น การเล่นเกม โดยใช้อุปกรณ์ที่คุ้นเคย หรือการทำอาหาร สามารถเข้าใจเรื่องราวในรูปแบบที่ง่าย ๆ ของนวนิยาย และสามารถติดตามเนื้อเรื่องในเรื่องสั้นที่มีโครงสร้างชัดเจน โดยอาจต้องใช้ความพยายาม และใช้พจนานุกรมเป็นประจำ สามารถเข้าใจจดหมายส่วนตัวเกี่ยวกับเหตุการณ์ ความรู้สึก และความปรารถนาได้ดี พอที่จะเขียนตอบกลับ</p> <p>I can understand the main points in straightforward factual texts on subjects of personal or professional interest well enough to talk about them afterwards. I can find and understand the information I need in brochures, leaflets and other short texts relating to my interests. I can understand the main points in short newspaper and magazine articles about current and familiar topics.</p>

ตารางที่ 2 คำอธิบายคุณลักษณะตามกรอบ CEFR จำแนกตามทักษะการฟัง และการอ่าน (ต่อ)

ทักษะ (skill)	คำอธิบายคุณลักษณะ (descriptors)
การอ่าน	
ระดับ B1 (ต่อ)	<p>I can follow simple instructions, for example for a game, using familiar types of equipment or cooking a meal. I can understand simplified versions of novels, and follow the story line in short stories with a clear structure, with some effort and regular use of a dictionary. I can understand private letters about events, feelings and wishes well enough to write back.</p>
ระดับ B2	<p>สามารถอ่านเรื่องได้อย่างอิสระเป็นอย่างมาก โดยใช้พจนานุกรมและแหล่งข้อมูลอ้างอิงอื่น ๆ เมื่อจำเป็น สามารถเข้าใจเนื้อหาและสาระสำคัญของข่าว บทความ และรายงานในหัวข้อที่เกี่ยวข้องกับความสนใจหรืองานได้อย่างรวดเร็ว และตัดสินใจว่าควรจะอ่านแบบเจาะลึกด้วยหรือไม่ สามารถเข้าใจบทความ รายงาน และบทวิจารณ์ ที่ผู้เขียนแสดงมุมมองเฉพาะ เช่น การวิจารณ์ทางการเมือง การวิพากษ์วิจารณ์นิตยสาร บทละคร ภาพยนตร์ เป็นต้น สามารถเข้าใจคำแนะนำในการใช้คู่มือที่ยาว ๆ เช่น อ่านคู่มือการใช้งานทีวีหรือกล้องดิจิทัลสำหรับติดตั้งซอฟต์แวร์ที่อ่านส่วนที่ยากได้ สามารถอ่านเรื่องสั้นและนวนิยายที่เขียนด้วยภาษาและลีลาที่ตรงไปตรงมา โดยใช้พจนานุกรม หากคุ้นเคยกับเรื่องราว และ/หรือผู้เขียน สามารถเข้าใจสาระสำคัญในจดหมายที่เป็นทางการ และไม่เป็นทางการที่เกี่ยวข้องกับความสนใจส่วนตัวและอาชีพ โดยใช้พจนานุกรมเป็นบางครั้ง</p> <p>I can read with a large degree of independence, using dictionaries and other reference sources selectively when necessary. I can rapidly grasp the content and the significance of news, articles and reports on topics connected with my interests or my job, and decide if a closer reading is worthwhile. I can understand articles, reports and reviews in which the writers express specific points of view (e.g., political commentary, critiques of exhibitions, plays, films, etc). I can understand lengthy instructions, for example in a user manual for a TV or digital camera, for installing software, as long as I can reread difficult sections. I can read short stories and novels written in a straightforward language and style, making use of a dictionary, if I am familiar with the story and/or the writer. I can understand the main points in formal and informal letters relating to my personal and professional interests, with occasional use of a dictionary.</p>

ที่มา: EAQUALS (2008)

สำหรับคำบรรยายคุณลักษณะของทักษะโครงสร้างภาษาใน EAQUALS Bank of Descriptors ไม่มีการกำหนดไว้อย่างชัดเจน เนื่องจากการวัดระดับทักษะภาษาอังกฤษตามกรอบ CEFR มุ่งเน้นการวัดความสามารถทางภาษา 4 ทักษะ ได้แก่ การฟัง การอ่าน การพูด และการเขียน อย่างไรก็ตาม ผู้วิจัยได้นำ Core Inventory for General English ในส่วนของ Grammar ซึ่งเป็นแนวทางสำหรับการสอนภาษาอังกฤษตามกรอบ CEFR ที่พัฒนาโดย British Council and EAQUALS (2015) มาใช้เป็นกรอบแนวทางในการพัฒนาแบบประเมินความสอดคล้องทักษะโครงสร้างภาษารายละเอียดแสดงดังตารางที่ 3

ตารางที่ 3 Core Inventory for General English ในส่วนของ Grammar

ทักษะ (skill)	Grammar
โครงสร้างภาษา	
ระดับ A2	Adjectives – comparative, – use of than and definite article Adjectives – superlative – use of definite article Adverbial phrases of time, place and frequency – including word order Adverbs of frequency Articles – with countable and uncountable nouns Countables and Uncountables: much/many Future Time (will and going to) Gerunds Going to Imperatives Modals – can/could Modals – have to Modals – should Past continuous Past simple Phrasal verbs – common Possessives – use of ‘s, s’ Prepositional phrases (place, time and movement) Prepositions of time: on/in/at Present continuous Present continuous for future Present perfect Questions Verb + ing/infinitive: like/ want-would like Wh-questions in past Zero and 1st conditional

ตารางที่ 3 Core Inventory for General English ในส่วนของ Grammar (ต่อ)

ทักษะ (skill)	Grammar
โครงสร้างภาษา	
ระดับ B1	Adverbs Broader range of intensifiers such as too, enough Comparatives and superlatives Complex question tags Conditionals, 2nd and 3rd Connecting words expressing cause and effect, contrast etc. Future continuous Modals - must/can't deduction Modals – might, may, will, probably Modals – should have/might have/etc Modals: must/have to Past continuous Past perfect Past simple Past tense responses Phrasal verbs, extended Present perfect continuous Present perfect/past simple Reported speech (range of tenses) Simple passive Wh- questions in the past Will and going to, for prediction

ตารางที่ 3 Core Inventory for General English ในส่วนของ Grammar (ต่อ)

ทักษะ (skill)	Grammar
โครงสร้างภาษา	
ระดับ B2	Adjectives and adverbs Future continuous Future perfect Future perfect continuous Mixed conditionals Modals – can't have, needn't have Modals of deduction and speculation Narrative tenses Passives Past perfect Past perfect continuous Phrasal verbs, extended Relative clauses Reported speech Will and going to, for prediction Wish Would expressing habits, in the past

ที่มา: British Council and EAQUALS (2015)

2.1.2 กรอบ CEFR กับการจัดการเรียนการสอนภาษาอังกฤษของประเทศไทย

กรอบ CEFR เป็นกรอบที่ใช้ประเมินความสามารถทางภาษาที่สหภาพยุโรปจัดทำขึ้น โดยมีวัตถุประสงค์เพื่อใช้เป็นแนวทางในการจัดการเรียนรู้ การสอน และการประเมินภาษาที่สองหรือภาษาต่างประเทศ กรอบ CEFR ถือได้ว่าเป็นกรอบมาตรฐานทางภาษาที่ได้รับการยอมรับจากหลากหลายประเทศ จึงทำให้ประเทศไทยเล็งเห็นถึงความสำคัญของการนำกรอบ CEFR มาใช้ในปี พ.ศ. 2557 กระทรวงศึกษาธิการมีนโยบายเร่งปฏิรูปการเรียนรู้ทั้งระบบให้สัมพันธ์เชื่อมโยงกัน เพื่อยกระดับคุณภาพการศึกษาและพัฒนาศักยภาพของผู้เรียน โดยเฉพาะการสร้างเสริมทักษะการใช้ภาษาอังกฤษให้ผู้เรียนสามารถใช้ภาษาอังกฤษเพื่อการสื่อสารและใช้เป็นเครื่องมือในการแสวงหาองค์ความรู้เพื่อการพัฒนาตนเอง อันจะนำไปสู่การเพิ่มขีดความสามารถในการแข่งขันของประเทศ และการเตรียมความพร้อมรองรับการเข้าสู่ประชาคมอาเซียน ดังนั้น เพื่อให้บรรลุเป้าหมายดังกล่าว กระทรวงศึกษาธิการจึงมีการกำหนดนโยบายปฏิรูปการเรียนการสอนภาษาอังกฤษในระดับการศึกษา

ขั้นพื้นฐาน โดยใช้กรอบอ้างอิงความสามารถทางภาษาอังกฤษที่เป็นสากล คือ CEFR เป็นกรอบความคิดหลักในการจัดการเรียนการสอนภาษาอังกฤษของประเทศไทย ทั้งในการออกแบบหลักสูตร การพัฒนาการเรียนการสอน การทดสอบ การวัดผล การพัฒนาครู รวมถึงการกำหนดเป้าหมายการเรียนรู้

การนำกรอบ CEFR มาใช้ในการปฏิรูปการเรียนการสอนภาษาอังกฤษสำหรับการศึกษาระดับขั้นพื้นฐานนั้น กระทรวงศึกษาธิการได้กำหนดแนวทางการดำเนินการดังนี้ (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2557)

1. ใช้ CEFR เป็นกรอบความคิดหลักในการกำหนดเป้าหมายการจัดการเรียนรู้หรือพัฒนาในเบื้องต้น กระทรวงศึกษาธิการกำหนดเป้าหมายการพัฒนาระดับความสามารถทางภาษาอังกฤษของผู้เรียนไว้ คือ ผู้สำเร็จการศึกษาระดับประถมศึกษา (ป.6) ต้องมีระดับความสามารถทางภาษาอังกฤษตามกรอบ CEFR ที่ระดับ A1 ผู้สำเร็จการศึกษาภาคบังคับ (ม.3) ต้องมีระดับความสามารถทางภาษาอังกฤษตามกรอบ CEFR ที่ระดับ A2 และผู้สำเร็จการศึกษาระดับขั้นพื้นฐาน (ม.6 หรือ ปวช.) ต้องมีระดับความสามารถทางภาษาอังกฤษตามกรอบ CEFR ที่ระดับ B1 ในการประเมินหรือตรวจสอบผลการจัดการศึกษา หรือผลการพัฒนาผู้เรียนในแต่ละระดับข้างต้น หน่วยงานที่รับผิดชอบในแต่ละระดับควรได้มีการทดสอบหรือวัดผล โดยใช้แบบทดสอบมาตรฐานที่เทียบเคียงผลคะแนนกับระดับความสามารถทางภาษาตามกรอบ CEFR เพื่อตรวจสอบว่าผู้เรียนมีผลสัมฤทธิ์การเรียนภาษาผ่านเกณฑ์ระดับความสามารถที่กำหนดหรือไม่

2. ใช้พัฒนาหลักสูตรการเรียนการสอนภาษาอังกฤษ โดยนำระดับความสามารถทางภาษาที่กรอบ CEFR มากำหนดเป้าหมายของหลักสูตร และใช้คำอธิบายความสามารถทางภาษาของระดับนั้น ๆ มากำหนดกรอบเนื้อหาสาระที่จะใช้ในการจัดการเรียนการสอนตามหลักสูตร ดังคำอธิบายความสามารถในตารางที่ 1

3. ใช้ในการจัดการเรียนการสอน โดยนำระดับความสามารถทางภาษาและคำอธิบายความสามารถทางภาษาที่กรอบ CEFR กำหนดไว้แต่ละระดับ มาพิจารณาจัดกระบวนการเรียนรู้ เพื่อให้ผู้เรียนสามารถแสดงออก ซึ่งทักษะทางภาษาและองค์ความรู้ตามที่ระบุไว้ เช่น ในระดับ A1 ผู้สอนต้องจัดกระบวนการเรียนรู้เพื่อให้ผู้เรียนสามารถเข้าใจ ใช้ภาษา แนะนำ ถาม-ตอบ ปฏิสัมพันธ์ พูดคุยในเรื่องที่กำหนดไว้ในคำอธิบายดังตารางที่ 1 ได้ การเรียนการสอนจึงต้องเน้นให้ผู้เรียนได้ฟัง และพูดสื่อสารเป็นหลัก ผู้เรียนจึงจะมีความสามารถตามที่กำหนด

4. ใช้ในการทดสอบ และการวัดผล โดยใช้แบบทดสอบหรือแบบวัดที่สามารถเทียบเคียงผลได้กับกรอบ CEFR เพื่อให้ได้ข้อมูลระดับความสามารถของผู้เรียนหรือผู้เข้ารับการทดสอบ เพื่อการจัดกระบวนการเรียนการสอน กิจกรรม หรือสื่อให้สอดคล้องกับความต้องการจำเป็น

อันจะนำไปสู่การพัฒนาผู้เรียนหรือผู้เข้ารับการทดสอบให้มีความสามารถตามเป้าหมายหรือเกณฑ์ที่กำหนด

5. ใช้ในการพัฒนาครู โดยดำเนินการดังต่อไปนี้ (1) ใช้เครื่องมือในการประเมินตนเอง (self- assessment checklist) ตามกรอบ CEFR เพื่อเตรียมความพร้อมก่อนเข้ารับการทดสอบ และประเมินความก้าวหน้า ความสามารถทางภาษาอังกฤษอย่างต่อเนื่อง (2) ประเมินความสามารถในการใช้ภาษาอังกฤษของครูก่อนการพัฒนา โดยใช้แบบทดสอบมาตรฐานตามกรอบ CEFR ในการตรวจสอบระดับความสามารถของครู (3) จัดทำฐานข้อมูล และกลุ่มครูตามระดับเพื่อวางแผนพัฒนา และติดตามความก้าวหน้าในการเข้ารับการพัฒนาตามระดับความสามารถ (4) กำหนดเป้าหมาย ความสามารถด้านภาษาตามกรอบ CEFR การพัฒนาครูแต่ละกลุ่ม เพื่อนำมาจัดหลักสูตร และกระบวนการพัฒนาครูให้มีความสามารถในการใช้ภาษาผ่านเกณฑ์และบรรลุเป้าหมายที่กำหนด และ (5) ใช้แบบทดสอบมาตรฐานตามกรอบ CEFR ทดสอบหลังการพัฒนาเพื่อประเมินหลักสูตร การพัฒนากระบวนการพัฒนา และความสามารถของครูเทียบเคียงกับเป้าหมายที่กำหนด รวมทั้งจัดกิจกรรมการพัฒนาอย่างต่อเนื่อง

จากที่กล่าวมาแสดงให้เห็นว่า การศึกษาระดับขั้นพื้นฐานของประเทศไทย ได้ใช้กรอบ CEFR ในการพัฒนาผู้เรียน พัฒนาหลักสูตรการเรียนการสอนภาษาอังกฤษ การจัดการเรียนการสอน การทดสอบและการวัดผล ตลอดจนใช้ในการพัฒนาผู้สอนให้สามารถใช้ภาษาอังกฤษเพื่อการสื่อสาร และใช้เป็นเครื่องมือในการแสวงหาความรู้ อันจะนำไปสู่การเพิ่มขีดความสามารถในการแข่งขันของประเทศ

สำหรับการศึกษาระดับอุดมศึกษาประเทศไทยก็มีนโยบายการยกระดับมาตรฐานภาษาอังกฤษเช่นเดียวกัน โดยสำนักงานคณะกรรมการการอุดมศึกษาแห่งชาติ กระทรวงศึกษาธิการ ปัจจุบัน คือ สำนักงานปลัดกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม มีประกาศคณะกรรมการการอุดมศึกษา เรื่อง นโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา เป็นไปตามมติที่ประชุมคณะกรรมการการอุดมศึกษา ในการประชุมครั้งที่ 3/2559 เมื่อวันที่ 22 มีนาคม พ.ศ. 2559 ว่า

“ตามที่รัฐบาลมีนโยบายด้านการศึกษาและการเรียนรู้ โดยจัดให้มีการปฏิรูปการศึกษา และการเรียนรู้เพื่อสร้างคุณภาพของคนไทยให้สามารถเรียนรู้ พัฒนาการให้เต็มตามศักยภาพ และรัฐมนตรีว่าการกระทรวงศึกษาธิการมีนโยบายยกระดับมาตรฐานภาษาอังกฤษในทุกหลักสูตร เพื่อพัฒนาการเรียนการสอนภาษาอังกฤษ ให้ผู้เรียนมีความรู้ความสามารถใช้ภาษาอังกฤษเป็น เครื่องมือศึกษาค้นคว้าองค์ความรู้ที่เป็นสากลและก้าวหน้าทันโลก เพื่อรองรับการเข้าสู่ประชาคมอาเซียน และการเพิ่มขีดความสามารถในการแข่งขันประเทศ จึงกำหนดให้สถาบันอุดมศึกษาดำเนินการดังนี้

ข้อ 1 ให้กำหนดนโยบายและเป้าหมายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา ในทุกหลักสูตรและทุกระดับการศึกษา เพื่อเป็นแนวทางในการพัฒนาทักษะความสามารถใช้ภาษาอังกฤษของนิสิตนักศึกษาให้เป็นบัณฑิตที่มีความพร้อมทั้งวิชาการ วิชาชีพ และทักษะการสื่อสารภาษาอังกฤษในระดับที่ใช้งานได้ (working knowledge)

ข้อ 2 ให้จัดทำแผนเพื่อดำเนินการให้เป็นไปตามนโยบายและเป้าหมาย โดยมีตัวชี้วัด และมีการประเมินผลที่ชัดเจน

ข้อ 3 ให้พิจารณาปรับปรุงการจัดการเรียนการสอนในรายวิชาภาษาอังกฤษ โดยมุ่งผลสัมฤทธิ์ตามเป้าหมายที่กำหนด

ข้อ 4 ให้พิจารณาการจัดกิจกรรมนอกหลักสูตร กระบวนการ สื่อ และ/หรือสิ่งแวดล้อม ที่จะเปิดโอกาสและเสริมแรงจูงใจให้นิสิตนักศึกษาสามารถพัฒนาทักษะการใช้ภาษาอังกฤษได้ด้วยตนเอง

ข้อ 5 ให้สถาบันอุดมศึกษาพิจารณาจัดให้นิสิตนักศึกษาทุกคน ทดสอบความรู้ภาษาอังกฤษตามแบบทดสอบมาตรฐานระดับอุดมศึกษาที่สถาบันสร้างขึ้น หรือที่เห็นสมควรจะนำมาใช้วัดสมรรถภาพทางภาษาอังกฤษ โดยสามารถเทียบเคียงผลกับกรอบ CEFR หรือมาตรฐานอื่น เพื่อให้ทราบระดับความสามารถของนิสิตนักศึกษาแต่ละคน และสถาบันอุดมศึกษาอาจพิจารณานำผลการทดสอบความรู้ทางภาษาอังกฤษบันทึกในใบรับรองผลการศึกษาหรือจัดทำเป็นประกาศนียบัตร โดยเริ่มตั้งแต่ปีการศึกษา 2559 เป็นต้นไป” (สำนักงานคณะกรรมการการอุดมศึกษา, 2559)

จากประกาศฉบับดังกล่าวทำให้สถาบันอุดมศึกษาทุกแห่งในประเทศไทย รวมถึงมหาวิทยาลัยสุโขทัยธรรมมาธิราช ได้กำหนดแนวทางในการยกระดับมาตรฐานภาษาอังกฤษของนักศึกษาในทุกหลักสูตรทั้งระดับปริญญาตรี และระดับบัณฑิตศึกษา เพื่อพัฒนาการเรียนการสอนภาษาอังกฤษ ให้ผู้เรียนมีความรู้ความสามารถใช้ภาษาอังกฤษเป็นเครื่องมือในการศึกษาค้นคว้าองค์ความรู้ที่เป็นสากลและก้าวทันโลก เพื่อรองรับการเข้าสู่ประชาคมอาเซียน และเป็นการเพิ่มขีดความสามารถในการแข่งขันหรือเพิ่มคุณสมบัติของคนไทยให้สอดคล้องกับความต้องการของตลาดแรงงาน

ต่อมาในปี พ.ศ. 2567 คณะรัฐมนตรีในการประชุม เมื่อวันที่ 28 พฤศจิกายน พ.ศ. 2566 ได้มอบหมายให้กระทรวงศึกษาธิการเป็นหน่วยงานหลักร่วมกับกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม และหน่วยงานที่เกี่ยวข้องของพิจารณากำหนดมาตรการหรือโครงการในการจัดการเรียนการสอนภาษาอังกฤษในทุกระดับการศึกษา เพื่อแก้ไขปัญหาการเรียนรู้อะไรและความสามารถในการใช้ภาษาอังกฤษของคนไทย ดังนั้น เพื่อเป็นการยกระดับความรู้ความสามารถในการใช้ภาษาอังกฤษของผู้สำเร็จการศึกษาระดับอุดมศึกษาให้ดีขึ้นได้อย่างเป็นรูปธรรม เป็นไปตามนโยบายการเสริมสร้างความรู้ความสามารถทางการใช้ภาษาอังกฤษของไทยตามมติคณะรัฐมนตรี

จึงจำเป็นต้องออกประกาศ เรื่อง นโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา พ.ศ. 2567 ซึ่งประกาศเมื่อวันที่ 17 มกราคม พ.ศ. 2567 สรุปใจความสำคัญได้ดังนี้

“ข้อ 4 ให้สถาบันอุดมศึกษากำหนดนโยบายและเป้าหมายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา เพื่อเป็นแนวทางในการพัฒนาทักษะความสามารถการใช้ภาษาอังกฤษของนิสิตนักศึกษาให้เป็นบัณฑิตที่มีความพร้อมทั้งทางด้านวิชาการ วิชาชีพ และทักษะการสื่อสารภาษาอังกฤษในระดับที่ใช้งานได้

ข้อ 5 ให้สถาบันอุดมศึกษาจัดทำแผนเพื่อดำเนินการให้เป็นไปตามนโยบายและเป้าหมายโดยมีตัวชี้วัดและมีการประเมินผลที่ชัดเจน

ข้อ 6 ให้สถาบันอุดมศึกษาพิจารณาปรับปรุงการจัดการเรียนการสอนในรายวิชาภาษาอังกฤษโดยมุ่งผลสัมฤทธิ์ตามเป้าหมายที่กำหนด

ข้อ 7 ให้สถาบันอุดมศึกษาพิจารณาจัดกิจกรรมเสริมหลักสูตร จัดสภาพแวดล้อมทั้งในห้องเรียนของรายวิชาต่าง ๆ และนอกห้องเรียน ที่ส่งเสริมการใช้ภาษาอังกฤษ ตลอดจนจัดให้มีสื่อการเรียนรู้ที่นิสิตนักศึกษาสามารถเข้าถึงได้โดยอาจเป็นรูปแบบเอกสารหรือสื่อการเรียนรู้ออนไลน์ (โปรแกรมภาษาอังกฤษออนไลน์) เพื่อเปิดโอกาสและสร้างแรงจูงใจให้นิสิตนักศึกษาสามารถพัฒนาทักษะการใช้ภาษาอังกฤษได้ด้วยตนเอง

ข้อ 8 ให้สถาบันอุดมศึกษาพิจารณาจัดให้นิสิตนักศึกษาทุกคน สอบวัดความรู้และทักษะภาษาอังกฤษก่อนสำเร็จการศึกษา ตามแบบทดสอบที่สถาบันอุดมศึกษาพัฒนาขึ้นหรือแบบทดสอบตามมาตรฐานสากลอื่น ๆ โดยเทียบเคียงผลกับกรอบ CEFR เพื่อให้ทราบถึงผลสัมฤทธิ์ด้านทักษะภาษาอังกฤษของนิสิตนักศึกษาแต่ละคน และสถาบันอุดมศึกษาอาจพิจารณานำผลการทดสอบความรู้และทักษะภาษาอังกฤษบันทึกในใบรับรองผลการศึกษา หรือจัดทำเป็นประกาศนียบัตร ทั้งนี้ เป้าหมายผลสัมฤทธิ์ด้านทักษะภาษาอังกฤษ มีดังนี้

(1) ระดับอนุปริญญา ควรกำหนดเป้าหมายผลสัมฤทธิ์ด้านทักษะภาษาอังกฤษก่อนสำเร็จการศึกษา เทียบเคียงผลกับกรอบ CEFR ในระดับเกณฑ์ตั้งแต่ B1 ขึ้นไป

(2) ระดับปริญญาตรี ควรกำหนดเป้าหมายผลสัมฤทธิ์ด้านทักษะภาษาอังกฤษก่อนสำเร็จการศึกษา เทียบเคียงผลกับกรอบ CEFR ในระดับเกณฑ์ตั้งแต่ B2 ขึ้นไป

(3) ระดับบัณฑิตศึกษา ควรกำหนดเป้าหมายผลสัมฤทธิ์ด้านทักษะภาษาอังกฤษก่อนสำเร็จการศึกษา เทียบเคียงผลกับกรอบ CEFR ในระดับเกณฑ์ตั้งแต่ C1 ขึ้นไป” (คณะกรรมการมาตรฐานการอุดมศึกษา, 2567)

ทั้งนี้ เป้าหมายผลสัมฤทธิ์ดังกล่าว ไม่ได้เป็นเงื่อนไขในการสำเร็จการศึกษาของนิสิตนักศึกษา แต่จะเป็นแนวทางในการประเมินแผนดำเนินการของสถาบันอุดมศึกษา

กล่าวโดยสรุป ประเทศไทยได้ให้ความสำคัญกับการจัดการเรียนการสอนภาษาอังกฤษ ทั้ง 2 ระดับ คือ การศึกษาขั้นพื้นฐาน และการศึกษาระดับอุดมศึกษา โดยใช้กรอบ CEFR เป็นกรอบความคิดหลักในการจัดการเรียนการสอนภาษาอังกฤษของประเทศไทย รวมทั้งรัฐบาลไทยต้องการยกระดับมาตรฐานภาษาอังกฤษ จึงมีนโยบายเร่งพัฒนาทักษะภาษาอังกฤษ โดยเฉพาะในระดับอุดมศึกษาให้ได้ตามมาตรฐานสากล เพื่อเตรียมความพร้อมของนิสิตนักศึกษาให้กลายเป็นทรัพยากรบุคคลที่มีความสามารถเป็นที่ต้องการของตลาดแรงงาน สอดคล้องกับนโยบายของประเทศไทยด้านการท่องเที่ยว การค้า การลงทุน รวมถึงภาคการศึกษาวิจัยต่าง ๆ

2.2 การสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราช

เนื้อหาในตอนนี้ แบ่งออกเป็น 3 หัวข้อ ได้แก่ 2.2.1 การยกระดับมาตรฐานภาษาอังกฤษของนักศึกษามหาวิทยาลัยสุโขทัยธรรมมาธิราช 2.2.2 ความเป็นมาของการสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราช และ 2.2.3 แบบทดสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราช รายละเอียดแต่ละหัวข้อมีดังนี้

2.2.1 การยกระดับมาตรฐานภาษาอังกฤษของนักศึกษามหาวิทยาลัยสุโขทัยธรรมมาธิราช

ตามที่รัฐมนตรีว่าการกระทรวงศึกษาธิการมีนโยบายยกระดับมาตรฐานภาษาอังกฤษในทุกหลักสูตรเพื่อพัฒนาการเรียนการสอนภาษาอังกฤษให้ผู้เรียนมีความรู้ความสามารถใช้ภาษาอังกฤษเป็นเครื่องมือศึกษาค้นคว้าองค์ความรู้ที่เป็นสากลและก้าวทันโลก คณะกรรมการการอุดมศึกษาจึงได้มีประกาศ เรื่อง นโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา ลงวันที่ 12 เมษายน พ.ศ. 2559 โดยให้สถาบันอุดมศึกษากำหนดนโยบายการยกระดับมาตรฐานภาษาอังกฤษในทุกหลักสูตร และทุกระดับการศึกษา เพื่อเป็นแนวทางในการพัฒนาทักษะการใช้ภาษาอังกฤษของนักศึกษาให้เป็นบัณฑิตที่มีความพร้อมทางวิชาการ วิชาชีพ และทักษะการสื่อสารภาษาอังกฤษในระดับที่ใช้งานได้ มหาวิทยาลัยสุโขทัยธรรมมาธิราช (มสธ.) จึงกำหนดแนวทางในการยกระดับมาตรฐานภาษาอังกฤษของนักศึกษาระดับปริญญาตรี ปริญญาโท และปริญญาเอก ดังนี้ (มหาวิทยาลัยสุโขทัยธรรมมาธิราช, 2565: ออนไลน์)

ระดับปริญญาตรี

มหาวิทยาลัยสุโขทัยธรรมมาธิราชให้ออกประกาศ เรื่อง การยกระดับมาตรฐานภาษาอังกฤษสำหรับนักศึกษาระดับปริญญาตรี สรุปประเด็นสำคัญดังนี้ (มหาวิทยาลัยสุโขทัยธรรมมาธิราช, 2562)

ข้อ 4 นักศึกษาทุกคนที่เข้าศึกษาตั้งแต่ ปีการศึกษา 2560 เป็นต้นไป ต้องเข้ารับการทดสอบความรู้ภาษาอังกฤษก่อนสำเร็จการศึกษา

นักศึกษาที่เข้าศึกษา ก่อน ปีการศึกษา 2560 หรือนักศึกษาที่ได้เปลี่ยนโครงสร้างหลักสูตร ปีการศึกษา 2560 เป็นต้นไป ไม่ต้องเข้ารับการทดสอบความรู้ภาษาอังกฤษก่อนสำเร็จการศึกษา โดยนักศึกษาสามารถเข้ารับการทดสอบความรู้ภาษาอังกฤษได้ตามความสมัครใจ

ข้อ 5 ให้นักศึกษาเข้ารับการทดสอบความรู้ภาษาอังกฤษ STOU-EPT ตามวันและเวลาที่มหาวิทยาลัยกำหนดและให้นักศึกษายื่นผลการทดสอบความรู้ภาษาอังกฤษ พร้อมหลักฐานที่เกี่ยวข้องให้สำนักทะเบียนและวัดผล ภายในวันสุดท้ายของวันสอบไล่ของภาคการศึกษาที่จะสำเร็จการศึกษา

กรณีนักศึกษาที่มีผลการทดสอบความรู้ภาษาอังกฤษจากสถาบันทดสอบภาษาอังกฤษที่ได้มาตรฐานเทียบเคียง CEFR หรือมาตรฐานอื่น เช่น TOEFL IELTS และ TOEIC หรือจากสถาบันทดสอบภาษาที่ได้มาตรฐาน หรือจากสถาบันอุดมศึกษาของรัฐสามารถนำมาใช้เป็นผลการทดสอบความรู้ภาษาอังกฤษได้ โดยให้นักศึกษายื่นผลการทดสอบความรู้ภาษาอังกฤษดังกล่าวพร้อมหลักฐานที่เกี่ยวข้องให้สำนักทะเบียนและวัดผล ภายในวันสุดท้ายของวันสอบไล่ของภาคการศึกษาที่จะสำเร็จการศึกษา ทั้งนี้ต้องเป็นผลการทดสอบที่สอบมาแล้วไม่เกินสองปีนับถึงวันที่ยื่นผลการทดสอบ

ข้อ 6 กรณีที่นักศึกษาได้ยื่นผลการทดสอบความรู้ภาษาอังกฤษของสถาบันทดสอบภาษาอังกฤษอื่นนอกเหนือจากที่กำหนดไว้ในข้อ 5 ต้องได้รับความเห็นชอบจากคณะกรรมการ

คณะกรรมการอาจพิจารณาแต่งตั้งคณะทำงานเพื่อทำหน้าที่พิจารณาให้ความเห็นชอบผลการทดสอบความรู้ภาษาอังกฤษของนักศึกษาที่ได้ยื่นไว้ตามวรรคหนึ่งแทนก็ได้

ในกรณีมีประเด็นข้อสงสัยหรือมีปัญหาในการพิจารณาผลการทดสอบความรู้ภาษาอังกฤษให้คณะทำงานตามวรรคสอง เสนอความเห็นต่อคณะกรรมการพิจารณาเป็นรายกรณี ความเห็นของคณะกรรมการถือเป็นที่สุด

ข้อ 7 มหาวิทยาลัยจะบันทึกผลการทดสอบความรู้ภาษาอังกฤษของนักศึกษาไว้ในใบรายงานผลการศึกษา (transcript) ว่าได้เข้ารับการทดสอบความรู้ภาษาอังกฤษแล้ว ทั้งนี้ผลการทดสอบความรู้ภาษาอังกฤษดังกล่าวไม่มีผลต่อการสำเร็จการศึกษา

ข้อ 8 นักศึกษาควรพัฒนาทักษะภาษาอังกฤษอย่างต่อเนื่องจากแหล่งเรียนรู้ที่มหาวิทยาลัยจัดให้ในรูปแบบบทเรียนอิเล็กทรอนิกส์ (e-Learning) หรือสื่ออื่น ๆ ตามความเหมาะสม เพื่อให้มีความพร้อมด้านภาษาอังกฤษในระดับที่ใช้งานได้

ระดับปริญญาโท

มหาวิทยาลัยสุโขทัยธรรมมาธิราชให้ออกประกาศ เรื่อง การยกระดับมาตรฐานความรู้ภาษาอังกฤษสำหรับนักศึกษาระดับปริญญาโท สรุปประเด็นได้ดังนี้ (มหาวิทยาลัยสุโขทัยธรรมมาธิราช, 2560)

ข้อ 1 ให้นักศึกษาระดับปริญญาโทที่เข้าศึกษาตั้งแต่ปีการศึกษา 2560 เป็นต้นไปต้องส่งผลการทดสอบความรู้ภาษาอังกฤษจากสถาบันทดสอบภาษาที่ได้มาตรฐาน หรือผลการทดสอบ STOU-EPT โดยสามารถส่งผลการทดสอบภาษาอังกฤษได้ภายในระยะเวลาที่ศึกษาตามหลักสูตร

ข้อ 2 กรณีที่ไม่มีผลทดสอบความรู้ภาษาอังกฤษจากสถาบันทดสอบภาษาที่ได้มาตรฐาน หรือผลการทดสอบ STOU-EPT สามารถเลือกเรียนชุดวิชา 11701 ภาษาอังกฤษสำหรับบัณฑิตศึกษา 1 หรือ

ข้อ 3 เข้ารับการอบรมเข้มพิเศษภาษาอังกฤษที่มหาวิทยาลัยสุโขทัยธรรมมาธิราชเป็นผู้จัด (intensive training english course)

ข้อ 4 ระดับคะแนนทดสอบความรู้ภาษาอังกฤษให้ถือว่าเป็นการยกระดับความรู้ภาษาอังกฤษของนักศึกษาระดับปริญญาโทตามเกณฑ์ของแต่ละสาขาวิชาที่กำหนด

เกณฑ์มาตรฐานความรู้ภาษาอังกฤษระดับปริญญาโท กรณีสอบ STOU-EPT ส่วนใหญ่สาขาวิชากำหนดให้น้อย 1 part ไม่ต่ำกว่า B1 จาก 3 part แต่มีบางหลักสูตรกำหนดให้แต่ละ part ไม่ต่ำกว่า B1 เป็นต้น

ระดับปริญญาเอก

มหาวิทยาลัยสุโขทัยธรรมมาธิราชให้ออกประกาศเรื่อง เกณฑ์มาตรฐานความรู้ภาษาอังกฤษระดับปริญญาเอก พ.ศ. 2561 สรุปประเด็นสำคัญได้ดังนี้ (มหาวิทยาลัยสุโขทัยธรรมมาธิราช, 2561)

ข้อ 1 นักศึกษาระดับปริญญาเอกต้องสอบผ่านความรู้ภาษาอังกฤษ และต้องมีคะแนนทดสอบความรู้ภาษาอังกฤษตามเกณฑ์ใดเกณฑ์หนึ่งดังนี้

(1) คะแนนผลการทดสอบความรู้ภาษาอังกฤษไม่เกิน 2 ปี นับแต่วันสอบของศูนย์ทดสอบทางภาษาดังนี้

ก. TOEFL (Computer based) ไม่ต่ำกว่า 173 คะแนน หรือ

ข. TOEFL (Internet based) ไม่ต่ำกว่า 62 คะแนน หรือ

ค. IELTS ไม่ต่ำกว่า 5.5 คะแนน หรือ

ง. TU-GET ไม่ต่ำกว่า 500 คะแนน หรือ

จ. CU-TEP ไม่ต่ำกว่า 70 คะแนน หรือ

ฉ. TOEIC ไม่ต่ำกว่า 650 คะแนน หรือ

ช. STOU-EPT ไม่ต่ำกว่า 500 คะแนน หรือ

ซ. STOU-EPT (E-Testing) แต่ละ Part ไม่ต่ำกว่าระดับ B2

สำหรับผลการทดสอบของศูนย์ทดสอบทางภาษาของสถาบันอื่น ๆ นอกจากที่กำหนดไว้ในวรรคหนึ่ง ต้องได้รับความเห็นชอบจากคณะกรรมการบัณฑิตศึกษามหาวิทยาลัยก่อน

ข้อ 2 คะแนนสอบรายวิชาภาษาอังกฤษสำหรับบัณฑิตศึกษา 2 ที่มีผลการสอบ ไม่เกิน 2 ปีนับตั้งแต่วันสอบ และได้คะแนนแต่ละ part จะต้องได้ไม่ต่ำกว่าระดับ B2

ข้อ 3 คะแนนการอบรมเข้มพิเศษ Intensive Training English Course ที่มีผลการอบรม ไม่เกิน 2 ปีนับตั้งแต่วันที่เข้ารับการอบรม และได้คะแนนการอบรม แต่ละ Part จะต้องได้ไม่ต่ำกว่าระดับ B2

นอกจากนั้น เพื่อเป็นการส่งเสริมการพัฒนาทักษะภาษาอังกฤษให้กับนักศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราชได้จัดเตรียมตัวอย่างข้อสอบวัดระดับความรู้ภาษาอังกฤษ STOU-EPT เพื่อให้นักศึกษาได้ทดสอบความรู้ก่อนที่จะทำการทดสอบจริง โดยเมื่อนักศึกษาเข้าศึกษาอย่างน้อย 10 โมดูล นักศึกษาจะมีสิทธิ์ลงทะเบียนทำแบบทดสอบจาก Grad Studies เพื่อประเมินตนเอง ทั้งนี้ไม่สามารถนำผลคะแนนจากการทดสอบนี้มายื่นก่อนสำเร็จการศึกษาเพื่อเพิ่มข้อความลงในใบรายงานผลการศึกษาว่านักศึกษาได้เข้ารับการทดสอบภาษาอังกฤษแล้ว

เมื่อนักศึกษาพร้อมเข้ารับการทดสอบภาษาอังกฤษมหาวิทยาลัยสุโขทัยธรรมาธิราชได้จัดให้มีการทดสอบความรู้ภาษาอังกฤษ STOU-EPT ดังนี้

1) กำหนดช่วงวันและเวลาที่มหาวิทยาลัยจัดสอบ ณ มหาวิทยาลัยสุโขทัยธรรมาธิราช หรือศูนย์วิทยพัฒนามหาวิทยาลัยสุโขทัยธรรมาธิราช ทั้ง 10 ศูนย์

2) การแจ้งผลการสอบข้อสอบความรู้ภาษาอังกฤษ STOU-EPT เมื่อนักศึกษาเข้าสอบเรียบร้อยแล้ว จะได้รับทราบผลการสอบเป็นรายบุคคล โดยกรรมการคุมสอบจะพิมพ์ใบรายงานผลและลงนามกำกับให้ โดยในใบรายงานผลจะแสดงให้เห็นว่าแต่ละส่วนของแบบทดสอบนักศึกษาได้รับผลสอบอยู่ในระดับใด และจะมีลายเซ็นผู้อำนวยการสำนักทะเบียนและวัดผลพิมพ์เป็นลายน้ำกำกับไว้

กล่าวโดยสรุป มหาวิทยาลัยสุโขทัยธรรมาธิราช กำหนดให้นักศึกษาระดับปริญญาตรีที่สมัครเข้าศึกษาตั้งแต่ภาคต้น ปีการศึกษา 2560 จะต้องผ่านการทดสอบความรู้ภาษาอังกฤษทุกคน ซึ่งผลการทดสอบฯ จะไม่จำกัดเกณฑ์มาตรฐาน/คะแนน/ระดับผลการทดสอบ ส่วนนักศึกษาระดับปริญญาโท และปริญญาเอก ต้องส่งผลการทดสอบความรู้ภาษาอังกฤษจากสถาบันทดสอบภาษาที่ได้มาตรฐาน หรือผลการทดสอบ STOU-EPT ตามเกณฑ์ที่แต่ละสาขาวิชากำหนด

2.2.2 ความเป็นมาของการสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราช

คณาจารย์สาขาวิชาศิลปศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช ได้พัฒนาแบบทดสอบ STOU-EPT เพื่อวัดความรู้ภาษาอังกฤษให้กับผู้ที่สมัครเข้าศึกษาต่อในระดับบัณฑิตศึกษาของมหาวิทยาลัยสุโขทัยธรรมาธิราช ตั้งแต่ปี พ.ศ. 2542 เป็นต้นมา มีรูปแบบการทดสอบแบบดั้งเดิม หรือที่ใช้กระดาษและดินสอ (paper and pencil test) ในการทดสอบ โดยแบบทดสอบวัดทักษะ 5 ด้าน ได้แก่ การสนทนา (dialogue) คำศัพท์ (vocabulary) โครงสร้างภาษา (structure) การเขียน (writing) และการอ่าน (reading) รวมข้อสอบทั้งหมด 100 ข้อ คะแนนเต็ม 1,000 คะแนน ต่อมาในปี

พ.ศ. 2553 คณาจารย์สาขาวิชาศิลปศาสตร์ได้พัฒนาแบบทดสอบวัดความรู้ภาษาอังกฤษ STOU-EPT โดยใช้กรอบ CEFR ซึ่งวัด 3 ทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน รวมข้อสอบทั้งสิ้น 100 ข้อ การพัฒนาแบบทดสอบ STOU-EPT ใหม่นี้ ทำให้รูปแบบการทดสอบแบบดั้งเดิมไม่รองรับทักษะการฟัง มหาวิทยาลัยสุโขทัยธรรมาธิราชจึงต้องปรับเปลี่ยนรูปแบบการทดสอบเป็นการทดสอบผ่านระบบอิเล็กทรอนิกส์ (สุภมาส อังศุโชติ และคณะ, 2562)

ในปี พ.ศ. 2556 สิริรัตน์ วิภาสศิลป์ และคณะ ได้พัฒนารูปแบบการบริหารการสอบวัดความรู้ภาษาอังกฤษ STOU-EPT ผ่านสื่ออิเล็กทรอนิกส์ ผลการศึกษา พบว่า (1) รูปแบบการบริหารการสอบความรู้ภาษาอังกฤษ STOU-EPT มี 3 วิธี ได้แก่ การสอบโดยใช้ CD-ROM เครื่องเล่น MP3 และคอมพิวเตอร์ (2) ผลการสังเคราะห์รูปแบบการบริหารการสอบความรู้ภาษาอังกฤษ STOU-EPT ด้วยคอมพิวเตอร์ โดยประยุกต์ใช้เครื่องมือของ Moodle (3) ผลการทดสอบรูปแบบการบริหารการสอบครั้งที่ 1 พบว่า เครื่องมือของ Moodle สามารถรองรับรูปแบบของข้อสอบ STOU-EPT แบบใหม่ได้ แต่พบปัญหาสำคัญ ได้แก่ ผู้เข้าสอบไม่เข้าใจคำอธิบายการเข้าสู่ระบบสอบ ข้อสอบ Part 1: การฟัง ไม่สามารถควบคุมจำนวนครั้งในการฟังคลิปเสียงตามเงื่อนไขที่กำหนด ข้อสอบ Part 3: การอ่าน ไม่สามารถแสดงข้อคำถามและเนื้อหาในหน้าจอเดียวกัน และ (4) ผลการประเมินรูปแบบการบริหารการสอบของผู้ทรงคุณวุฒิ ครั้งที่ 1 มีข้อเสนอแนะให้ปรับแก้วิธีทัศน์ให้กระชับและครอบคลุมเนื้อหาสำหรับระบบสอบให้ปรับแก้โดยแสดงข้อมูลบนหน้าจอให้มีเฉพาะที่จำเป็น แสดงเวลาที่เหลือในการทำข้อสอบ ผลการทดสอบรูปแบบการบริหารการสอบ ครั้งที่ 2 พบว่า สามารถควบคุมการฟังคลิปเสียงตามที่กำหนดได้ และผลการสัมมนาผู้ทรงคุณวุฒิ ครั้งที่ 2 มีข้อเสนอแนะให้ผู้เข้าสอบสามารถฟังคลิปเสียงได้ 1 หรือ 2 ครั้ง ตามความต้องการ ให้ผู้เข้าสอบมีโอกาสได้ศึกษาขั้นตอน และวิธีการทำข้อสอบ เพื่อสร้างความคุ้นเคยในการทำข้อสอบก่อนการสอบในวันจริง ดังนั้น รูปแบบการบริหารการสอบที่พัฒนาจากผลการทดสอบรูปแบบการบริหารการสอบ ครั้งที่ 1 และครั้งที่ 2 รวมทั้งข้อเสนอแนะของผู้ทรงคุณวุฒิ จึงได้นำมากำหนดวิธีการดำเนินการได้ 3 ขั้นตอน ได้แก่ ขั้นตอนที่ 1 การเตรียมการก่อนการทดสอบ ประกอบด้วย การรับสมัคร การเตรียมแบบทดสอบด้วยเครื่องมือของ Moodle การเตรียมห้องสอบที่มีอุปกรณ์คอมพิวเตอร์ และระบบรักษาความปลอดภัย ขั้นตอนที่ 2 การดำเนินการในวันสอบ โดยใช้ห้องปฏิบัติการคอมพิวเตอร์ของมหาวิทยาลัยเป็นห้องสอบ โดยใช้ระเบียบการสอบของมหาวิทยาลัย ผู้เข้าสอบสามารถทราบคะแนนในแต่ละ Part อย่างไม่เป็นทางการ และขั้นตอนที่ 3 การดำเนินการภายหลังการทดสอบ ประกอบด้วย การประมวลผล การประกาศผลสอบทางเว็บไซต์ของมหาวิทยาลัย และแจ้งผลสอบทางไปรษณีย์ (สิริรัตน์ วิภาสศิลป์ และคณะ, 2558)

จากการทดสอบรูปแบบการบริหารการสอบงานวิจัยข้างต้น ทำให้พบปัญหาเกี่ยวกับการบริหารการสอบ STOU-EPT บนระบบ Moodle เช่น ข้อสอบ Part 1: การฟัง ไม่สามารถควบคุมจำนวนครั้งในการฟังคลิปเสียงตามเงื่อนไขที่กำหนด ข้อสอบ Part 3: การอ่าน ไม่สามารถแสดงข้อ

คำถามและเนื้อหาในหน้าจอดีเหมือนกัน แบบทดสอบ STOU-EPT มีข้อสอบหลากหลายรูปแบบ และไม่มีระบบรับสมัครสอบ จึงทำให้ในปี พ.ศ. 2557 มหาวิทยาลัยสุโขทัยธรรมาธิราชได้มีโครงการจัดจ้างที่ปรึกษาพัฒนาระบบสอบตามความพร้อมของนักศึกษารายบุคคลด้วยคอมพิวเตอร์ ระยะที่ 2 จึงได้พัฒนาระบบการสอบ STOU-EPT ผ่านระบบอิเล็กทรอนิกส์เข้าไว้โครงการดังกล่าวด้วย

การพัฒนาระบบการสอบ STOU-EPT ผ่านระบบอิเล็กทรอนิกส์ ดำเนินการแล้วเสร็จในปี พ.ศ. 2559 และเริ่มดำเนินการสอบมาตั้งแต่ พ.ศ. 2560 จนถึงปัจจุบัน การสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราชผ่านระบบอิเล็กทรอนิกส์ หรือเรียกว่า STOU-EPT (E-Testing) เป็นการสอบที่วัดทักษะทางภาษาอังกฤษได้อย่างหลากหลาย ในรูปแบบมัลติมีเดีย ผู้สอบสามารถทราบผลได้ทันทีหลังสอบเสร็จ ด้วยการประมวลผลการสอบผ่านระบบคอมพิวเตอร์ โดยเกณฑ์การประเมินผลเป็นไปตามกรอบ CEFR ซึ่งได้รับการยอมรับอย่างกว้างขวางว่าเป็นมาตรฐานในการจัดลำดับความสามารถทางภาษาของแต่ละบุคคล โดยวัดความสามารถทางภาษาอังกฤษ 3 ด้าน ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน วัตถุประสงค์หลักของการจัดสอบ STOU-EPT เพื่อวัดความรู้ภาษาอังกฤษ และเพื่อให้ผู้เข้าสอบได้ทราบระดับความรู้ของตนเองอย่างรวดเร็ว ทั้งนี้ นำไปสู่การพัฒนาความสามารถทางภาษาอังกฤษในระดับที่สูงขึ้นได้อย่างมีประสิทธิภาพ โดยจัดสอบให้แก่ผู้ประสงค์สมัครเข้าศึกษาต่อระดับบัณฑิตศึกษา และนักศึกษาระดับปริญญาตรีของมหาวิทยาลัยสุโขทัยธรรมาธิราช และผู้สนใจทั่วไปที่ประสงค์จะทดสอบความรู้ภาษาอังกฤษ และใช้ผลการสอบเป็นหลักฐานอ้างอิงระดับความรู้ภาษาอังกฤษสำหรับการสมัครงาน สมัครเรียน ฯลฯ (ตามข้อกำหนดของหน่วยงานหรือสถาบันการศึกษานั้น ๆ)

ระบบการสอบวัดความรู้ภาษาอังกฤษ STOU-EPT ผ่านระบบอิเล็กทรอนิกส์ของมหาวิทยาลัยสุโขทัยธรรมาธิราช ประกอบด้วย 5 ระบบย่อย ได้แก่ ระบบลงทะเบียนสอบ ระบบคลังข้อสอบ ระบบจัดการการสอบ ระบบสอบ และระบบควบคุมการสอบ ซึ่งระบบดังกล่าวได้พัฒนาแล้วเสร็จ และทดลองใช้ในปี พ.ศ. 2559 และเริ่มดำเนินการจัดสอบมาตั้งแต่ปี พ.ศ. 2560 ตั้งแต่ใช้ระบบนี้สำนักทะเบียนและวัดผล ยังไม่ได้มีการประเมินระบบสอบ สุภมาส อังศุโชติ และคณะ (2562) จึงได้ทำการประเมินระบบการสอบวัดความรู้ภาษาอังกฤษ STOU-EPT ผ่านระบบอิเล็กทรอนิกส์ โดยมีวัตถุประสงค์เพื่อประเมินระบบการสอบวัดความรู้ภาษาอังกฤษ STOU-EPT ผ่านระบบอิเล็กทรอนิกส์ และนำผลการประเมินไปปรับปรุงระบบการสอบวัดความรู้ภาษาอังกฤษ STOU-EPT ผ่านระบบอิเล็กทรอนิกส์ กลุ่มตัวอย่างเป็นผู้แทนจากสำนักทะเบียนและวัดผล จำนวน 45 คน ผู้แทนจากศูนย์สารสนเทศ จำนวน 3 คน ผู้แทนจากศูนย์วิทย์พัฒนา มสธ. ลำปาง นครสวรรค์ อุตรธานี และนครศรีธรรมราช แห่งละ จำนวน 2 คน รวมเป็น 8 คน คณะที่ปรึกษาจากมหาวิทยาลัยเกษตรศาสตร์ จำนวน 6 คน อาจารย์วัดผลที่ใช้งานระบบคลังข้อสอบ STOU-EPT จำนวน 4 คน และเจ้าหน้าที่ศูนย์วิจัยและพัฒนาแบบทดสอบที่ใช้งานระบบคลังข้อสอบ STOU-EPT จำนวน 2 คน นักศึกษาที่เข้า

สอบวัดความรู้ภาษาอังกฤษ STOU-EPT จำนวน 131 คน เครื่องมือที่ใช้เป็นแบบสอบถาม และแบบสัมภาษณ์ และวิเคราะห์ข้อมูลโดยใช้สถิติพื้นฐาน และการวิเคราะห์เนื้อหา ผลการประเมิน พบว่าระบบลงทะเบียนสอบ ระบบคลังข้อสอบ ระบบจัดการการสอบ ระบบสอบ ระบบควบคุมการสอบ สามารถใช้งานได้สะดวก รวดเร็ว คู่มือมีความถูกต้อง เป็นไปตามขั้นตอนการปฏิบัติงาน อ่านง่าย สามารถทำตามได้ แต่มีข้อควรปรับปรุง ได้แก่ (1) มหาวิทยาลัยควรกำหนดค่าตอบแทนให้กับชาวต่างชาติที่บันทึกเสียงในส่วนการฟัง ในอัตราที่จูงใจ (2) เสียงสนทนาในการฟัง ควรใช้ชาวต่างชาติ เพื่อให้ข้อสอบมีมาตรฐาน (3) ควรปรับ DVD แนะนำการสอบในส่วนของการฟังเพื่อให้สอดคล้องกับข้อสอบ (4) มหาวิทยาลัยควรจัดหาจอคอมพิวเตอร์ wide screen เพื่อให้รองรับกับการปรับแอปพลิเคชันระบบสอบให้นำเสนอโจทย์ร่วมในส่วนการอ่าน แสดงบทอ่านในลักษณะอยู่ทางซ้ายของจอคอมพิวเตอร์ และโจทย์อยู่ทางขวา เนื่องจากบทอ่านยาว มีหลายหน้า การแบ่งครึ่งบนล่างทำให้ผู้สอบอ่านไม่สะดวก (5) อนุญาตให้ผู้เข้าสอบทดสอบ ส่วนที่ 1 การฟัง และ (6) ปรับช่วงเวลาการลงชื่อเข้าสอบและการตรวจบัตรเข้าสอบ เป็นหลังจากทำข้อสอบส่วนที่ 1 การฟัง เสร็จแล้ว และเก็บ e-ticket เพื่อป้องกันผู้เข้าสอบคัดลอกข้อสอบ ส่วนผลการประเมินที่นำมาปรับปรุงส่วนใหญ่เป็นการปรับปรุงระบบสอบ ประกอบด้วย การตรวจสอบการเชื่อมต่อของเครื่องคอมพิวเตอร์ทุกเครื่องกับระบบสอบ และปรับระดับเสียงของเครื่องคอมพิวเตอร์และที่หูฟังให้สามารถได้ยินเสียงชัดเจน การปรับ DVD แนะนำการสอบให้สั้นกระชับ ในส่วนของการฟังให้สอดคล้องกับข้อสอบ จัดทำสรุปคำแนะนำการสอบพิมพ์ใส่กระดาษ ติดที่โต๊ะคอมพิวเตอร์เพื่อไว้อ่านทบทวน เพิ่มปุ่มปลดล็อกข้อสอบ ส่วนการฟังให้สามารถกลับไปทำข้อสอบย้อนหลังได้เป็นรายบุคคล และปุ่มเพิ่มเวลาการทำข้อสอบ ส่วนการฟัง สำนักทะเบียนและวัดผลได้นำผลการวิจัยนี้มาใช้ในการปรับปรุงระบบการสอบวัดความรู้ภาษาอังกฤษ STOU-EPT ให้สมบูรณ์ยิ่งขึ้น

ข้อสอบวัดความรู้ภาษาอังกฤษ STOU-EPT ที่อยู่ในคลังข้อสอบถูกนำมาใช้ระยะหนึ่งแล้ว แต่หลังจากนำข้อสอบ STOU-EPT มาใช้ สำนักทะเบียนและวัดผล ยังไม่ได้มีการตรวจสอบคุณภาพข้อสอบ ศจี จิระโร และ สุภมาส อังศุโชติ (2563) จึงได้ทำการตรวจสอบคุณภาพข้อสอบวัดสมิทธิภาพในการใช้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมิกราช โดยมีวัตถุประสงค์เพื่อตรวจสอบคุณภาพข้อสอบ STOU-EPT ตามกรอบ CEFR ในประเด็นเกี่ยวกับกระบวนการสร้างแบบทดสอบ โครงสร้างแบบทดสอบแต่ละ part ตามเกณฑ์การสร้างข้อสอบของ CEFR ความยากง่ายและอำนาจของข้อสอบ สัมประสิทธิ์ความเที่ยงของแบบทดสอบแต่ละส่วน และเพื่อตรวจสอบความตรงตามเกณฑ์สัมพันธของแบบทดสอบ STOU-EPT กับแบบทดสอบที่ผ่านการเทียบระดับความสามารถตามเกณฑ์ CEFR แล้ว ข้อมูล มี 2 ประเภท ได้แก่ ข้อมูลปฐมภูมิ คือ ผู้เชี่ยวชาญที่มีประสบการณ์เกี่ยวกับการทดสอบภาษาอังกฤษตามกรอบ CEFR จำนวน 4 คน คณะกรรมการออกข้อสอบ จำนวน 1 คน อาสาสมัครเข้าสอบวัดความรู้ภาษาอังกฤษ STOU-EPT การสอบ SWU-SET และ IELTS รวม 30 คน

และข้อมูลทฤษฎี คือ ผลการตอบข้อสอบของผู้เข้าสอบวัดความรู้ภาษาอังกฤษ STOU-EPT ทั้ง 4 แบบ รวม 800 คน ผลการวิจัย พบว่า 1) การตรวจสอบคุณภาพข้อสอบ STOU-EPT ตามกรอบ CEFR ได้ผลการวิจัยดังนี้ (1) ขั้นตอนการพัฒนา แบบทดสอบมีการดำเนินการเริ่มจากมหาวิทยาลัยได้ส่งตัวแทนกรรมการออกข้อสอบ เข้ารับการฝึกอบรมการพัฒนาแบบทดสอบมาตรฐานภาษาอังกฤษ ณ สถาบันที่มีความเชี่ยวชาญ ในประเทศสหราชอาณาจักร และได้พัฒนากรอบของแบบทดสอบในแต่ละทักษะตามกรอบ CEFR โดยมีผู้เชี่ยวชาญ ซึ่งเป็นวิทยากรอบรมเป็นผู้ตรวจสอบและให้ข้อเสนอแนะ (2) การพิจารณาโครงสร้างข้อสอบได้ผ่านความเห็นชอบจากผู้เชี่ยวชาญ เพื่อให้สอดคล้องกับกรอบ CEFR และการดำเนินการของมหาวิทยาลัย การตรวจสอบคุณภาพของข้อสอบก่อนนำไปใช้ พิจารณาความถูกต้อง เหมาะสม และความสอดคล้องตามกรอบ CEFR โดยนักวิชาการที่มีชื่อเสียงด้านการวัด และประเมินผลทางภาษาของประเทศสหราชอาณาจักร การใช้โปรแกรม Lexitutor ตรวจสอบความเหมาะสมของระดับภาษา และ AWL list ตรวจสอบความเหมาะสมของการใช้คำศัพท์ทางวิชาการ ข้อสอบของแบบทดสอบ ฉบับที่ 1 ฉบับที่ 2 และฉบับที่ 3 ส่วนมากมีระดับสูงกว่าเกณฑ์ CEFR โดยฉบับที่ 2 มีจำนวนข้อสอบที่สูงกว่าระดับที่กำหนดมากที่สุด รองลงมา คือ ฉบับที่ 3 และฉบับที่ 1 ตามลำดับ ส่วนฉบับที่ 4 ข้อสอบส่วนมากสอดคล้องกับระดับที่กำหนด (3) ความยากและอำนาจจำแนกของแบบทดสอบ พบว่า แบบทดสอบฉบับที่ 1 ข้อสอบที่อยู่ในเกณฑ์ใช้ได้ 64 ข้อ มีค่าความยากระหว่าง .11 ถึง .79 และอำนาจจำแนกรายข้อระหว่าง -.35 ถึง .59 ฉบับที่ 2 มีค่าความยากระหว่าง .20 ถึง .91 และอำนาจจำแนกรายข้อระหว่าง -.12 ถึง .69 ข้อสอบที่อยู่ในเกณฑ์ใช้ได้ 77 ข้อ ฉบับที่ 3 มีค่าความยากระหว่าง .17 ถึง .94 และอำนาจจำแนกรายข้อระหว่าง -.09 ถึง .65 ฉบับที่ 4 มีค่าความยากระหว่าง .05 ถึง .79 และอำนาจจำแนกรายข้อระหว่าง -.15 ถึง .63 ข้อสอบที่อยู่ในเกณฑ์ใช้ได้ 73 ข้อ และ (4) ผลการวิเคราะห์สัมประสิทธิ์ความเที่ยงของแบบทดสอบทั้ง 4 ฉบับ โดยใช้สูตร KR-20 Part 3 การอ่าน ทั้ง 4 ฉบับ มีความเที่ยงสูงสุด ค่าความเที่ยงอยู่ระหว่าง .853 ถึง .893 โดยแบบทดสอบฉบับที่ 3 มีความเที่ยงสูงสุด เท่ากับ .893 ส่วน Part 2 การฟัง (Part I) มีค่าความเที่ยงต่ำสุด เท่ากับ .350 และ 2) การวิเคราะห์ความสัมพันธ์ระดับ CEFR ของ STOU-EPT ที่สัมพันธ์กับ IELTS และ SWU-SET โดยสูตร Kendall's tau-b พบว่า ระดับ CEFR ของ STOU-EPT สัมพันธ์กับ IELTS และ SWU-SET ทั้งทักษะการฟัง โครงสร้างภาษา และการอ่าน โดยระดับ CEFR ของ STOU-EPT มีความสัมพันธ์กับ IELTS และ SWU-SET ในระดับสูงทุกด้าน โดยด้านที่มีความสัมพันธ์สูงสุดทั้ง IELTS และ SWU-SET คือ ด้านการอ่าน เมื่อพิจารณาแต่ละการทดสอบ พบว่า ระดับ CEFR ของ STOU-EPT มีความสัมพันธ์กับ IELTS สูงกว่า SWU-SET ในทุกด้าน จากการวิเคราะห์ความตรงตามเกณฑ์สัมพันธ์ระหว่าง STOU-EPT และแบบทดสอบมาตรฐาน พบว่า ระดับ A2 มีความสอดคล้องกันน้อย โดย STOU-EPT ให้ระดับที่สูงกว่า ทั้งนี้อาจเกิดจากความคลาดเคลื่อนจากเกณฑ์การแปลความหมายคะแนน โดยเฉพาะช่วงคะแนนระดับ B1 ที่ค่อนข้างกว้างเกินไป

ต่อมา ศุภรัตน์ อิงชาติเจริญ (2567) ได้ตรวจสอบคุณสมบัติทางจิตมิติของข้อสอบ STOU-EPT โดยมีวัตถุประสงค์การวิจัย 2 ข้อ ได้แก่ 1) วิเคราะห์คุณภาพของ STOU-EPT ใน 3 พารามิเตอร์ ได้แก่ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าโอกาสการเดา (c) ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) และ 2) ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ STOU-EPT ผลการวิจัย พบว่า คุณภาพข้อสอบ STOU-EPT ทักษะการฟัง ค่า a อยู่ในช่วง -9.62 ถึง 4.50 ผ่านเกณฑ์ $a > 0$ จำนวน 22 ข้อ (คิดเป็น 88%) ค่า b อยู่ในช่วง -1.80 ถึง 8.19 ผ่านเกณฑ์ $-2 \leq b \leq +2$ จำนวน 22 ข้อ (คิดเป็น 88%) และค่า c อยู่ในช่วง 0 ถึง .61 ทักษะโครงสร้างไวยากรณ์ ค่า a อยู่ในช่วง -3.40 ถึง 7.65 ผ่านเกณฑ์ $a > 0$ จำนวน 31 ข้อ (คิดเป็น 88.57%) ค่า b อยู่ในช่วง -4.21 ถึง 3.06 ผ่านเกณฑ์ $-2 \leq b \leq +2$ จำนวน 22 ข้อ (คิดเป็น 62.86%) และค่า c อยู่ในช่วง 0 ถึง .44 และทักษะการอ่าน ค่า a อยู่ในช่วง -8.75 ถึง 7.87 ผ่านเกณฑ์ $a > 0$ จำนวน 36 ข้อ (คิดเป็น 90%) ค่า b อยู่ในช่วง -2.02 ถึง 23.75 ผ่านเกณฑ์ $-2 \leq b \leq +2$ จำนวน 35 ข้อ (คิดเป็น 87.50%) และค่า c อยู่ในช่วง -0 ถึง .63 และผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ STOU-EPT ด้วยวิธี LR, SIBTEST และ MH โดยเปรียบเทียบจำนวนข้อที่พบ DIF ตรงกัน ตั้งแต่ 2 วิธีขึ้นไป พบว่า ทักษะการฟังมี DIF ระหว่างเพศ (หญิง = focal) ภูมิภาค (กรุงเทพมหานคร และปริมณฑล = focal) และประสบการณ์ในการสอบ (> 1 ครั้ง = focal) จำนวนอย่างละ 3 ข้อ เท่ากัน (คิดเป็น 12%) ทักษะโครงสร้างไวยากรณ์ มี DIF ระหว่างเพศ ภูมิภาค และประสบการณ์ในการสอบ จำนวน 3, 5 และ 1 ข้อ ตามลำดับ (คิดเป็น 8.57%, 14.29%, และ 2.86%) และทักษะการอ่าน มี DIF ระหว่างเพศ ภูมิภาค และประสบการณ์ในการสอบ จำนวน 1, 3 และ 0 ข้อ ตามลำดับ (คิดเป็น 2.50%, 7.50%, และ 0%) ผู้พัฒนาแบบทดสอบ STOU-EPT ควรนำผลการวิจัยไปใช้ปรับปรุงคุณภาพข้อสอบให้มีมาตรฐานมากขึ้น โดยเฉพาะการปรับโครงสร้างของข้อสอบให้สอดคล้องกับหลักการพัฒนาแบบทดสอบสากล และลดปัจจัยที่อาจทำให้เกิด DIF เช่น คำศัพท์ที่เกี่ยวข้องกับเพศชายหรือหญิง ภาษาที่ใช้เขียนข้อสอบ ลักษณะเนื้อหาที่นำมาสร้างข้อสอบ

แม้จะมีการพัฒนาและประเมินคุณภาพข้อสอบ STOU-EPT มาอย่างต่อเนื่อง แต่ยังคงมีประเด็นที่ต้องศึกษาเพิ่มเติมเพื่อให้ข้อสอบมีความเป็นมาตรฐานมากขึ้น โดยเฉพาะการตรวจสอบความสอดคล้องระหว่างข้อสอบ STOU-EPT กับกรอบ CEFR รวมถึงการกำหนดคะแนนจุดตัด (cut scores) โดยใช้กระบวนการกำหนดมาตรฐาน (standard setting) ที่แม่นยำขึ้น เพื่อลดความคลาดเคลื่อนในการแปลผลคะแนนและให้แน่ใจว่าข้อสอบสามารถจำแนกผู้สอบได้อย่างถูกต้องและเป็นธรรม การวิจัยในครั้งนี้ จึงต้องมีการเทียบเคียงคะแนนแบบทดสอบ STOU-EPT กับกรอบ CEFR เพื่อเสริมสร้างความน่าเชื่อถือของผลการสอบ และสามารถใช้อ้างอิงระดับความสามารถทางภาษาอังกฤษได้อย่างแม่นยำยิ่งขึ้น

2.2.3 แบบทดสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราช

ในส่วนนี้นำเสนอกระบวนการพัฒนาแบบทดสอบ และโครงสร้างและการแปลผลคะแนนแบบทดสอบ STOU-EPT โดยมีรายละเอียดดังนี้

1) กระบวนการพัฒนาแบบทดสอบ

ในการพัฒนาแบบทดสอบ STOU-EPT มหาวิทยาลัยสุโขทัยธรรมมาธิราชได้ส่งรองศาสตราจารย์ ดร.อลิสวา วานิชดี อาจารย์ประจำแขนงวิชาภาษาอังกฤษ สาขาวิชาศิลปศาสตร์ เข้าฝึกอบรมการพัฒนาแบบสอบมาตรฐานภาษาอังกฤษ ณ Centre for Research in English Language Learning and Assessment University of Bedfordshire ในสหราชอาณาจักร โดยวิทยากรฝึกอบรม คือ Dr.Cyril Weir และคณะวิทยากร ซึ่งเป็นนักวิชาการที่มีชื่อเสียงด้านการวัดและประเมินผลของสหราชอาณาจักร โดยเนื้อหาการฝึกอบรมเน้นเกณฑ์มาตรฐานตามกรอบ CEFR และการพัฒนาแบบทดสอบมาตรฐานตามกรอบ CEFR ทุกทักษะ และการกำหนดเกณฑ์เทียบเคียงแบบทดสอบกับเกณฑ์ของกรอบ CEFR ระหว่างการอบรม รองศาสตราจารย์ ดร.อลิสวา วานิชดี ได้พัฒนาลักษณะเฉพาะของแบบทดสอบ (test specifications) ของมหาวิทยาลัยสุโขทัยธรรมมาธิราช ในแต่ละทักษะตามกรอบ CEFR โดยมีวิทยากรฝึกอบรมเป็นผู้ตรวจสอบและให้ข้อเสนอแนะในการพัฒนาลักษณะเฉพาะของแบบทดสอบ การพัฒนาข้อสอบ STOU-EPT ตามกรอบ CEFR ครอบคลุม 3 ทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน ส่วนทักษะการพูด สถาบันภาษาหลายแห่งไม่ได้วัดทักษะนี้ เนื่องจากมีข้อจำกัดในการตรวจให้คะแนน และค่าใช้จ่ายค่อนข้างสูง ดังนั้น กรรมการออกข้อสอบ STOU-EPT พิจารณาแล้วไม่ควรวัดทักษะการพูด

หลังจากฝึกอบรม รองศาสตราจารย์ ดร.อลิสวา วานิชดี ได้มาถ่ายทอดความรู้ให้กับอาจารย์แขนงวิชาภาษาอังกฤษ สาขาวิชาศิลปศาสตร์ มหาวิทยาลัยสุโขทัยธรรมมาธิราช และมีการแต่งตั้งคณะกรรมการออกข้อสอบ STOU-EPT โดยออกข้อสอบ จำนวน 4 แบบ แบบละ 100 ข้อ รวม 400 ข้อ ส่วนอีก 1 ฉบับ เป็นแบบทดสอบฉบับคูชานานกับแบบทดสอบทั้ง 4 ฉบับข้างต้น โดยกระบวนการออกข้อสอบทั้งหมดแล้วเสร็จในปี พ.ศ. 2553 (ศจี จิระโร และ สุภมาศ อังคุโชติ, 2563)

จุดประสงค์ของการพัฒนาแบบทดสอบ STOU-EPT มี 2 ประการ ดังนี้

1. เพื่อใช้เป็นเครื่องมือประเมินความสามารถทางภาษาอังกฤษในการสมัครเข้าศึกษาต่อในระดับปริญญาตรีและระดับบัณฑิตศึกษาของมหาวิทยาลัยสุโขทัยธรรมมาธิราช
2. เพื่อใช้เป็นหลักฐานแสดงความสามารถทางภาษาอังกฤษสำหรับบุคคลทั่วไปอันสามารถนำไปใช้ประกอบการสมัครงาน การสมัครเรียน หรือเพื่อวัตถุประสงค์อื่นที่เกี่ยวข้อง

ขั้นตอนการพัฒนาแบบทดสอบ STOU-EPT

แบบทดสอบ STOU-EPT ได้รับการพัฒนาโดยคณะกรรมการผู้เชี่ยวชาญของมหาวิทยาลัยสุโขทัยธรรมาราช โดยมีขั้นตอนสำคัญ 5 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การกำหนดกรอบและการสร้างข้อสอบ คณะกรรมการพัฒนาแบบทดสอบได้รับการอบรมการสร้างข้อสอบตามกรอบ CEFR จาก Dr. Cyril Weir และทีมงานจาก Centre for Research in English Language Learning and Assessment, University of Bedfordshire สหราชอาณาจักร โดยเน้นการออกแบบและทำลักษณะเฉพาะของแบบทดสอบ เพื่อใช้เป็นแนวทางในการออกข้อสอบ หลังจากนั้นจึงสร้างข้อสอบให้มีความถูกต้องทางภาษา เนื้อหาสอดคล้องกับลักษณะเฉพาะของแบบทดสอบ และเหมาะกับระดับความสามารถที่กำหนด

ขั้นตอนที่ 2 การตรวจสอบคุณภาพข้อสอบโดยผู้เชี่ยวชาญ ข้อสอบที่สร้างขึ้นจะตรวจสอบโดยผู้เชี่ยวชาญด้านภาษาและการวัดผล เพื่อประเมินความชัดเจน ความถูกต้องทางภาษา และความสอดคล้องกับระดับความสามารถตามกรอบ CEFR นอกจากนี้ ยังมีการใช้โปรแกรม Lextutor และ Academic Word List (AWL) เพื่อตรวจสอบความซับซ้อนของภาษาและความเหมาะสมของคำศัพท์

ขั้นตอนที่ 3 การทดลองใช้ข้อสอบ ข้อสอบที่ผ่านการตรวจสอบจะนำไปทดลองใช้ภายใต้เงื่อนไขมาตรฐานเพื่อเก็บข้อมูลเชิงประจักษ์และทำการวิเคราะห์คุณภาพข้อสอบรายข้อ โดยใช้ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) เพื่อคำนวณค่าความยาก (difficulty) และค่าจำแนก (discrimination) ของข้อสอบ

ขั้นตอนที่ 4 การปรับปรุงและแก้ไขข้อสอบ ข้อสอบที่มีคุณภาพต่ำหรือไม่เป็นไปตามเกณฑ์ที่กำหนดจะต้องได้รับการปรับปรุง แก้ไข หรือตัดออก โดยพิจารณาจากผลการวิเคราะห์เชิงสถิติและข้อเสนอแนะจากผู้เชี่ยวชาญ

ขั้นตอนที่ 5 การจัดเก็บข้อสอบและการรักษามาตรฐาน ข้อสอบที่ผ่านการรับรองคุณภาพจะถูกจัดเก็บในคลังข้อสอบที่มีระบบรักษาความปลอดภัย เพื่อใช้เป็นฐานข้อมูลสำหรับการจัดทำข้อสอบในอนาคต โดยมีการควบคุมคุณภาพเพื่อให้ข้อสอบที่ใช้ในแต่ละครั้งมีมาตรฐานเดียวกัน

2) โครงสร้างและการแปลผลคะแนนแบบทดสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาราช

2.1) การกำหนดโครงสร้างของแบบทดสอบ STOU-EPT

การกำหนดโครงสร้างของแบบทดสอบ STOU-EPT ให้น้ำหนักความสำคัญกับทักษะการอ่านมากที่สุด (มีข้อสอบ 40 ข้อ) เนื่องจากเป้าหมายหลักของกลุ่มผู้สอบเพื่อนำผลการสอบไปใช้ในการเข้าศึกษาต่อระดับบัณฑิตศึกษา ส่วนนักศึกษาระดับปริญญาตรี เป็นกลุ่มที่ใช้ทักษะการอ่านภาษาอังกฤษเพื่อการเรียนรู้เป็นหลัก รองลงมา คือ ทักษะโครงสร้างภาษา (มีข้อสอบ 35 ข้อ)

และทักษะที่ให้ความสำคัญน้อยที่สุด คือ ทักษะการฟัง (มีข้อสอบ 25 ข้อ) การสร้างข้อสอบและการกำหนดน้ำหนักคะแนนที่ใช้ในการกำหนดระดับความสามารถแต่ละทักษะเป็นไปตามแนวของ Cambridge Assessment English เป็นสถาบันทางภาษาของมหาวิทยาลัยเคมบริดจ์ที่ได้รับการยอมรับในระดับสากล (ศจี จิระโร และ สุภมาส อังศุโชติ, 2563)

2.2) โครงสร้างของแบบทดสอบ STOU-EPT

โครงสร้างของแบบทดสอบ STOU-EPT ประกอบด้วย 3 ทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน แสดงดังตารางที่ 4 (ศจี จิระโร และ สุภมาส อังศุโชติ, 2563)

ตารางที่ 4 โครงสร้างของแบบทดสอบ STOU-EPT

ทักษะ	รายละเอียดข้อสอบ	ลักษณะข้อสอบ	จำนวนข้อสอบ
Part 1: Listening (ข้อ 1-25 รวม 25 ข้อ)	Text 1	Shot Conversation โดยให้ฟังการสนทนา แล้วตอบข้อสอบแบบเลือกตอบ	8
	Text 2	ให้ฟังเสียงแล้วจับคู่กับสิ่งได้ฟัง	6
	Text 3	ให้ฟังเสียงแล้วตอบข้อสอบแบบเติมคำ	6
	Text 4	ให้ฟังเสียงแล้วตอบข้อสอบแบบถูก-ผิด	5
Part 2: Structure (ข้อ 26-60 รวม 35 ข้อ)		ข้อสอบแบบเลือกตอบ ให้เลือกคำตอบที่ถูกต้องที่สุด	35
Part 3: Reading (ข้อ 61-100 รวม 40 ข้อ)	Passage 1	เรียงลำดับเหตุการณ์ที่เกิดขึ้นก่อนหลัง	5
	Passage 2	ข้อสอบแบบเลือกตอบ	5
	Passage 3	ข้อสอบแบบเลือกตอบ	10
	Passage 4	ข้อสอบแบบเลือกตอบ	10
	Passage 5	ข้อสอบมี 2 ลักษณะ คือ ข้อ 91 – 93 ให้นำคำตอบที่กำหนดให้มาตอบคำถาม และข้อ 94 – 100 เป็นข้อสอบแบบเลือกตอบ	10
รวมจำนวนข้อสอบทั้ง 3 Part			100

2.3) การแปลผลคะแนนแบบทดสอบ STOU-EPT

การแปลผลคะแนนแบบทดสอบ STOU-EPT ตามกรอบ CEFR โดยระบุระดับความสามารถในแต่ละช่วงคะแนนของแต่ละส่วน ได้แก่ ระดับ A2 B1 และ B2 หลังสอบเสร็จผู้สอบจะทราบคะแนน และระดับความสามารถตามเกณฑ์ CEFR รายละเอียดแสดงดังตารางที่ 5

ตารางที่ 5 การแปลผลคะแนนการสอบ STOU-EPT

Marks	Descriptions	CEFR based level
<i>Part I Listening (25 Marks)</i>		
0	No attempt made.	-
1-8	The candidate probably has sufficient performance to understand everyday English and short, simple, clear, personal and public messages.	A2
9-20	The candidate probably has sufficient performance to access straightforward factual information and language used in educational contexts.	B1
21-25	The candidate probably has sufficient performance to access standard spoken language in educational contexts and various content-based lectures.	B2
<i>Part II Structure (35 Marks)</i>		
0	No attempt made.	-
1-10	The candidate probably can use some simple structures correctly but still systematically makes basic mistakes.	A2
11- 23	The candidate probably can use some simple structures correctly but still systematically makes fewer basic mistakes than A2 level.	B1
24-35	The candidate probably can show sufficient knowledge for straightforward situations in educational contexts.	B2

ตารางที่ 5 การแปลผลคะแนนการสอบ STOU-EPT (ต่อ)

Marks	Descriptions	CEFR based level
<i>Part III Reading (40 Marks)</i>		
0	No attempt made.	-
1-7	The candidate probably can access simple documents on everyday matters and simple informational sources.	A2
8-21	The candidate probably can access expository and informative sources, and other English texts in commonly encountered educational contexts.	B1
22-40	The candidate probably can access English texts that might be encountered most in educational contexts.	B2

ที่มา: การแปลผลคะแนนการสอบ STOU-EPT (สำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช, 2565: ออนไลน์)

2.3 แนวคิดเกี่ยวกับการประเมินความสอดคล้องของเครื่องมือวัดหรือเครื่องมือประเมิน

เนื้อหาในตอนนี้ แบ่งออกเป็น 2 หัวข้อ ได้แก่ 2.3.1 รูปแบบของความสอดคล้อง และ 2.3.2 การประเมินความสอดคล้อง รายละเอียดแต่ละหัวข้อมีดังนี้

2.3.1 รูปแบบของความสอดคล้อง

ความสอดคล้องของมาตรฐานและเครื่องมือประเมินสามารถทำได้ 2 รูปแบบ ได้แก่ ความสอดคล้องในแนวนอน (horizontal alignment) เป็นความสอดคล้องของเครื่องมือประเมินกับมาตรฐานวิชาใดวิชาหนึ่งของระดับชั้นใดชั้นหนึ่ง เช่น วิชาวิทยาศาสตร์ ชั้น ม.1 โดยวัตถุประสงค์ของการวิเคราะห์ความสอดคล้องในแนวนอน เพื่อให้มั่นใจว่าเครื่องมือวัดหรือเครื่องมือประเมินมีเนื้อหาสอดคล้องกับมาตรฐานที่กำหนดทั้งด้านความครอบคลุมและความลึกของเนื้อหา และความสอดคล้องในแนวตั้ง (vertical alignment) เป็นความสอดคล้องของเครื่องมือประเมินกับมาตรฐานในวิชาใดวิชาหนึ่งของระดับชั้นต่าง ๆ ต่อเนื่องกัน สาเหตุที่จำเป็นต้องศึกษาความสอดคล้องในแนวตั้ง คือ ความต้องตรวจสอบความเหมาะสมของการวัดพัฒนาการของผู้เรียน เพื่อให้มั่นใจว่าผลการประเมินสามารถบ่งบอกถึงพัฒนาการของผู้เรียนได้อย่างเหมาะสม (สังวรณั ังค์กระโทก, 2561)

2.3.2 การประเมินความสอดคล้อง

1) วิธีการประเมินความสอดคล้องของมาตรฐานกับเครื่องมือประเมิน ที่นิยมใช้มี 3 วิธี ได้แก่ การทำเป็นขั้นตอน การประเมินโดยผู้เชี่ยวชาญ และการวิเคราะห์เอกสาร รายละเอียดแต่ละวิธีมีดังนี้ (สังวรรณ ังตกระโทก, 2561)

1.1) การทำเป็นขั้นตอน (sequential development) แนวคิดของวิธีนี้ คือ มาตรฐานและเครื่องมือประเมินพัฒนาต่อเนื่องกัน คือ ต้องพัฒนาเนื้อหาของแต่ละวิชาก่อน แล้วผู้สร้างเครื่องมือประเมินหรือผู้สร้างแบบทดสอบจะเข้ามาตราฐานเนื้อหาเป็นพิมพ์เขียว สำหรับกำหนดโครงสร้างและเนื้อหาของการประเมินผู้เรียน ระหว่างพัฒนาเครื่องมือจะต้องบอกได้ว่าข้อสอบแต่ละข้อสอดคล้องกับมาตรฐานเพียงใด หลังจากนั้นนำผลประเมินความสอดคล้องไปประชุมกลุ่มผู้เขียนข้อสอบเพื่อให้เกิดความมั่นใจในคุณภาพของเครื่องมือประเมิน และลดคำวิพากษ์เรื่องคุณภาพข้อสอบจากนักการศึกษา ผู้เชี่ยวชาญ และสังคม จุดเด่นของวิธีการนี้ คือ ตรงไปตรงมา ไม่ซับซ้อน

1.2) การประเมินโดยผู้เชี่ยวชาญ (expert review) วิธีนี้ใช้ความคิดเห็นของผู้เชี่ยวชาญประเมินความสอดคล้อง ผู้เชี่ยวชาญ ประกอบด้วย ผู้เชี่ยวชาญด้านเนื้อหา ด้านการสร้างเครื่องมือประเมินผลการศึกษา การประเมินวิธีนี้มักจะใช้เมื่อมาตรฐานด้านเนื้อหา และเครื่องมือประเมินได้พัฒนาขึ้นทั้ง 2 อย่างแล้ว หลักการสำคัญ คือ การให้ผู้เชี่ยวชาญประเมินข้อสอบในเครื่องมือประเมินแต่ละข้อเทียบกับมาตรฐาน อย่างไรก็ตาม ในทางปฏิบัติวิธีการนี้มีความหลากหลาย ไม่มีกฎเกณฑ์แน่นอนว่าจะต้องดำเนินการอย่างไรบ้าง ขึ้นอยู่กับบริบท วิธีการประเมินในกลุ่มนี้ เช่น วิธีการของเวบบ์ วิธีการของพอร์เตอร์

1.3) การวิเคราะห์เอกสาร (document analysis) เป็นการใช้คณะผู้เชี่ยวชาญวิเคราะห์เนื้อหา และโครงสร้างในเครื่องมือประเมิน โดยมีแนวทางใช้รหัสเนื้อหาและโครงสร้างเมื่อนำผลการวิเคราะห์ของผู้เชี่ยวชาญมาวิเคราะห์ร่วมกันจะสามารถวัดความสอดคล้องออกมาเป็นเชิงปริมาณ การประเมินความสอดคล้องวิธีนี้มักจะใช้ในการประเมินด้านคณิตศาสตร์และวิทยาศาสตร์ (TIMSS)

2) วิธีการวิเคราะห์ความสอดคล้องที่นิยมใช้ มี 3 วิธี ได้แก่ (สังวรรณ ังตกระโทก, 2561)

2.1) วิธีการของเวบบ์ วิธีการนี้ได้พัฒนาขึ้นในปี ค.ศ. 1997 โดยเวบบ์ (Webb) ซึ่งวิธีนี้เป็นวิธีที่นิยมใช้แพร่หลายในแวดวงการวัดและประเมินผลทางการศึกษา เกณฑ์ในการวิเคราะห์ความสอดคล้องของเวบบ์ มี 5 เกณฑ์ ได้แก่ (1) เนื้อหา (content focus) เป็นการพิจารณาเนื้อหาตามมาตรฐาน (2) การเปลี่ยนแปลงตามระดับชั้นและอายุ (articulation across grades and ages) เป็นการพิจารณาพัฒนาการการเปลี่ยนแปลงของความรู้ในช่วงเวลาหนึ่ง ๆ (3) ความยุติธรรม (equity and fairness) เป็นการพิจารณาความเป็นธรรมและยุติธรรมในกลุ่มประชากรกลุ่มต่าง ๆ (4) การจัดการเรียนการสอน (pedagogical implications) เป็นการพิจารณาปัจจัยเอื้อให้ผู้เรียนเกิดการ

เรียนรู้ เช่น ทักษะคิดต่อวิชา สื่อการสอน และ (5) ระบบ (system applicability) ต้องการให้การประเมินและมาตรฐานสอดคล้องกันตามความเป็นจริง สมเหตุสมผล จัดการได้

อย่างไรก็ตาม ในทางปฏิบัติเกณฑ์เหล่านี้ได้นำไปใช้เพียงบางเกณฑ์เท่านั้น กล่าวคือ ด้านการวัดและประเมินที่อิงมาตรฐาน เว็บสนใจเพียงเกณฑ์ด้านเนื้อหาเท่านั้น โดยเกณฑ์ประเมินความสอดคล้องด้านเนื้อหา แบ่งออกเป็น 5 เกณฑ์ ได้แก่ ความสอดคล้องของเนื้อหา (categorical occurrence) ความลึกของความรู้ (depth of knowledge: DOK) ความกว้างของความรู้ (range of knowledge correspondence) ความสมดุล (balance of representation) และแหล่งที่มาของความท้าทาย (source of challenge) (Webb, 1997) มีรายละเอียดดังนี้

(1) ความสอดคล้องของเนื้อหา เป็นการประเมินว่าเนื้อหาที่มีในมาตรฐานและเครื่องมือประเมินตรงกันหรือเหมือนกันอย่างไร หรือการประเมินด้วยเครื่องมือวัดต่าง ๆ วัดตรงกับเนื้อหาที่กำหนดในมาตรฐานหรือไม่

(2) ความสอดคล้องด้านความลึกของความรู้ เป็นการเปรียบเทียบความซับซ้อนของเนื้อหาที่กำหนดในมาตรฐานและความซับซ้อนของเนื้อหาที่วัดโดยเครื่องมือประเมินหรือในข้อสอบสอดคล้องกันหรือไม่ ซึ่งความสอดคล้องด้านความลึกของความรู้ แบ่งออกเป็น 4 ระดับ ได้แก่

ระดับ 1 ความจำ (recall) เป็นระดับความรู้ความสามารถทางสมองที่ต้องการให้ผู้เรียนจำข้อมูลและบอกสิ่งที่เคยเห็นมา เช่น ข้อเท็จจริง นิยาม วิธีการ คุณสมบัติ ความรู้ระดับนี้ไม่ซับซ้อน เช่น การใช้สูตร การทำตามคำสั่ง โดยตัวอย่างคำที่แสดงถึงความจำ เช่น ระบุ จำได้ ใช้ คำนวณ

ระดับ 2 การมีทักษะและมโนทัศน์ (skills and concepts) เป็นเรื่องการตัดสินใจเกี่ยวกับการแก้ปัญหา ซึ่งการแก้ปัญหานี้ต้องดำเนินการมากกว่า 1 ขั้นตอน ดังนั้น ทักษะนี้เป็นการใช้ทักษะกระบวนการทางสมองสูงกว่าระดับ 1 ความจำ โดยตัวอย่างคำที่แสดงถึงการมีทักษะและมโนทัศน์ เช่น จำแนก สร้าง ประเมิน รวบรวม นำเสนอ เปรียบเทียบ

ระดับ 3 การคิดเชิงกลยุทธ์ (strategic thinking) เป็นการคิดที่ต้องมีเหตุผล มีการวางแผน และมีการใช้หลักฐานสนับสนุนอย่างสมเหตุสมผล และเป็นการคิดสูงกว่าระดับ 1 ความจำ และระดับ 2 การมีทักษะและมโนทัศน์ โดยตัวอย่างคำที่แสดงถึงการคิดเชิงกลยุทธ์ เช่น การสรุป การวิพากษ์ด้วยหลักฐานอย่างมีเหตุผล การแก้ปัญหาที่ไม่ใช่ปัญหาในชีวิตประจำวัน

ระดับ 4 การคิดหลายขั้นตอน (extended thinking) เป็นการคิดขั้นสูงและซับซ้อน ต้องการให้ผู้สอบเชื่อมโยงประเด็นหลายส่วนเข้าด้วยกัน เช่น สามารถสรุปอ้างอิงผลการศึกษาและนำไปใช้แก้ปัญหาในสถานการณ์ใหม่ได้ โดยตัวอย่างคำที่แสดงถึงการคิดหลายขั้นตอน เช่น งานที่ต้องใช้การวิเคราะห์ ที่ต้องอาศัยการทำอย่างต่อเนื่อง มีการทำหลายขั้นตอน

(3) ความสอดคล้องด้านความกว้างของความรู้ เป็นการเปรียบเทียบความกว้างของความรู้ (span of knowledge) หรือความหลากหลายของเนื้อหาที่มีในมาตรฐานกับความหลากหลายของความรู้ที่วัดในเครื่องมือประเมินว่าสอดคล้องกันเพียงใด

(4) ความสอดคล้องด้านความสมดุล เป็นการประเมินว่าการเน้นของเนื้อหาในมาตรฐาน และการเน้นของเนื้อหาในเครื่องมือประเมินเน้นสอดคล้องกันหรือไม่ หรือเครื่องมือประเมินเน้นวัดเนื้อหาในมาตรฐานเท่าเทียมกันหรือไม่ ไม่เน้นเนื้อหาใดเนื้อหาหนึ่งมากเกินไป

(5) แหล่งที่มาของความท้าทาย เป็นเกณฑ์ที่ใช้ตรวจสอบว่าข้อสอบมีความยากเกินความจำเป็นอันเนื่องมาจากปัจจัยอื่นที่ไม่เกี่ยวข้องกับทักษะที่ต้องการวัด เช่น อคติทางวัฒนธรรม หรือความรู้เฉพาะทาง ข้อสอบที่มีปัญหานี้อาจทำให้ผู้สอบบางคนไม่สามารถตอบคำถามได้ หรือทำคะแนนได้น้อยกว่าความสามารถที่แท้จริง แม้ว่าจะมีความเข้าใจและทักษะที่จำเป็นก็ตาม

วิธีการประเมินความสอดคล้องของเว็บ ถูกนำไปใช้ในงานวิจัยจำนวนมาก เช่น Roach and Elliott (2004) ศึกษาความสอดคล้องของแบบทดสอบทางเลือกในรัฐวิสคอนซิน และไอดาโฮ โดยใช้โมเดลของเว็บร่วมกับกระบวนการบุ๊คมาร์คที่ได้รับการปรับปรุง (Modified Bookmark Procedure) ผลการศึกษา พบว่า แบบทดสอบส่วนใหญ่สอดคล้องกับมาตรฐานทางวิชาการ โดยเฉพาะด้านเนื้อหา ความลึกของความรู้ และความสมดุล อย่างไรก็ตาม การประเมินทางเลือกสำหรับผู้เรียนที่มีความต้องการพิเศษมีระดับความลึกของความรู้ต่ำกว่ามาตรฐานทั่วไป และกระบวนการบุ๊คมาร์คที่ได้รับการปรับปรุงสามารถกำหนดเกณฑ์คะแนนได้อย่างเป็นธรรมและแม่นยำ ต่อมา Roach, Elliott, and Webb (2005) ได้วิเคราะห์ความสอดคล้องของ Wisconsin Alternate Assessment (WAA) ซึ่งเป็นแบบทดสอบทางเลือกที่พัฒนาขึ้นสำหรับผู้เรียนที่มีความพิการ โดยเปรียบเทียบกับ Wisconsin Model Academic Standards เพื่อประเมินว่าข้อสอบสามารถสะท้อนแนวคิดและทักษะที่ระบุไว้ในมาตรฐานของรัฐได้มากน้อยเพียงใด การประเมินใช้แนวทางการวิเคราะห์ความสอดคล้องของเนื้อหาตามหลักเกณฑ์ของเว็บ ซึ่งพิจารณาปัจจัยสำคัญ 4 ด้าน ได้แก่ ด้านเนื้อหา ความลึกของความรู้ ความสมดุล และความกว้างของความรู้ ผลการศึกษา พบว่า WAA มีระดับความสอดคล้องที่ยอมรับได้ในด้านเนื้อหา ความลึกของความรู้ และความสมดุล อย่างไรก็ตาม ผลการประเมินด้านความลึกของความรู้ยังต่ำกว่ามาตรฐาน เนื่องจาก WAA ถูกออกแบบมาให้เหมาะสมกับผู้เรียนที่มีความต้องการพิเศษ ดังนั้น เนื้อหาส่วนใหญ่จะเน้นไปที่ทักษะพื้นฐานมากกว่าทักษะที่ซับซ้อน ในทำนองเดียวกัน Roach, Wixson, and Talapatra (2010) ได้ประเมินระดับความสอดคล้องระหว่าง Indiana's Kindergarten Content Standards กับข้อคำถามของแบบประเมิน Indiana Standards Tool for Alternate Reporting (ISTAR) โดยความสอดคล้อง หมายถึง ระดับที่มาตรฐานและแบบประเมินสามารถทำงานร่วมกันเพื่อช่วยให้ผู้สอนสนับสนุนการเรียนรู้และพัฒนาการเรียนรู้ของผู้เรียนได้อย่างมีประสิทธิภาพ การศึกษานี้จึงได้ประยุกต์แนวทางการประเมิน

ความสอดคล้องของเว็บ โดยใช้ผู้เชี่ยวชาญ จำนวน 13 คน ผลการศึกษา พบว่า ISTAR มีความสอดคล้องกับเกณฑ์ของเว็บด้านความลึกของความรู้ และความกว้างของความรู้ไม่สม่ำเสมอ อย่างไรก็ตาม ผลการศึกษาที่น่าสนใจ คือ ระดับ F1 และ F2 ของ ISTAR มีความสอดคล้องสูงกับมาตรฐานของอนุบาล แสดงให้เห็นว่า ISTAR สามารถใช้ประเมินพัฒนาการของทักษะและองค์ความรู้ของเด็กได้อย่างต่อเนื่องตามลำดับพัฒนาการ Forte (2017) ศึกษาความสอดคล้องของการประเมินขนาดใหญ่ที่อิงตามมาตรฐานโดยใช้โมเดลของเว็บ โดยวิเคราะห์เนื้อหา ความลึกของความรู้ และความสมดุลของข้อสอบ ผลการวิจัย พบว่า แม้อัตราข้อสอบมาตรฐานของรัฐครอบคลุมเนื้อหาตามข้อกำหนด แต่บางส่วนยังขาดความลึกของความรู้ที่เหมาะสม และการใช้ข้อสอบจากหลายแหล่งและการกำหนดเกณฑ์การให้คะแนนที่แตกต่างกัน ส่งผลต่อความแม่นยำในการวิเคราะห์ผลการทดสอบ งานวิจัยนี้จึงเน้นย้ำถึงความสำคัญของการพัฒนาแบบทดสอบที่สะท้อนผลสัมฤทธิ์ของผู้เรียนได้อย่างถูกต้องและเป็นธรรม นอกจากนี้ Traynor et al. (2020) นำวิธีของเว็บมาใช้ในการวิเคราะห์ความสอดคล้องของแบบทดสอบมากกว่า 100 ชุดในสหรัฐอเมริกา เพื่อปรับปรุงคุณภาพข้อสอบโดยใช้ข้อมูลระดับข้อสอบ (item-level data) วิเคราะห์ความสอดคล้องด้านความลึกของความรู้ และเปรียบเทียบการให้คะแนนของผู้เชี่ยวชาญภายนอกกับข้อมูลภายใน ผลการศึกษา พบว่า ข้อสอบบางข้อมีปัญหาด้านแหล่งที่มาของความท้าทาย ซึ่งอาจทำให้ผู้สอบได้คะแนนคลาดเคลื่อนจากปัจจัยที่ไม่เกี่ยวข้องกับความสมารถที่ต้องการวัด ข้อสอบที่มีความสอดคล้องต่ำกว่าร้อยละ 75 เมื่อเทียบกับมาตรฐานจำเป็นต้องได้รับการปรับปรุง นอกจากนี้ พบว่าความคลาดเคลื่อนในการกำหนดระดับความลึกของความรู้ อาจเกิดจากการตีความที่แตกต่างกัน ดังนั้น ควรมีการฝึกการประเมินความสอดคล้องก่อนทำการประเมินจริง อีกทั้งขอเสนอแนะจากผู้เชี่ยวชาญยังช่วยลดอคติทางวัฒนธรรมและเพิ่มความเป็นธรรมของข้อสอบได้

หลักการสำคัญของวิธีการประเมินความสอดคล้องตามแนวคิดของเว็บได้มีการนำไปประยุกต์ใช้ทั้งโดยตรงและโดยอ้อมในการศึกษาที่เกี่ยวข้องกับการประเมินความสอดคล้องตามกรอบ CEFR ในปัจจุบัน เช่น Tangsakul and Poonpon (2024) ใช้ CEFR Content Analysis Grid เพื่อจำแนกความครอบคลุมของข้อสอบตามระดับ CEFR ซึ่งสะท้อนเกณฑ์ของเว็บในด้านความสอดคล้องของเนื้อหา และความกว้างของความรู้ ผลการศึกษาแสดงให้เห็นว่า การทำแผนที่ข้อสอบกับตัวบรรยาย CEFR อย่างเป็นระบบ (content mapping) ช่วยยืนยันความตรงเชิงเนื้อหาว่าข้อสอบสอดคล้องกับมาตรฐาน CEFR และทำให้เห็นภาพความครอบคลุมของเนื้อหาอย่างมีหลักฐานรองรับ ซึ่งเน้นถึงความจำเป็นในการนำกระบวนการประเมินความสอดคล้องไปใช้ในการตรวจสอบความตรงตามกรอบ CEFR

อย่างไรก็ตาม กรอบแนวคิดของเว็บไม่มีการนำไปใช้อย่างแพร่หลายในการประเมินภาษาอังกฤษในประเทศไทย การวิจัยครั้งนี้ จึงได้เติมเต็มช่องว่างดังกล่าวโดยนำกรอบแนวคิดดังกล่าวมาประยุกต์ใช้กับแบบทดสอบวัดระดับความสามารถทางภาษาที่อ้างอิงกรอบ CEFR ที่มีความสำคัญสูง (high-stakes) เนื่องจากกรอบของเว็บช่วยตรวจสอบได้อย่างเป็นระบบว่าข้อสอบครอบคลุมหัวข้อสำคัญตามมาตรฐานครบถ้วนเพียงใด และข้อสอบมีการกระจายระดับ DOK อย่างเหมาะสมหรือไม่ (Webb, 1997)

2.2) วิธีการของพอร์เตอร์และสมิธสัน (Porter & Smithson) หรือวิธีการสำรวจหลักสูตรที่นำไปใช้จริง (Surveys of Enacted Curriculum: SEC) วิธีการนี้ได้พัฒนาขึ้นโดยพอร์เตอร์และสมิธสัน เป็นวิธีประเมินที่ใช้ผู้เชี่ยวชาญ โดยการวัดความสอดคล้องเริ่มจากการแบ่งประเภทของเนื้อหาออกเป็นเรื่อง ๆ ในแต่ละเรื่องจะทำการประเมินความลึกของความรู้ ซึ่งเป็นทักษะด้านสมรรถนะที่ต้องการให้ผู้เรียนบรรลุ เช่น การรู้ความจำ ความเข้าใจ การนำไปใช้ การวัดความสอดคล้องวิธีนี้จะนำผลการประเมินของผู้เชี่ยวชาญมาสร้างเมตริกซ์ 2 เมตริกซ์ เช่น เมตริกซ์ของมาตรฐานหรือเนื้อหาตามมาตรฐาน (X) และเมตริกซ์ของเครื่องมือประเมิน (Y) แล้วนำมาคำนวณดัชนีความสอดคล้อง

SEC เป็นเครื่องมือที่ได้รับการออกแบบให้สามารถเก็บข้อมูลเกี่ยวกับแนวปฏิบัติในการสอนและความสอดคล้องของหลักสูตรกับมาตรฐานทางวิชาการในระดับชั้นเรียน โดย Blank (2002) ร่วมมือกับสภาผู้อำนวยการการศึกษาของรัฐ (Council of Chief State School Officers: CCSSO) และ Wisconsin Center for Education Research (WCER) เพื่อพัฒนาแนวทางที่สามารถเก็บรวบรวมและรายงานข้อมูลเกี่ยวกับหลักสูตรที่ถูกนำไปใช้จริง ในวิชาคณิตศาสตร์และวิทยาศาสตร์ระดับ K-12 แบบสำรวจนี้ได้รับการออกแบบให้เป็นเครื่องมือที่ช่วยเก็บข้อมูลเกี่ยวกับการสอนคณิตศาสตร์และวิทยาศาสตร์อย่างเป็นระบบและเชื่อถือได้ ผ่านการรายงานโดยผู้สอน ข้อมูลที่ได้จากแบบสำรวจสามารถช่วยให้หน่วยงานระดับรัฐ เขตพื้นที่การศึกษา และสถาบันการศึกษาวิเคราะห์แนวปฏิบัติในการสอน รวมถึงการพัฒนาวิชาชีพของครู โดยเชื่อมโยงกับมาตรฐานเนื้อหาและเป้าหมายของระบบการศึกษาเพื่อการปรับปรุงคุณภาพการเรียนการสอน การศึกษาที่เกี่ยวข้องได้นำเสนอผลการใช้ SEC ในระยะแรก และแสดงให้เห็นว่าระบบการรายงานข้อมูลจากแบบสำรวจเป็นเครื่องมือที่ทรงพลังสำหรับนักการศึกษาและผู้ประเมินด้านการศึกษา ในการวิเคราะห์และพัฒนา นโยบายทางการศึกษา รวมถึงส่งเสริมการยกระดับผลสัมฤทธิ์ทางการเรียนของผู้เรียนได้อย่างมีประสิทธิภาพ ด้วยความสำคัญของการพัฒนามาตรฐานการศึกษา Polikoff et al. (2019) ได้ปรับปรุงแบบสำรวจ SEC และวิธีการใช้ SEC เพื่อติดตามการดำเนินงานของมาตรฐานใหม่ที่เน้นความร่วมมือทางวิชาการและอาชีพ การพัฒนานี้ประกอบด้วยการประชุมผู้เชี่ยวชาญด้านเนื้อหา การสัมภาษณ์เชิงลึก เพื่อตรวจสอบความถูกต้อง และการทดสอบนำร่องขนาดเล็ก ผลการศึกษา พบว่า SEC เป็นเครื่องมือที่มีประสิทธิภาพในการประเมินความสอดคล้องของหลักสูตรกับมาตรฐานทางวิชาการ และ

ช่วยให้ผู้กำหนดนโยบาย นักวิจัย และนักการศึกษาใช้ข้อมูลในการปรับปรุงหลักสูตร ระบบการเรียน การสอน และการพัฒนาผู้สอนได้อย่างเป็นระบบ สำหรับประเทศไทยมีการประยุกต์แนวทางที่ คล้ายคลึงกันในงานวิจัยของ พงศม ใประหมาน (2559) ได้ศึกษาความสอดคล้องระหว่างข้อสอบ ระดับชาติวิชาภาษาฝรั่งเศสกับกรอบมาตรฐาน CEFR โดยใช้วิธีแผนที่ข้อสอบ ผลการวิจัย พบว่า ข้อสอบที่สอดคล้องกับกรอบมาตรฐาน CEFR มี 16 ข้อ ซึ่งทั้งหมดวัดสมรรถนะการอ่านในระดับภาษา A2 และวัดทักษะทางสมองขั้นวิเคราะห์/สืบสวน นอกจากนี้ ผลการประเมินโดยใช้วิธีแผนที่ข้อสอบกับ ผลการประเมินโดยผู้เชี่ยวชาญมีความสอดคล้องกัน อย่างไรก็ตาม ปัจจัยที่ส่งผลต่อความไม่สอดคล้อง มี 3 ด้าน ได้แก่ (1) ความไม่สัมพันธ์ระหว่างเนื้อหาที่ใช้ออกข้อสอบกับเนื้อหาของมาตรฐาน (2) การ วัดทักษะทางสมองที่แตกต่างกัน และ (3) การใช้ภาษาและไวยากรณ์ในการสร้างข้อสอบที่ไม่สัมพันธ์ กับมาตรฐาน และไม่สัมพันธ์กับระดับความสามารถทางภาษาของผู้เรียน

โดยสรุปแนวทางของพอร์เตอร์และสมิธชันหรือ SEC เป็นเครื่องมือที่มี ประสิทธิภาพในการวิเคราะห์ความสอดคล้องระหว่างหลักสูตรกับมาตรฐานทางการศึกษา ซึ่งสามารถ นำไปใช้ในการวิจัยด้านการศึกษาเพื่อปรับปรุงการสอน การวัดผล และการพัฒนาหลักสูตรให้มีความ สอดคล้องกับมาตรฐานทางวิชาการ

2.3) วิธีการของบริษัท Achieve หรือ Achieve model วิธีการนี้พัฒนาขึ้นเพื่อ เปรียบเทียบความสอดคล้องของมาตรฐานการศึกษาของรัฐต่าง ๆ ผลการเปรียบเทียบสามารถนำไปใช้ พัฒนาบุคลากรทางการศึกษา การปฏิรูปการศึกษา และการสร้างเครื่องมือประเมิน โดยมีเกณฑ์ ประเมินความสอดคล้อง 4 เกณฑ์ ดังนี้ (Rothman et al., 2002)

(1) จุดเน้นของเนื้อหา (content centrality) เกณฑ์นี้เป็นการวิเคราะห์เชิงลึก เกี่ยวกับความสอดคล้องระหว่างเนื้อหาของแต่ละคำถามในข้อสอบกับเนื้อหาของมาตรฐานที่เกี่ยวข้อง โดยพิจารณาระดับหรือคุณภาพของความสอดคล้องดังกล่าว ผู้ประเมินจะต้องจัดข้อสอบแต่ละข้อให้ อยู่หนึ่งในสี่ประเภทตามระดับของความสอดคล้องที่กำหนดไว้

(2) จุดเน้นด้านผลงาน (performance centrality) เกณฑ์นี้มุ่งเน้นไปที่ระดับ ของความสอดคล้องระหว่างประเภทของระดับทางปัญญา (cognitive demand) ที่ข้อสอบแต่ละข้อ ต้องการให้ผู้สอบแสดงออกตามสมรรถนะที่ระบุไว้ในมาตรฐาน โดยข้อสอบแต่ละข้อต้องการให้ผู้เรียน ใช้ทักษะทางปัญญาในรูปแบบที่แตกต่างกัน เช่น การเลือก (select) การระบุ (identify) การ เปรียบเทียบ (compare) หรือการวิเคราะห์ (analyze) ผู้ประเมินจะต้องจัดข้อสอบแต่ละข้อให้อยู่ หนึ่งในสี่ประเภทตามระดับของความสอดคล้องที่กำหนดไว้

(3) ความท้าทาย (challenge) เกณฑ์นี้ถูกนำมาใช้กับกลุ่มของข้อสอบเพื่อ ตรวจสอบว่า ผู้สอบจะทำข้อสอบได้ดีต้องอาศัยการเรียนรู้เนื้อหาที่มีความท้าทายหรือไม่ ผู้ประเมินจะ พิจารณาสองปัจจัยในการประเมินข้อสอบตามเกณฑ์ความท้าทาย ได้แก่

(3.1) แหล่งที่มาของความท้าทาย (source of challenge) ใช้ในการวิเคราะห์ว่า ข้อสอบมีความยากเนื่องจากองค์ความรู้และทักษะที่ต้องการวัด หรือเกิดจากปัจจัยอื่นที่ไม่เกี่ยวข้องกับเนื้อหา เช่น การอาศัยความรู้พื้นฐานของผู้เรียนที่ไม่ได้เป็นส่วนหนึ่งของมาตรฐาน ผู้ประเมินจะให้คะแนนแต่ละข้อว่ามีแหล่งที่มาของความท้าทายเหมาะสมหรือไม่เหมาะสม

(3.2) ระดับความท้าทาย (level of challenge) ใช้ในการเปรียบเทียบระดับความสามารถที่ข้อสอบต้องการกับระดับความสามารถที่กำหนดในมาตรฐาน นอกจากการพิจารณาความสอดคล้องแล้ว ผู้ประเมินยังพิจารณาด้วยว่า กลุ่มของข้อสอบมีระดับความยากที่เหมาะสมสำหรับผู้เรียนในระดับชั้นที่กำหนดหรือไม่ โดยอิงจากมาตรฐาน หลักเกณฑ์การประเมินและเอกสารประกอบ ผู้ประเมินจะเขียนสรุปสั้น ๆ เกี่ยวกับระดับความท้าทายของชุดข้อสอบแต่ละชุดเพื่อสะท้อนความเหมาะสมในการวัดผลผู้เรียน

(4) ความสมดุลและความกว้าง (balance and range) ข้อสอบต้องครอบคลุมช่วงเนื้อหาทั้งหมดของมาตรฐาน พร้อมทั้งมีการกระจายความสำคัญของเนื้อหาอย่างเหมาะสม อย่างไรก็ตาม ไม่มีข้อสอบใดที่สามารถวัดองค์ความรู้และทักษะทั้งหมดที่ระบุไว้ในมาตรฐานได้อย่างสมบูรณ์ การประเมินความสมดุล และความกว้างของเนื้อหา จึงเป็นกระบวนการที่ให้ข้อมูลทั้งเชิงคุณภาพและเชิงปริมาณเกี่ยวกับทางเลือกที่ผู้พัฒนาแบบทดสอบใช้ในการออกแบบข้อสอบ

(4.1) ความสมดุล เป็นการเปรียบเทียบระหว่างการเน้นเนื้อหาที่ปรากฏในชุดข้อสอบกับการเน้นเนื้อหาที่กำหนดไว้ในมาตรฐาน นอกจากการพิจารณาความสอดคล้องแล้ว ผู้ประเมินยังพิจารณาด้วยว่า ชุดข้อสอบให้ความสำคัญกับเนื้อหาที่สำคัญในระดับชั้นนั้นหรือไม่ ผู้ประเมินจะต้องสรุปข้อมูลเกี่ยวกับความสมดุลของแต่ละชุดข้อสอบอย่างกระชับ

(4.2) ความกว้างของเนื้อหา เป็นตัวชี้วัดขอบเขตหรือความครอบคลุมของข้อสอบ ซึ่งพิจารณาจาก สัดส่วนเชิงปริมาณของเนื้อหาทั้งหมดที่ถูกลำมาทดสอบ เพื่อให้มั่นใจว่าข้อสอบได้วัดองค์ความรู้และทักษะที่จำเป็นตามที่กำหนดไว้ในมาตรฐานอย่างเพียงพอ

Rothman et al. (2002) กล่าวว่า การใช้วิธีการ Achieve ผู้ประเมินต้องมีความหลากหลาย เช่น เป็นผู้สอน ผู้เชี่ยวชาญด้านหลักสูตร และผู้เชี่ยวชาญในแต่ละสาขาวิชา ผู้ประเมินบางคนอาจมีประสบการณ์ในการพัฒนาแบบทดสอบขนาดใหญ่ มีประสบการณ์ในการพัฒนามาตรฐานทางการศึกษา ความหลากหลายของพื้นฐานความรู้และประสบการณ์มีประโยชน์อย่างมากในการตัดสินใจว่าข้อสอบเหมาะสมกับระดับชั้นเรียนหรือไม่ และผู้ประเมินจะต้องเข้าประชุมย่อย เพื่ออภิปรายเกี่ยวกับการตัดสินใจของตนเอง โดยมีผู้นำกลุ่มที่มีประสบการณ์เป็นผู้ดำเนินการอภิปราย ทั้งนี้ การตัดสินใจจะอยู่ภายใต้เครื่องมือและเกณฑ์การประเมินที่ใช้ในการตัดสินใจตัดสินความสอดคล้องในแต่ละมิติ

จากการศึกษาวิธีประเมินความสอดคล้องทั้ง 3 วิธี พบว่า วิธีของเว็บเป็นวิธีที่ได้รับความนิยมน้อยกว่า เนื่องจากมีเกณฑ์ครอบคลุมทั้งด้านเนื้อหา ความลึกของความรู้ ความกว้างของความรู้ ความสมดุล และแหล่งที่มาของความท้าทาย ซึ่งสามารถประยุกต์ได้ในหลายบริบทของการประเมินตั้งแต่การวิเคราะห์แบบทดสอบระดับชาติ การทดสอบทางเลือก ไปจนถึงการศึกษาความสอดคล้องของมาตรฐานในระดับต่าง ๆ ดังนั้น ในการศึกษาความสอดคล้องระหว่างแบบทดสอบ STOU-EPT กับกรอบ CEFR ผู้วิจัยจึงเลือกใช้วิธีประเมินความสอดคล้องด้วยวิธีของเว็บ โดยมุ่งเน้นประเมินความสอดคล้องใน 3 เกณฑ์หลัก ได้แก่ ความสอดคล้องของเนื้อหา ความสอดคล้องด้านความลึกของความรู้ และความสอดคล้องด้านความสมดุล ซึ่งทั้งสามเกณฑ์นี้เป็นหลักฐานสำคัญที่ช่วยยืนยันความน่าเชื่อถือของการตีความคะแนน และสอดคล้องกับแนวคิด Argument-Based Validity ของ Kane (2006) และมาตรฐานสากลของ AERA, APA, and NCME (2014) ที่ระบุว่า การมีหลักฐานที่แข็งแกร่งเป็นหัวใจสำคัญของความตรงของแบบทดสอบ ส่วนอีก 2 เกณฑ์ไม่ได้นำมาประเมิน ได้แก่ ความกว้างของความรู้ เนื่องจากแบบทดสอบ STOU-EPT พัฒนาตาม CEFR โดยตรงและกำหนดขอบเขตของเนื้อหาไว้ชัดเจน ส่วนแหล่งที่มาของความท้าทาย เป็นเกณฑ์ที่ใช้ตรวจสอบว่าข้อสอบมีความยากเกินความจำเป็นอันเนื่องมาจากปัจจัยอื่นที่ไม่เกี่ยวข้อง กับสิ่งที่ต้องการวัด (construct-irrelevant difficulty) เช่น อคติทางวัฒนธรรม หรือความรู้เฉพาะ เนื่องจากการออกแบบข้อสอบตามกรอบ CEFR มีเป้าหมายเพื่อลดปัจจัยดังกล่าวตั้งแต่ต้นอยู่แล้ว

อย่างไรก็ตาม แม้มีเหตุผลเชิงทฤษฎีรองรับการไม่นำเกณฑ์ความกว้างของความรู้และแหล่งที่มาของความท้าทายของเว็บมาใช้ในการศึกษาครั้งนี้ ก็ยังจำเป็นต้องใช้ความระมัดระวัง เนื่องจากการออกแบบข้อสอบที่อ้างอิงกรอบ CEFR เพียงอย่างเดียว ไม่ได้ยืนยันว่าข้อสอบจะครอบคลุมเนื้อหาอย่างเพียงพอหรือมีความเป็นธรรมในทุกมิติ อีกทั้งระดับความลึกของความรู้ อาจไม่สามารถสะท้อนความซับซ้อนที่หลากหลายของคำอธิบายคุณลักษณะ CEFR ได้ทั้งหมด (Alderson, 2007) ดังนั้น เพื่อเสริมหลักฐานสนับสนุนความตรงให้ครอบคลุมตามกรอบ Argument-Based Validity (Kane, 2006; AERA, APA, & NCME, 2014) การวิจัยในอนาคตควรพิจารณานำเกณฑ์ทั้ง 5 เกณฑ์ของเว็บมาใช้ และ/หรือจัดหาหลักฐานเพิ่มเติม เช่น การตรวจสอบความเป็นธรรมโดยใช้ DIF

2.4 มาตรฐานของการทดสอบทางการศึกษาและจิตวิทยา

เนื้อหาในส่วนนี้เกี่ยวกับมาตรฐานของการทดสอบทางการศึกษาและจิตวิทยา โดยแบ่งเนื้อหาออกเป็น 3 หัวข้อ ได้แก่ 2.4.1 จุดประสงค์ของมาตรฐาน และมาตรฐานของการทดสอบทางการศึกษาและจิตวิทยา 2.4.2 การตรวจสอบหลักฐานที่เกี่ยวข้องกับความตรงเชิงโครงสร้าง และ 2.4.3 มาตรฐานของการกำหนดคะแนนจุดตัด รายละเอียดของแต่ละหัวข้อมีดังนี้

2.4.1 จุดประสงค์ของมาตรฐาน และมาตรฐานของการทดสอบทางการศึกษาและจิตวิทยา

1) จุดประสงค์ของมาตรฐาน

สมาคมวิจัยทางการศึกษาแห่งอเมริกา (American Educational Research Association: AERA) กล่าวถึง จุดประสงค์ของมาตรฐาน เป็นการกำหนดเกณฑ์สำหรับการพัฒนาและการประเมินแบบทดสอบ และแนวปฏิบัติในการดำเนินการสอบ เพื่อเป็นแนวทางในการตรวจสอบความตรงของการตีความหมายของคะแนนการสอบ แม้ว่าการประเมินจะขึ้นอยู่กับ การตัดสินใจของผู้เชี่ยวชาญ มาตรฐานเป็นกรอบสำหรับอ้างอิงเพื่อให้มั่นใจว่าการพัฒนาและการประเมินแบบทดสอบ และการดำเนินการสอบดำเนินการถูกต้อง เหมาะสมตามกรอบมาตรฐาน ดังนั้น ผู้พัฒนาแบบทดสอบ ผู้ใช้แบบทดสอบ และผู้ที่เกี่ยวข้อง ควรปฏิบัติตามมาตรฐานเหล่านี้เพื่อรักษาคุณภาพของการวัดและการประเมินผล และส่งเสริมบุคคลอื่นให้ปฏิบัติตามเช่นกัน (AERA, 2014)

2) มาตรฐานของการทดสอบทางการศึกษาและจิตวิทยา

มาตรฐานของการทดสอบทางการศึกษาและจิตวิทยา เป็นชุดของมาตรฐานการทดสอบที่พัฒนาขึ้นโดยสมาคมวิจัยทางการศึกษาแห่งอเมริกา (AERA) สมาคมจิตวิทยาแห่งอเมริกา (APA) และสภาแห่งชาติด้านการวัดผลทางการศึกษา (NCME) โดยมาตรฐานของการทดสอบทางการศึกษาและจิตวิทยา ประกอบด้วย 3 ส่วน ได้แก่ 1) หลักการทดสอบ มีสาระสำคัญว่าด้วยความตรง ความเที่ยง/ความแม่นยำ/ความคลาดเคลื่อนในการวัด และความยุติธรรมในการทดสอบ 2) การปฏิบัติการทดสอบ มีสาระสำคัญว่าด้วย การออกแบบและพัฒนาแบบทดสอบ คะแนน/มาตรวัด/เกณฑ์ปกติ/การเชื่อมโยงคะแนน/คะแนนจุดตัด การบริหารการสอบ/คะแนน/การรายงาน/การตีความหมายของคะแนน เอกสารสนับสนุนสำหรับแบบทดสอบ สิทธิและความรับผิดชอบของผู้สอบ และสิทธิและความรับผิดชอบของผู้ใช้แบบทดสอบ และ 3) การนำแบบทดสอบไปใช้ มีสาระสำคัญว่าด้วยการทดสอบและประเมินทางจิตวิทยา การทดสอบในสถานที่ทำงานและการรับรอง การทดสอบและประเมินทางการศึกษา และการใช้แบบทดสอบสำหรับการประเมินโปรแกรม/การศึกษานโยบาย/ความรับผิดชอบ (AERA, 2014)

มาตรฐานของการทดสอบทางการศึกษาและจิตวิทยาที่กล่าวมาข้างต้น มีบทบาทสำคัญในการกำหนดแนวทางพัฒนาแบบทดสอบให้มีความตรง ความเที่ยง และความยุติธรรม ทั้งในด้านกระบวนการพัฒนา การบริหารการสอบ การให้คะแนน การแปลความหมายของคะแนน และการนำผลสอบไปใช้ให้เหมาะสม โดยมุ่งเน้นให้แบบทดสอบเป็นเครื่องมือที่มีคุณภาพ และสามารถวัดผลได้อย่างมีความยุติธรรมสำหรับผู้สอบทุกกลุ่ม

2.4.2 การตรวจสอบหลักฐานที่เกี่ยวข้องกับความตรงเชิงโครงสร้าง

การตรวจสอบหลักฐานที่เกี่ยวข้องกับความตรงเชิงโครงสร้าง ประกอบด้วย หลักฐานเกี่ยวกับเนื้อหา หลักฐานเกี่ยวกับกระบวนการตอบของผู้สอบ หลักฐานเกี่ยวกับโครงสร้างภายในของแบบทดสอบ หลักฐานเกี่ยวกับความสัมพันธ์ของคะแนนกับตัวแปรอื่น ๆ และหลักฐานเกี่ยวกับผลกระทบของการใช้คะแนน (สังวรณ ังตกระโทก, 2561) ซึ่งในการวิจัยครั้งนี้ ผู้วิจัยเน้นการตรวจสอบหลักฐานที่เกี่ยวข้องกับเนื้อหา และหลักฐานเกี่ยวกับโครงสร้างภายในของแบบทดสอบ มีรายละเอียดดังนี้

1) การตรวจสอบหลักฐานที่เกี่ยวข้องกับเนื้อหา ส่วนใหญ่เน้นการประเมินโดยใช้ผู้เชี่ยวชาญ หลักฐานเกี่ยวกับเนื้อหาที่ควรตรวจสอบ มี 8 ประเด็น ได้แก่ 1.1) การตรวจสอบโครงสร้างการสร้างแบบทดสอบ (test blueprint) การตรวจสอบนี้ใช้ผู้เชี่ยวชาญตรวจสอบความเหมาะสมของพิมพ์เขียวของแบบทดสอบว่าเหมาะสมหรือไม่ 1.2) ความเป็นตัวแทนของเนื้อหาในโครงสร้างแบบทดสอบ โดยพิจารณาว่าเนื้อหาที่นำมาออกข้อสอบสอดคล้องกับสิ่งที่ต้องการวัด และเป็นตัวแทนสิ่งที่ต้องการวัดหรือไม่ 1.3) โครงสร้างเนื้อหาของแบบทดสอบ (test specification) การสร้างและพัฒนาแบบทดสอบควรมีข้อสอบที่เป็นตัวแทนของเนื้อหาในพิมพ์เขียวเพียงพอ และไม่ควรถูกกำหนดเนื้อหาได้น้อยหรือมากเกินไป 1.4) ความสอดคล้องระหว่างข้อคำถามกับโครงสร้างเนื้อหา ข้อสอบในแบบทดสอบควรสอดคล้องกับเนื้อหาในด้านรูปแบบการตอบ และรูปแบบคำถาม เพื่อให้มั่นใจว่าแบบทดสอบวัดในสิ่งที่ต้องการจะวัดได้ดี 1.5) ความเป็นตัวแทนของข้อสอบแต่ละมาตรฐาน ข้อสอบในแต่ละมาตรฐานควรมีจำนวนเพียงพอที่จะวัดมาตรฐานนั้น ๆ โดยทั่วไปควรมีข้อสอบไม่น้อยกว่า 6 ข้อหรือมาตรฐาน เพราะหากน้อยเกินไปจะทำให้เกิดความคลาดเคลื่อนในการวัด จนไม่สามารถวัดและจำแนกผู้เรียนได้ 1.6) ความเกี่ยวข้องระหว่างเนื้อหาที่ต้องการทดสอบกับมาตรฐานการประเมินความเกี่ยวข้องระหว่างเนื้อหาที่ต้องการทดสอบกับมาตรฐาน ควรประเมินโดยใช้ดัชนีความสอดคล้อง (alignment) 1.7) คุณภาพของการเขียนตอบ อาจพิจารณาจากภาษาการใช้ถ้อยคำที่เหมาะสม เข้าใจง่าย ตรงประเด็น รวมทั้งรูปแบบข้อสอบสอดคล้องกับสิ่งที่ต้องการวัด และ 1.8) คุณสมบัติของผู้ออกข้อสอบ เป็นสิ่งสำคัญที่ควรประเมิน เพราะเกี่ยวข้องกับคุณภาพข้อสอบ โดยผู้ออกข้อสอบควรจะเป็นผู้เชี่ยวชาญด้านเนื้อหานั้น ๆ

กล่าวโดยสรุป การตรวจสอบหลักฐานเกี่ยวกับเนื้อหา มักจะใช้ผู้เชี่ยวชาญเป็นผู้ประเมิน ซึ่งควรครอบคลุมใน 8 ประเด็น ได้แก่ 1) การตรวจสอบโครงสร้างการสร้างแบบทดสอบ 2) ความเป็นตัวแทนของเนื้อหาในโครงสร้างแบบทดสอบ 3) โครงสร้างเนื้อหาของแบบทดสอบ 4) ความสอดคล้องระหว่างข้อคำถามกับโครงสร้างเนื้อหา 5) ความเป็นตัวแทนของข้อสอบแต่ละมาตรฐาน 6) ความเกี่ยวข้องระหว่างเนื้อหาที่ต้องการทดสอบกับมาตรฐาน 7) คุณภาพของการเขียนตอบ และ 8) คุณสมบัติของผู้ออกข้อสอบ

2) หลักฐานเกี่ยวกับโครงสร้างภายในของแบบทดสอบ การตรวจสอบความตรงของแบบทดสอบเป็นสิ่งสำคัญในการยืนยันความเชื่อถือได้ และความแม่นยำของการวัด โดยเฉพาะในบริบทของการศึกษา การรวบรวมหลักฐานที่เกี่ยวข้องกับโครงสร้างภายในของแบบทดสอบจะช่วยให้สามารถประเมินคุณภาพและประสิทธิภาพของการวัดได้อย่างถูกต้อง ประกอบด้วย 6 ประเด็นหลักดังนี้

2.1) การวิเคราะห์ข้อสอบ เป็นขั้นตอนแรกที่สำคัญในการประเมินคุณภาพของข้อสอบโดยใช้ทั้งทฤษฎีการทดสอบแบบดั้งเดิม (CTT) และทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อให้ได้ข้อมูลที่ชัดเจนเกี่ยวกับคุณสมบัติของข้อสอบ การประเมินคุณภาพของข้อสอบโดยใช้ IRT มีข้อได้เปรียบในการประเมินพารามิเตอร์ของข้อสอบ เช่น ความยาก (difficulty) อำนาจจำแนก (discrimination) และโอกาสเดา (guessing) จะไม่ขึ้นอยู่กับกลุ่มตัวอย่าง (sample independence) ซึ่งต่างจาก CTT ที่ค่าสถิติเปลี่ยนแปลงตามกลุ่มผู้สอบ (Embretson & Reise, 2000; van der Linden, 2016) การใช้ IRT ช่วยให้สามารถระบุข้อสอบที่มีคุณภาพสูง และสามารถคัดเลือกข้อสอบที่เหมาะสมสำหรับการวัดได้อย่างมีประสิทธิภาพ นอกจากนี้ IRT ยังช่วยตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (differential item functioning: DIF) ซึ่งเป็นการวิเคราะห์ว่า ข้อสอบแต่ละข้อทำงานแตกต่างกันในกลุ่มประชากรที่แตกต่างหรือไม่ โดยสามารถสร้างกราฟ Item Characteristic Curve (ICC) เพื่อประเมินความเหมาะสมของโมเดลได้อย่างชัดเจน

2.2) ความสัมพันธ์ของข้อสอบแต่ละข้อในฉบับเดียวกัน สะท้อนถึงความสอดคล้องภายใน (internal consistency) ของแบบทดสอบ ซึ่งเป็นสิ่งสำคัญในการยืนยันว่าข้อสอบทั้งหมดสามารถวัดคุณลักษณะเดียวกันได้ โดยทั่วไปจะใช้การวิเคราะห์องค์ประกอบหรือสหสัมพันธ์ระหว่างข้อสอบเพื่อประเมินความสอดคล้อง การใช้ค่าสัมประสิทธิ์ Cronbach's alpha เป็นวิธีที่นิยมในการประเมินความสอดคล้องภายใน โดยค่าที่สูงกว่า .70 ถือว่ามีความสอดคล้องในระดับที่ยอมรับได้ (Nunnally & Bernstein, 1994) นอกจากนี้ การใช้เทคนิคการวิเคราะห์อื่น ๆ เช่น การวิเคราะห์องค์ประกอบเชิงสำรวจ (exploratory factor analysis: EFA) และการวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis: CFA) ยังช่วยให้สามารถระบุโครงสร้างทางจิตมิติของแบบทดสอบ และยืนยันว่าข้อสอบมีความสอดคล้องกันในระดับที่เหมาะสม

2.3) ความสัมพันธ์ของคะแนนรายข้อ และคะแนนรวม เป็นอีกปัจจัยสำคัญในการตรวจสอบคุณภาพของแบบทดสอบ เนื่องจากข้อสอบที่มีคุณภาพควรมีความสัมพันธ์เชิงบวกกับคะแนนรวม ซึ่งสามารถวัดได้ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ เช่น Point-Biserial Correlation ค่าสัมประสิทธิ์ช่วยให้ทราบว่า ผู้ที่ตอบถูกมากกว่ามักจะได้คะแนนสูงกว่าผู้ที่ตอบผิด โดยทั่วไปแล้วค่าสัมประสิทธิ์ที่สูงกว่า .20 แสดงถึง ความสัมพันธ์ในระดับที่ยอมรับได้ (ศิริชัย กาญจนวาสี, 2563) การตรวจสอบความสัมพันธ์จะช่วยให้มั่นใจว่าข้อสอบมีคุณภาพ และสามารถจำแนกผู้สอบที่มีระดับความรู้แตกต่างกันได้อย่างชัดเจน

2.4) ความเที่ยงของมาตรคะแนนสะท้อนถึงความน่าเชื่อถือและคงเส้นคงวาของการวัด ซึ่งสามารถประเมินได้จากค่าสัมประสิทธิ์ความเที่ยง เช่น Cronbach's alpha หรือ Composite Reliability ค่าสัมประสิทธิ์ความเที่ยงที่สูงกว่า .70 แสดงถึง ความเที่ยงในระดับที่ยอมรับได้ (Dimitrov, 2003) นอกจากนี้ การตรวจสอบความเที่ยงยังสามารถทำได้โดยการเปรียบเทียบผลการวัดในช่วงเวลาที่แตกต่างกัน เพื่อดูว่าผลลัพธ์ยังคงเหมือนเดิมหรือไม่ การมีความเที่ยงสูงจะช่วยให้มั่นใจว่าเครื่องมือวัดนั้นสามารถใช้งานได้ดีในระยะยาว

2.5) ความคลาดเคลื่อนมาตรฐานของการวัด (standard error of measurement: SEM) เป็นเครื่องมือที่ใช้เพื่อระบุระดับความคลาดเคลื่อนที่ยอมรับได้ในการวัด ซึ่งเป็นข้อมูลสำคัญในการตีความผลการวัด SEM จะช่วยให้นักวิจัยเข้าใจถึงขอบเขตของค่าที่แท้จริง ซึ่งอาจเกิดขึ้นจากการวัด โดยทั่วไป SEM จะถูกคำนวณจากค่าความเที่ยง และรายงานเป็นช่วงคะแนน หาก SEM ต่ำแสดงว่า การวัดนั้นมีความแม่นยำสูง และผลลัพธ์มีแนวโน้มที่ใกล้เคียงกับค่าจริงมากขึ้น (Hambleton, Swaminathan, & Rogers, 1991) SEM ช่วยให้นักวิจัยสามารถตีความผลการวัดได้อย่างถูกต้อง และนำไปใช้ในการตัดสินใจในด้านการศึกษาอย่างมีหลักฐานรองรับ

2.6) การวิเคราะห์มิติของข้อสอบ เช่น การวิเคราะห์องค์ประกอบยืนยัน (CFA) เป็นวิธีที่ใช้เพื่อยืนยันโครงสร้างทางจิตมิติของแบบทดสอบ โดย CFA จะช่วยตรวจสอบว่าข้อสอบทั้งหมดทำงานร่วมกันเพื่อวัดคุณลักษณะเดียวกันหรือไม่ นอกจากนี้ ยังสามารถศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) เพื่อประเมินความลำเอียงในกลุ่มประชากร ซึ่งเป็นขั้นตอนสำคัญในการตรวจสอบความตรงเชิงโครงสร้าง (construct validity) ของแบบทดสอบ หากพบว่าข้อสอบบางข้อทำงานแตกต่างกันในกลุ่มประชากรที่แตกต่างกัน อาจต้องปรับปรุงหรือเปลี่ยนแปลงข้อสอบเพื่อให้เหมาะสมกับกลุ่มเป้าหมายมากขึ้น (Gulliksen, 1950)

การดำเนินการทั้ง 6 ข้อที่กล่าวมา มีเป้าหมายเพื่อยืนยันว่าแบบทดสอบนั้นมีหลักฐานทางจิตมิติที่แกร่ง และสามารถใช้งานได้อย่างมีประสิทธิภาพ

3) การเลือกใช้ทฤษฎีการทดสอบ เป็นสิ่งสำคัญในการประเมินความสามารถทางภาษาอังกฤษ ซึ่งมีลักษณะเฉพาะในการวัดทักษะที่หลากหลาย เช่น ทักษะการฟัง โครงสร้างภาษา

การอ่าน ข้อสอบในแบบทดสอบ ประกอบด้วย ข้อสอบเดี่ยว และกลุ่มข้อสอบ ที่ใช้สิ่งเร้าร่วมกันหรือที่เรียกว่า Testlet ซึ่งพบได้บ่อยในข้อสอบการอ่าน และการฟัง ที่มีข้อความหรือเสียงเป็นฐานในการตั้งคำถามหลายข้อ การออกแบบข้อสอบลักษณะนี้มักทำให้ข้อสอบมีความสัมพันธ์กันภายใน Testlet ส่งผลให้เกิดการละเมิดสมมติฐานของความเป็นอิสระของข้อสอบแต่ละข้อ (local independence) หมายถึง ข้อสอบแต่ละข้อไม่ควรเกี่ยวข้องหรือมีผลต่อกัน เช่น การทำข้อสอบข้อหนึ่งไม่ควรมีผลต่อการตอบข้อสอบข้ออื่น อย่างไรก็ตาม ข้อสอบใน Testlet มักมีความเชื่อมโยงกันโดยธรรมชาติ เนื่องจากใช้สิ่งเร้าร่วมกัน ส่งผลให้การประมาณค่าพารามิเตอร์ข้อสอบ เช่น ความยาก (difficulty) หรืออำนาจจำแนก (discrimination) อาจเกิดความคลาดเคลื่อนได้ ดังนั้น จึงจำเป็นต้องใช้วิธีการวิเคราะห์ที่เหมาะสม เพื่อตรวจสอบและควบคุมผลกระทบที่เกิดขึ้นจาก Testlet (Wainer, Bradlow, & Wang, 2007)

การวิเคราะห์มิติโครงสร้างของแบบทดสอบ เป็นขั้นตอนสำคัญในการตรวจสอบสมมติฐานเกี่ยวกับความเป็นเอกมิติของแบบทดสอบ โดยเฉพาะแบบทดสอบที่วัดความสามารถหลายด้าน เช่น การฟัง โครงสร้างภาษา การอ่าน การใช้โมเดลที่ซับซ้อน เช่น Multidimensional IRT (MIRT) หรือ Bifactor Model จะช่วยแยกแยะความสามารถแต่ละด้านได้อย่างชัดเจน (Reise, 2012) แม้ว่าจะมีประสิทธิภาพในการวิเคราะห์มิติของข้อสอบ แต่โมเดลดังกล่าวไม่ได้ออกแบบมาโดยตรงเพื่อจัดการกับปัญหา Local Dependence ที่เกิดจาก Testlet ซึ่งเป็นข้อจำกัดที่สำคัญ นอกจากนี้ การตรวจสอบความสัมพันธ์ระหว่างข้อสอบใน Testlet และข้อสอบเดี่ยวที่เป็นอิสระจากกัน สามารถทำได้โดยใช้สถิติ Q3 ซึ่งเป็นตัวชี้วัดความสัมพันธ์ระหว่างข้อสอบคู่หนึ่งในกลุ่ม Testlet โดยคำนวณจากส่วนที่เหลือของโมเดล (residuals) หากค่าความสัมพันธ์ Q3 สูงกว่าที่คาดการณ์ บ่งชี้ว่า ข้อสอบมีความสัมพันธ์กันเกินปกติอันเนื่องมาจากการใช้สิ่งเร้าร่วมกัน (Yen, 1983)

การเลือกใช้ Testlet Response Theory (TRT) และ Rasch Testlet Model จึงเป็นวิธีการที่เหมาะสมในการจัดการปัญหา Local Dependence ซึ่งเกิดขึ้นในข้อสอบลักษณะ Testlet ข้อได้เปรียบที่สำคัญของ TRT คือ ความสามารถในการแยกผลกระทบจาก Testlet ออกจากความสามารถที่แท้จริงของผู้สอบ ทำให้การวัดผลมีความแม่นยำมากขึ้น นอกจากนี้ TRT ยังมีความเรียบง่าย และการตีความได้ชัดเจน โดยคงสมมติฐานการวัดมิติเดียวเอาไว้ พร้อมทั้งสามารถประเมินความสามารถ และค่าพารามิเตอร์ของข้อสอบ เช่น ความยาก อำนาจจำแนก และโอกาสเดา ได้อย่างแม่นยำ โดยไม่ขึ้นอยู่กับกลุ่มตัวอย่าง (sample independence) (Hambleton et al., 1991; Wang & Wilson, 2005) ข้อได้เปรียบอีกประการหนึ่งของ TRT คือ ความยืดหยุ่นในการใช้งาน สามารถนำไปใช้กับข้อสอบเดี่ยว และ Testlet ได้หลากหลายประเภท เช่น ข้อสอบการอ่าน และการฟังที่ CTT ไม่สามารถรองรับได้ (Wainer, Bradlow, & Wang, 2007) ยิ่งไปกว่านั้น TRT ยังช่วยตรวจสอบความลำเอียงของข้อสอบ (DIF) เพื่อให้มั่นใจว่าข้อสอบมีความเป็นธรรมสำหรับทุกกลุ่มผู้สอบ อีกทั้ง

ยังสามารถควบคุมปัญหาการละเมิด Local Independence ได้อย่างแม่นยำผ่านการวิเคราะห์ความแปรปรวนภายใน Testlet ทำให้การวิเคราะห์ข้อสอบมีความน่าเชื่อถือสูงกว่าโมเดลอื่นๆ เช่น MIRT และ Bifactor Model ที่เน้นการแยกความสามารถหลายมิติแต่ไม่ได้เน้นการควบคุมความสัมพันธ์ภายใน Testlet นอกจากนี้ TRT ยังสนับสนุนการออกแบบข้อสอบเชิงโครงสร้าง (construct-based testing) ที่สามารถวัดมิติความสามารถซับซ้อนได้อย่างชัดเจน (van der Linden, 2016; Reise, 2012)

แบบทดสอบ STOU-EPT เป็นแบบทดสอบที่ออกแบบมาเพื่อวัดทักษะภาษาอังกฤษใน 3 ด้าน ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน โดยเฉพาะในส่วนของ การฟัง และการอ่าน ข้อสอบเหล่านี้มีลักษณะการใช้สิ่งเร้าร่วมกัน เช่น ในทักษะการฟัง ผู้สอบจะได้ฟังเสียง และทำข้อสอบในรูปแบบต่างๆ ได้แก่ การจับคู่ข้อมูลกับสิ่งที่ได้ยิน จำนวน 6 ข้อ การตอบคำถามแบบเติมคำจากเรื่องที่ได้ยิน จำนวน 6 ข้อ และการตอบคำถามในรูปแบบถูก-ผิด โดยพิจารณาว่าสิ่งที่ได้ยินนั้นถูกต้องหรือไม่ จำนวน 5 ข้อ สำหรับทักษะการอ่าน ข้อสอบจะมีเนื้อเรื่องให้ผู้สอบอ่าน ประมาณ 6 เรื่อง และแต่ละเรื่องจะมีคำถามให้ตอบ ประมาณ 6 ถึง 7 ข้อ ซึ่งข้อสอบในลักษณะดังกล่าว เรียกว่า Testlet หมายถึง กลุ่มข้อสอบที่ใช้สิ่งเร้าร่วมกัน เช่น เสียง หรือข้อความ ส่งผลให้ข้อสอบภายในกลุ่มมีความสัมพันธ์กัน

เนื่องจากลักษณะ Testlet นี้ อาจทำให้ข้อสอบมีปัญหาความสัมพันธ์กันภายในกลุ่ม (local dependence) ดังนั้น การวิจัยครั้งนี้ ผู้วิจัยจึงเลือกใช้ทฤษฎีการตอบสนองข้อสอบ Rasch Testlet Model ในการวิเคราะห์คุณภาพข้อสอบ โมเดลนี้ช่วยจัดการกับปัญหาความสัมพันธ์ของข้อสอบในกลุ่ม Testlet และช่วยให้การประเมินคุณภาพข้อสอบ เช่น ความยาก มีความแม่นยำมากขึ้น

ผลการวิเคราะห์คุณภาพข้อสอบด้วย Rasch Testlet Model นี้ ถูกนำไปให้ผู้เชี่ยวชาญใช้เป็นข้อมูลประกอบการตัดสินใจในการกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดด้วยวิธีการของแองกอฟที่ได้รับการปรับปรุง ในรอบที่ 2

2.4.3 มาตรฐานของการกำหนดคะแนนจุดตัด

มาตรฐานของการทดสอบทางการศึกษาและจิตวิทยา ในส่วนที่ 2.4.1 ได้ให้ความสำคัญกับคะแนนสอบ เนื่องจากคะแนนสามารถบ่งบอกถึงความสามารถของผู้สอบ และการที่จะแบ่งผู้สอบออกเป็น 2 กลุ่ม หรือมากกว่า 2 กลุ่ม ที่สะท้อนระดับความสามารถของผู้สอบที่ต่างกันได้นั้น จำเป็นจะต้องใช้วิธีการกำหนดเกณฑ์ในการตัดสินเกี่ยวกับระดับความสามารถของผู้สอบที่เรียกว่า การกำหนดมาตรฐาน (standard setting) เพื่อหาคะแนนจุดตัด

คะแนนจุดตัด เป็นเกณฑ์ที่ใช้ในการตัดสินผลสอบ หากคะแนนจุดตัดที่หามา มีความคลาดเคลื่อนสูงหรือจำแนกผู้สอบได้ไม่ถูกต้อง ก็จะส่งผลกระทบต่อ การตัดสินผลสอบ ดังนั้น การกำหนดคะแนนจุดตัด ควรเลือกใช้วิธีการที่เหมาะสม และดำเนินการตามวิธีที่เลือกอย่างถูกต้อง ครบถ้วน หรืออาจจะเลือกวิธีกำหนดคะแนนจุดตัดหลายวิธี แล้วพิจารณาว่าวิธีใดให้คะแนนจุดตัดที่

เหมาะสมที่สุด หากกำหนดคะแนนจุดตัดไม่เหมาะสม เช่น คะแนนต่ำหรือสูงกว่าความสามารถจริงของผู้สอบในแต่ละระดับความสามารถ อาจส่งผลกระทบต่อผู้สอบได้ ผลกระทบที่อาจเกิดขึ้นคือ คะแนนจุดตัดที่กำหนดทำให้คนสอบผ่านเกณฑ์ทั้ง ๆ ที่เขาไม่ควรจะสอบผ่าน ในทางกลับกันทำให้คนสอบไม่ผ่านเกณฑ์ทั้ง ๆ ที่เขาควรจะสอบผ่าน อันเนื่องมาจากความคลาดเคลื่อนของการกำหนดคะแนนจุดตัด สอดคล้องกับ สังวรรณ ังคกระโทก (2561) กล่าวว่า จุดตัดที่ดีควรมีความเคลื่อนไหวในการจำแนกผู้สอบต่ำ คือ สามารถจำแนกผู้สอบได้ถูกต้อง

นอกจากนั้น มาตรฐานของการทดสอบทางการศึกษาและจิตวิทยา กล่าวถึงมาตรฐานของคะแนนจุดตัดว่า 1) เมื่อมีการเสนอการแปลความหมายคะแนนที่เกี่ยวข้องกับคะแนนจุดตัด ผู้พัฒนาเครื่องมือประเมินจะต้องสามารถให้เหตุผล และอธิบายกระบวนการในการกำหนดคะแนนจุดตัดได้ชัดเจน 2) เมื่อมีการกำหนดคะแนนจุดตัด เช่น ผ่าน-ไม่ผ่าน หรือระดับความสามารถขึ้นอยู่กับ การตัดสินโดยตรงเกี่ยวกับความเหมาะสมของข้อสอบหรือความสามารถของแบบทดสอบ กระบวนการตัดสินควรถูกออกแบบเพื่อให้ผู้เข้าร่วมในการตัดสินสามารถนำความรู้และประสบการณ์ของตนในการให้เหตุผลในการกำหนดคะแนนจุดตัดอย่างเหมาะสม และ 3) การกำหนดคะแนนจุดตัดที่เป็นไปได้และเหมาะสม คะแนนจุดตัดจะต้องอธิบายความแตกต่างระหว่างกลุ่มผู้สอบได้อย่างแท้จริง การแปลความหมายคะแนนจุดตัดที่แตกต่างกัน ควรมีคำอธิบายที่มีเหตุผลสนับสนุนชัดเจน โดยใช้ข้อมูลเชิงประจักษ์เกี่ยวกับความสัมพันธ์ระหว่างคะแนนสอบกับเกณฑ์ที่กำหนดไว้ (AERA, 2014)

2.5 การกำหนดมาตรฐานเพื่อหาคะแนนจุดตัด

เนื้อหาในส่วนนี้ แบ่งออกเป็น 3 หัวข้อ ได้แก่ 2.5.1 มโนทัศน์เกี่ยวกับมาตรฐาน การกำหนดมาตรฐาน และคะแนนจุดตัด 2.5.2 กระบวนการกำหนดจุดตัดของคะแนน 2.5.3 วิธีการกำหนดคะแนนจุดตัด และ 2.5.4 งานวิจัยที่เทียบเคียงแบบทดสอบความรู้ภาษาอังกฤษด้วยวิธี Modified Angoff กับกรอบ CEFR รายละเอียดแต่ละหัวข้อมีดังนี้

2.5.1 มโนทัศน์เกี่ยวกับมาตรฐาน การกำหนดมาตรฐาน และคะแนนจุดตัด

Hambleton (1978) ได้ให้ความหมายของ มาตรฐาน (standard) หมายถึง คะแนนที่ได้จากการสอบที่ใช้ในการแบ่งผู้สอบออกเป็น 2 กลุ่ม ตามระดับความสามารถที่แตกต่างกัน คือ กลุ่มที่รอบรู้ (masters) และกลุ่มที่ไม่รอบรู้ (non-masters)

การกำหนดมาตรฐานเป็นวิธีการที่ใช้ในการกำหนดระดับของผลสัมฤทธิ์ หรือความสามารถ และคะแนนจุดตัดที่สอดคล้องกับระดับดังกล่าว คะแนนจุดตัดเป็นเพียงคะแนนที่ทำหน้าที่ในการจำแนกกลุ่มคนที่ได้คะแนนต่ำกว่าคะแนนจุดตัด และกลุ่มคนที่ได้คะแนนเท่ากับจุดตัดหรือสูงกว่าคะแนนจุดตัดไปยังระดับถัดไปและระดับที่สูงกว่า (Bejar, 2008) และสิ่งที่ควรคำนึงถึงในการกำหนดมาตรฐาน Millman (1973) กล่าวไว้ 5 ประเด็น ได้แก่ 1) ระดับความสามารถในการทำ

แบบทดสอบของคนอื่น ๆ โดยเฉพาะผลการสอบของกลุ่มที่ผ่านการยอมรับแล้ว ยื่อนำมาใช้เป็นเกณฑ์ในการพิจารณากำหนดมาตรฐาน 2) เนื้อหาสาระของข้อสอบในแบบสอบ เนื้อหาของข้อสอบเป็นส่วนสำคัญส่วนหนึ่งในการที่จะพิจารณากำหนดมาตรฐาน ข้อสอบที่วัดเนื้อหาที่มีความสำคัญต้องให้ผู้สอบทำได้ในสัดส่วนที่สูง แต่ในส่วนของเนื้อหาที่ไม่สำคัญนัก หรือมีความยากเกินไปเกณฑ์มาตรฐานก็ควรลดต่ำลง 3) ผลของการศึกษาที่จะตามมาจากการกำหนดเกณฑ์มาตรฐาน ถ้าหากกำหนดเกณฑ์มาตรฐานต่ำเกินไปอาจมีผลเสีย คือ ทำให้ผู้เรียนไม่สามารถเรียนบทเรียนที่สูงต่อไปได้ แต่ถ้ากำหนดมาตรฐานสูงเกินไป จะมีผลทำให้ผู้ไม่รอบรู้ตกค้างเป็นจำนวนมาก 4) ผลทางจิตวิทยา และสภาพทางการเงิน การกำหนดมาตรฐาน ควรคำนึงถึงผลทางจิตใจที่เกิดขึ้นกับผู้สอบ เช่น แรงจูงใจ ความเบื่อหน่าย ค่าใช้จ่ายในการสอนซ่อมเสริมให้กับผู้ที่ไม่ผ่านเกณฑ์ และ 5) ความคลาดเคลื่อนในการเดา และการสุ่มข้อสอบ การกำหนดมาตรฐานควรคำนึงถึงโอกาสที่ผู้สอบจะเดาถูกรวมทั้งลักษณะการสุ่มข้อสอบจากประชากรข้อสอบทั้งหมดในขอบเขตเนื้อหาที่กำหนดขึ้นด้วย

Berk (1986) กล่าวว่า คะแนนจุดตัด เป็นจุดที่แบ่งคะแนนของกลุ่มที่จัดว่าเป็นผู้รอบรู้ กับคะแนนของกลุ่มที่จัดว่าเป็นผู้ไม่รอบรู้ คะแนนจุดตัดเป็นค่าที่ได้จากการสังเกต ดังนั้น การกำหนดคะแนนจุดตัด ย่อมมีความคลาดเคลื่อนมาตรฐาน (standard error) ความคลาดเคลื่อน แบ่งออกเป็น 2 ประเภท ได้แก่ ประเภทที่ 1 ผู้สอบที่มีความรอบรู้แท้จริง แต่ถูกจัดประเภทไม่มีความรอบรู้ ความคลาดเคลื่อนลักษณะนี้ เกิดจากการกำหนดคะแนนจุดตัดสูงเกินไป ทำให้เป็นผู้สอบที่มีความรอบรู้สอบไม่ผ่านเกณฑ์ที่กำหนด และประเภทที่ 2 ผู้สอบที่ไม่มีความรอบรู้อย่างแท้จริง แต่ถูกจัดประเภทมีความรอบรู้ ความคลาดเคลื่อนลักษณะนี้ เกิดจากการกำหนดคะแนนจุดตัดต่ำเกินไป ทำให้ผู้สอบที่ไม่มีความรอบรู้ผ่านเกณฑ์ที่กำหนด

ดังนั้น การกำหนดคะแนนจุดตัด ควรกำหนดคะแนนแต่ละกลุ่ม โดยอาจจะแบ่งออกเป็น 2 กลุ่ม หรือมากกว่า 2 กลุ่ม ที่สะท้อนระดับความสามารถของผู้สอบหรือผู้เรียนที่ต่างกันได้อย่างชัดเจน

2.5.2 กระบวนการกำหนดจุดตัดของคะแนน

กระบวนการกำหนดจุดตัดของคะแนน เพื่อใช้จำแนกผู้สอบออกเป็นกลุ่ม ๆ ตามคะแนนวัดความรู้ของผู้สอบ โดยทั่วไปการกำหนดจุดตัดเป็นกระบวนการระดมความคิดเห็นและข้อเสนอแนะจากผู้เชี่ยวชาญ ซึ่งขั้นตอนการกำหนดจุดตัด มี 12 ขั้นตอน ดังนี้ (สังวรณัฏ จัดกระโทก, 2561)

1. กำหนดเป้าหมายของการกำหนดจุดตัด ผู้ทำการทดสอบต้องกำหนดเป้าหมายของการทดสอบให้ชัดเจนว่าต้องการใช้ผลการทดสอบเพื่ออะไร การกำหนดเป้าหมายของการทดสอบมีส่วนสำคัญในการเลือกเนื้อหา และความยากง่ายของการทดสอบเพื่อให้การทดสอบนั้นบรรลุเป้าหมายของการจัดสอบ ผู้ทำการทดสอบต้องพิจารณาว่า จำเป็นต้องหาจุดตัดสำหรับประเมินครั้งนั้นหรือไม่

ถ้าต้องการจุดตัดจะต้องระบุเป้าหมายของการใช้จุดตัดให้ชัดเจนว่าต้องการใช้เพื่อจุดมุ่งหมายใด ซึ่งเป้าหมายการกำหนดจุดตัดมีความสำคัญมาก เพราะมีผลต่อการพิจารณาว่าควรกำหนดจุดตัดให้สูงหรือต่ำ ดังนั้น ขั้นตอนแรกของกระบวนการกำหนดจุดตัด ต้องมีการชี้แจงผู้เชี่ยวชาญและผู้เข้าร่วมให้เข้าใจเป้าหมายของการกำหนดจุดตัด เพื่อให้ทุกคนเข้าใจกรอบการทำงานตรงกันตั้งแต่เริ่มต้น โดยเป้าหมายการกำหนดจุดตัด ได้แก่ 1) ให้คำแนะนำ 2) หาตัวอย่างหรือแนวปฏิบัติที่เป็นเลิศ 3) ประเมินคุณภาพการศึกษา และ 4) รับรองความรู้

2. การคัดเลือกผู้เชี่ยวชาญ การกำหนดจุดตัดเป็นการใช้กลุ่มผู้เชี่ยวชาญมาร่วมกัน กำหนดจุดตัดสำหรับการประเมิน จุดตัดที่ได้จากการระดมความคิดเห็นไม่ใช่จุดตัดที่ถูกต้องสมบูรณ์ แต่เป็นความคิดเห็นของผู้เชี่ยวชาญ การหาจุดตัดโดยกลุ่มผู้เชี่ยวชาญที่มีลักษณะต่างกันจะได้ค่าจุดตัดต่างกัน ดังนั้น คุณลักษณะของผู้เชี่ยวชาญจึงมีความสำคัญต่อการกำหนดจุดตัด

3. การเลือกวิธีการกำหนดจุดตัด วิธีการกำหนดจุดตัดมีหลายวิธี แต่ละวิธีมีข้อดี และความซับซ้อนแตกต่างกัน ผู้ใช้ควรเลือกให้เหมาะกับสิ่งที่ต้องการประเมิน รวมทั้งควรเลือกให้เหมาะกับงบประมาณ และทรัพยากร

4. การกำหนดระดับความสามารถของผู้สอบ การกำหนดมาตรฐานผลงาน มี 3 องค์ประกอบ ได้แก่ จำนวนและชื่อของระดับความสามารถ คำอธิบายความสามารถแต่ละระดับ และคะแนนจุดตัดที่ใช้จำแนกผู้สอบออกเป็นกลุ่มตามความสามารถของผู้สอบ การกำหนดมาตรฐานผลงานมีความสำคัญต่อการรายงานผลการประเมินอิงมาตรฐาน ในการรายงานอิงมาตรฐานต้องรายงานผลอย่างละเอียด เพื่อให้ผู้ใช้ผลการประเมินทราบว่า ผู้ที่ผ่านหรือไม่ผ่านมาตรฐานมีความรู้ความสามารถใบบ้าง ผู้ที่ไม่ผ่านมาตรฐานควรทราบว่าจะบกพร่องเรื่องใด ซึ่งการรายงานผลประเมินที่ชัดเจนช่วยให้ผู้ใช้ผลประเมินนำผลประเมินไปใช้ และปรับปรุงการเรียนการสอนได้อย่างเหมาะสม

5. กำหนดคำอธิบายระดับความสามารถ ขั้นตอนนี้เป็นกรเขียนอธิบายระดับความสามารถต่าง ๆ การเขียนอธิบายระดับความสามารถของผู้สอบแต่ละระดับจะต้องเขียนให้แตกต่างกันอย่างชัดเจน ซึ่งมีลักษณะเหมือนกับการทำเกณฑ์การให้คะแนนรูปิก

6. การฝึกอบรมผู้เชี่ยวชาญเกี่ยวกับการกำหนดจุดตัด ขั้นตอนนี้จะดำเนินการก่อนการกำหนดจุดตัด เพื่อให้ความรู้กับผู้เชี่ยวชาญเกี่ยวกับแนวทางการกำหนดจุดตัดของวิธีที่เลือกใช้ รวมทั้งให้ผู้เชี่ยวชาญเข้าใจวิธีการทำงาน และทำความเข้าใจของผู้เชี่ยวชาญให้ตรงกัน นอกจากนี้ระหว่างกรฝึกอบรมควรให้ผู้เชี่ยวชาญได้ฝึกซ้อมการกำหนดจุดตัด เปิดโอกาสให้ผู้เชี่ยวชาญได้ซักถาม แลกเปลี่ยนความรู้ และวิจารณ์ผลลัพธ์ของการกำหนดจุดตัด

7. การรวบรวมคำตอบหรือความคิดเห็นจากผู้เชี่ยวชาญ หลังจากผู้เชี่ยวชาญได้ฝึกอบรมการกำหนดจุดตัดแล้ว จะต้องรวบรวมผลการตอบของผู้เชี่ยวชาญมาวิเคราะห์ เพื่อประเมินว่าผู้เชี่ยวชาญมีความคิดเห็นสอดคล้องกันเพียงใด โดยผลการตอบมักจะนำมาหาค่ากลาง เช่น ค่าเฉลี่ย หรือมัธยฐาน และส่วนเบี่ยงเบนมาตรฐาน

8. การนำเสนอผลให้ผู้เชี่ยวชาญทราบ และให้ข้อมูลย้อนกลับ ผู้ดำเนินงานจะนำผลการตอบของผู้เชี่ยวชาญมาวิเคราะห์ และนำเสนอผลวิเคราะห์ให้ผู้เชี่ยวชาญทราบ โดยผู้เชี่ยวชาญจะทราบผลการตอบของตนเอง และผลการตอบของกลุ่มโดยภาพรวม และมีการเปิดโอกาสให้ผู้เชี่ยวชาญแลกเปลี่ยนความรู้ โดยอาจจะให้ผู้เชี่ยวชาญที่ตอบต่างจากคนอื่น อธิบายเหตุผลของการตอบว่าทำไมถึงตอบเช่นนั้น และให้โอกาสผู้เชี่ยวชาญท่านอื่น ๆ ร่วมแลกเปลี่ยนประสบการณ์กัน

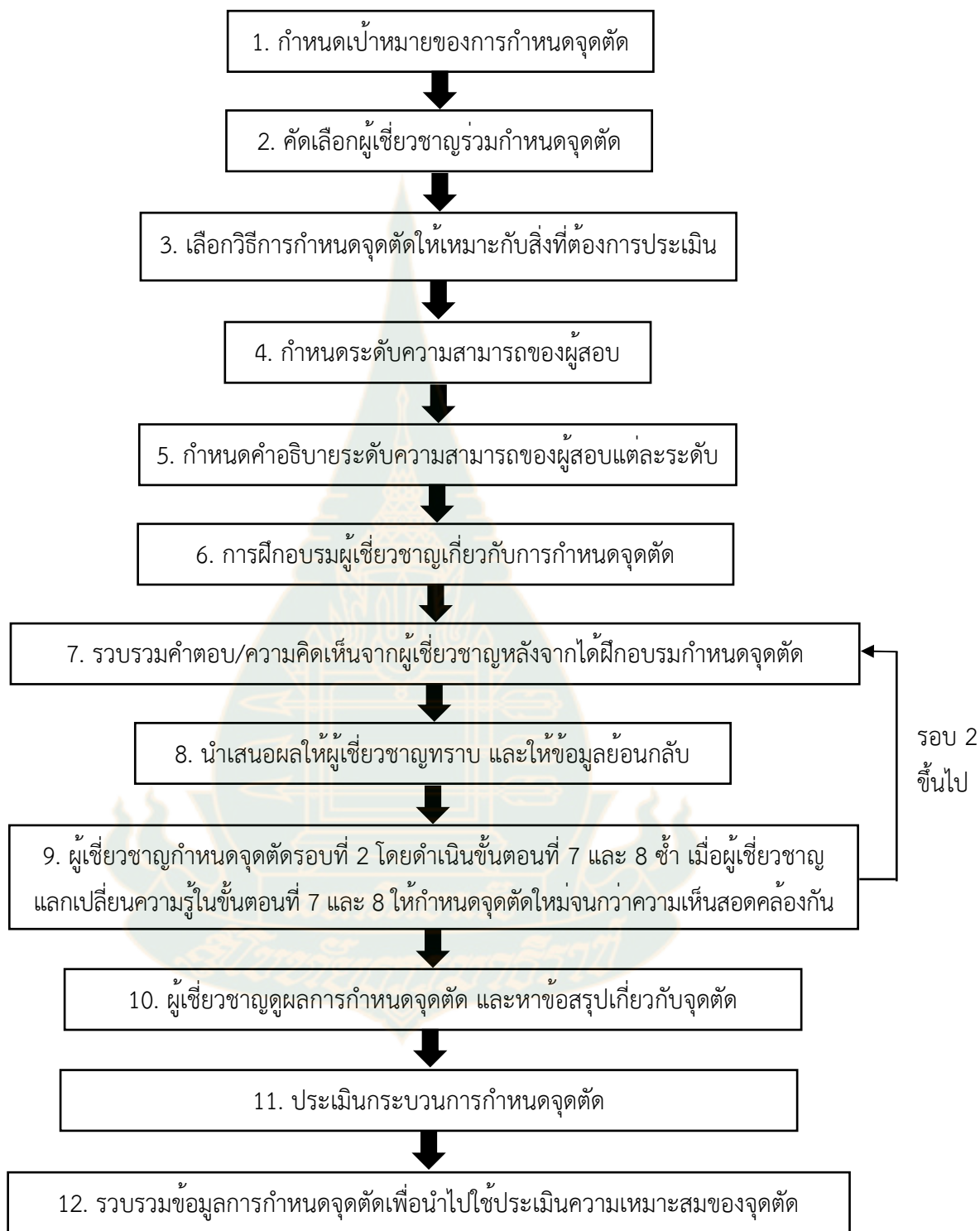
9. การให้ผู้เชี่ยวชาญกำหนดจุดตัดรอบที่ 2 และดำเนินงานขั้นตอนที่ 7 และขั้นตอนที่ 8 ซ้ำ เมื่อผู้เชี่ยวชาญแลกเปลี่ยนความรู้กันในขั้นตอนที่ 7 และ 8 เสร็จแล้ว ก็ให้ผู้เชี่ยวชาญแต่ละคนกำหนดจุดตัดใหม่อีกรอบ จนกว่าความเห็นของผู้เชี่ยวชาญจะสอดคล้องกันหรือมีฉันทามติเกิดขึ้น

10. การให้ผู้เชี่ยวชาญดูผลการกำหนดจุดตัด และหาข้อสรุปเกี่ยวกับจุดตัด หลังจากผู้เชี่ยวชาญทำการกำหนดจุดตัดเสร็จแล้ว จะนำผลการตอบของผู้เชี่ยวชาญแต่ละคนมาหาจุดตัดของกลุ่ม คือ ค่ากลาง และส่วนเบี่ยงเบนมาตรฐาน ซึ่งเป็นค่าสถิติของความเห็นของผู้เชี่ยวชาญรอบสุดท้ายที่จะนำไปใช้เป็นจุดตัด

11. การประเมินกระบวนการกำหนดจุดตัด เป็นการประเมินความเหมาะสมของกระบวนการกำหนดจุดตัดที่ได้ดำเนินการแล้ว สามารถทำได้ 3 วิธี ได้แก่ 1) การให้ผู้เชี่ยวชาญที่อยู่ในกระบวนการกำหนดจุดตัดประเมินความเหมาะสมของขั้นตอนการทำงาน 2) การให้ผู้ประเมินภายนอกที่ไม่ได้อยู่ในกระบวนการกำหนดจุดตัดประเมินความเหมาะสมของขั้นตอนการทำงาน และ 3) การให้ผู้เชี่ยวชาญที่อยู่ในกระบวนการกำหนดจุดตัด และผู้ประเมินภายนอกที่ไม่ได้อยู่ในกระบวนการกำหนดจุดตัดประเมินความเหมาะสมของขั้นตอนการทำงาน

12. การรวบรวมข้อมูลที่เกี่ยวข้องกับการกำหนดจุดตัดเพื่อนำไปใช้ประเมินความเหมาะสมของจุดตัดที่ได้ หลังจากกระบวนการกำหนดจุดตัดเสร็จสิ้น ผู้ที่กำหนดจุดตัดควรเก็บรวบรวมข้อมูลต่าง ๆ ที่จำเป็นสำหรับการประเมินความเหมาะสมของจุดตัดที่ได้จากกระบวนการกลุ่มของผู้เชี่ยวชาญ เช่น คุณสมบัติและจำนวนผู้ร่วมกำหนดจุดตัด ระยะเวลาในการกำหนดจุดตัด

โดยสรุป กระบวนการกำหนดจุดตัด มี 12 ขั้นตอน แสดงดังภาพที่ 2



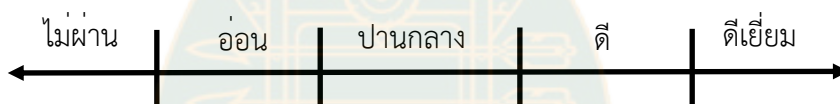
ภาพที่ 2 กระบวนการกำหนดจุดตัด

2.5.3 วิธีการกำหนดคะแนนจุดตัด

นักวิชาการจำนวนมากได้พัฒนาวิธีการกำหนดคะแนนจุดตัดที่มีความน่าเชื่อถือและแม่นยำ ในการจำแนกผู้สอบตามระดับความสามารถ วิธีการกำหนดคะแนนจุดตัดมีหลากหลายวิธีแต่ละวิธีมีขอบเขตการนำไปใช้ จุดเด่น และข้อจำกัดที่แตกต่างกัน (Barman, 2008) เช่น วิธีแองกอฟ (Angoff Method) วิธีบุ๊กมาร์ค (Bookmark Method) วิธีกลุ่มตรงข้าม (Contrasting Groups Method) วิธีเลือกงานผู้สอบ (Examinee Paper Selections Method) วิธีองค์รวม (Holistic Method) Plake and Cizek (2012) กล่าวว่า วิธีแองกอฟ เป็นวิธีที่นิยมนำมาใช้ในการกำหนดคะแนนจุดตัดอย่างแพร่หลาย วิธีการนี้พัฒนาโดย Angoff (1971) เป็นวิธีที่อาศัยหลักการของความน่าจะเป็นที่ผู้สอบซึ่งมีสมรรถภาพขั้นต่ำสุดที่จะยอมรับได้ตอบข้อสอบถูก โดยนำไปให้ผู้เชี่ยวชาญด้านเนื้อหาวิชาพิจารณาค่าความน่าจะเป็นในการตอบข้อสอบจะถูกนำมาใช้เป็นคะแนนจุดตัด

กระบวนการกำหนดคะแนนจุดตัดด้วยวิธีของแองกอฟ ประกอบด้วย 7 ขั้นตอน ดังนี้ (สังวรณ ังตกระโทก, 2561)

1. กำหนดจำนวนระดับความสามารถ เป็นการจำแนกกลุ่มผู้สอบตามผลการสอบออกเป็นกลุ่ม ๆ ซึ่งการกำหนดกลุ่มผู้สอบนี้จะช่วยให้เราทราบว่าต้องหาจุดตัดกี่จุด เช่น ไม่ผ่าน อ่อน ปานกลาง ดี ดีเยี่ยม ดังนั้น ในการหาจุดตัดจะต้องหา 4 จุด แสดงดังภาพที่ 3



ภาพที่ 3 ตัวอย่างการจำแนกกลุ่มผู้สอบ

2. อธิบายคุณลักษณะของผู้สอบที่มีความสามารถ ณ จุดคาบเกี่ยวระหว่างความสามารถสองระดับ เช่น ผู้สอบที่มีความสามารถก้ำกึ่งระหว่าง “ไม่ผ่าน” และ “อ่อน” ผู้สอบที่มีความสามารถระดับก้ำกึ่ง (borderline examinees) หรือผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบหรือระดับก้ำกึ่งระหว่างความสามารถสองระดับ เช่น ในการที่จะจำแนกผู้เรียนออกเป็นกลุ่มไม่ผ่านและกลุ่มอ่อน แสดงดังภาพที่ 4 ผู้เชี่ยวชาญจะต้องร่วมกันอธิบายลักษณะของผู้สอบที่มีความสามารถในระดับอ่อนขั้นต่ำที่สุดที่ยอมรับได้ ซึ่งเป็นผู้สอบที่มีความสามารถระดับก้ำกึ่ง หรือเป็นผู้สอบที่มีความสามารถของเขาอยู่ระหว่างบริเวณชายขอบของระดับความสามารถสองประเภทที่กำลังพิจารณาหาจุดตัด หรือเพื่อให้เข้าใจง่ายขึ้น คือ เมื่อนำผู้สอบที่มีระดับผลการเรียนอ่อนมาเรียงลำดับด้วยคะแนนสอบที่เขาทำได้ ผู้สอบคนที่อยู่ในกลุ่มที่มีความสามารถระดับอ่อนคนสุดท้ายมีลักษณะอย่างไร คุณลักษณะของผู้สอบกลุ่มนี้เป็นกลุ่มที่ผู้เชี่ยวชาญต้องอธิบายให้ชัดเจน ในทางทฤษฎีเชื่อว่าหากรู้

คะแนนของผู้สอบ ณ จุดนี้ จะถือว่าคะแนนของพวกเขา คือ จุดตัดที่สามารถใช้แยกกลุ่มผู้สอบกลุ่มไม่ผ่าน และอ่อน ออกจากกัน



ภาพที่ 4 การระบุผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบของความสามารถระดับไม่ผ่านกับอ่อน

3. ผู้เชี่ยวชาญอ่านข้อสอบทีละข้อ เพื่อทำความเข้าใจเนื้อหา และความยากง่ายของข้อสอบ ในการหาจุดตัดนั้น ผู้เชี่ยวชาญจะต้องหาจุดตัดทีละครั้ง เช่น หากต้องการหาจุดตัดระหว่าง “ไม่ผ่าน” กับ “อ่อน” ผู้เชี่ยวชาญแต่ละคนจะต้องอ่านข้อสอบทีละข้อ พร้อมกับจินตนาการว่า หากมีผู้สอบที่มีความสามารถระดับอ่อนที่ยอมรับได้ จำนวน 100 คน คนเหล่านี้จะตอบข้อสอบแต่ละข้อถูกกี่คน ทำเช่นนี้ให้ครบทุกข้อ

4. นำผลการตอบของผู้เชี่ยวชาญมาหาค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ซึ่งค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานนี้ เป็นสถิติของการประเมินเพื่อหาจุดตัด ในรอบที่ 1 โดยจุดตัดคำนวณจากค่าเฉลี่ยรวมรอบที่ 1 \times จำนวนข้อสอบ

5. เปิดโอกาสให้ผู้เชี่ยวชาญแสดงความคิดเห็นเกี่ยวกับผลประเมิน

6. หาจุดตัดรอบที่ 2 ตามขั้นตอนที่ 3 ถึงขั้นตอนที่ 5 โดยค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานที่ได้ จะเป็นค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ในรอบที่ 2

7. คำนวณจุดตัด การคำนวณจุดตัดจะใช้ค่าเฉลี่ยความคิดเห็นของผู้เชี่ยวชาญ ในรอบที่ 2 ซึ่งเป็นสัดส่วนของจำนวนข้อสอบทั้งหมดที่ผู้มีความสามารถ ณ จุดตัด ควรได้ โดยจุดตัดคำนวณจาก ค่าเฉลี่ยรวมรอบ 2 \times จำนวนข้อสอบ

ต่อมาวิธีการของแองกอฟแบบดั้งเดิมได้รับการปรับปรุงโดย Livingston and Zieky (1982 อ้างถึงใน สุรีพร อนุศาสนนันท์, 2550) เรียกว่า วิธีการของแองกอฟที่ได้รับการปรับปรุง เนื่องจากต้องการเพิ่มความเที่ยงของผู้ตัดสินในการประมาณโอกาสการตอบข้อสอบถูก โดยเพิ่มความสอดคล้องในการตัดสินของผู้ตัดสิน (intra-judge consistency) และความสอดคล้องในการตัดสินระหว่างผู้ตัดสิน (inter-judge consistency)

Reckase (2000a, cite in Brabdin, 2004 อ้างถึงใน สุรีพร อนุศาสนนันท์, 2550) กล่าวว่า วิธีแองกอฟได้รับการปรับปรุงอีกหลายครั้ง แต่ยังไม่มียุติเกี่ยวกับกระบวนการของแองกอฟ แต่สามารถสรุปขั้นตอนกว้าง ๆ ได้แก่ 1) คัดเลือกผู้เชี่ยวชาญ (panelists) ที่มีความรู้

เกี่ยวกับเนื้อหาข้อสอบ 2) ฝึกอบรมผู้เชี่ยวชาญ เกี่ยวกับวิธีการพิจารณาคะแนนจุดตัด 3) กำหนดนิยามของระดับความสามารถของผู้สอบ 4) ให้ผู้เชี่ยวชาญประเมินความสามารถขั้นต่ำที่ผู้สอบควรมีเพื่อให้ได้ค่าความน่าจะเป็นของการตอบถูกในแต่ละข้อสอบ จนครบทุกข้อ 5) อภิปรายและปรับแก้การประเมินในแต่ละรอบ โดยใช้ข้อมูลประกอบ เช่น ค่าความยากของข้อสอบ 6) ดำเนินการประมาณ 2-3 รอบเพื่อให้ได้ค่าประมาณที่มีความน่าเชื่อถือ และ 7) คำนวณคะแนนจุดตัดสุดท้ายจากข้อมูลที่ได้

อย่างไรก็ตาม แม้ว่ามีการพัฒนาวิธีการกำหนดคะแนนจุดตัดในรูปแบบอื่น ๆ เช่น วิธีใช่/ไม่ใช่ของแองกอฟ (Yes/No Angoff Method) (Impara & Plake, 1997) วิธีปรับขยายของแองกอฟ (Extended Angoff Method) (Hambleton & Plake, 1995) เพื่อให้เหมาะกับลักษณะของข้อสอบที่แตกต่างกัน แต่โดยภาพรวมวิธีการของแองกอฟที่ได้รับการปรับปรุงยังคงได้รับความนิยมมากที่สุด โดยเฉพาะกรณีข้อสอบแบบเลือกตอบ ซึ่งต้องการความแม่นยำในการจำแนกระดับความสามารถของผู้สอบ

2.5.4 งานวิจัยที่เทียบเคียงแบบทดสอบความรู้ภาษาอังกฤษด้วยวิธีการของแองกอฟที่ได้รับการปรับปรุงกับกรอบ CEFR

นอกจากการนำวิธีการของแองกอฟที่ได้รับการปรับปรุง มาใช้ในการกำหนดคะแนนจุดตัดในข้อสอบทั่วไปแล้ว วิธีนี้ยังถูกนำไปใช้ในกระบวนการเทียบคะแนนแบบทดสอบวัดระดับความรู้ภาษาอังกฤษกับกรอบ CEFR ตัวอย่าง

Tannenbaum and Baron (2011) ได้เทียบคะแนนแบบทดสอบ TOEFL ITP กับกรอบ CEFR โดยใช้วิธีการของแองกอฟที่ได้รับการปรับปรุง ซึ่งในการตัดสินความสามารถในแต่ละระดับ ใช้ผู้เชี่ยวชาญ จำนวน 16 คน เป็นเพศหญิง 11 คน และเพศชาย 7 คน ซึ่งมีประสบการณ์ในการสอนภาษาอังกฤษ และคุ้นเคยกับแบบทดสอบ TOEFL ในการหาคะแนนจุดตัดของข้อสอบแต่ละข้อเทียบกับระดับความสามารถด้านภาษาอังกฤษตามกรอบ CEFR ระดับ A2 B1 และ B2 ดำเนินการ 3 รอบ โดยวิธีการของแองกอฟที่ได้รับการปรับปรุง เป็นวิธีที่พิจารณาว่าผู้สอบที่มีความสามารถต่ำสุดในแต่ละระดับมีโอกาสทำข้อสอบข้อที่ i ได้กี่เปอร์เซ็นต์ โดยทำเช่นนี้ให้ครบทุกข้อ จากนั้นจึงนำคะแนนจุดตัดแต่ละรอบมาหาค่าเฉลี่ย สำหรับการหาคะแนนจุดตัดระดับ C1 ใช้วิธีประมาณการจากการเทียบคะแนนกับข้อสอบ TOEFL PBT และ TOEFL CBT ผลการศึกษา พบว่า ผู้สอบได้คะแนนระดับต่าง ๆ เมื่อเทียบกับกรอบ CEFR สรุปได้ดังนี้

ระดับ	การฟัง	โครงสร้างและการเขียน	การอ่าน	รวม
A2	38	32	31	337
B1	47	43	48	460
B2	54	53	56	543
C1	64	64	63	627

Baron and Papageorgiou (2016) ได้ทำคะแนนจุดตัดแบบทดสอบความสามารถทางภาษาสำหรับการจัดตำแหน่งผู้เรียนที่เรียนภาษาอังกฤษเป็นภาษาที่สอง (English as a Second Language: ESL) ในการศึกษาระดับมัธยมศึกษา การศึกษานี้มีจุดมุ่งหมายเพื่อหาคะแนนจุดตัดของแบบทดสอบ TOEFL Junior เพื่อใช้ในการตัดสินใจ 3 ประเด็น ได้แก่ 1) การปฏิเสธการรับเข้าศึกษาไม่รับผู้เรียนเพราะพวกเขาไม่มีทักษะภาษาอังกฤษเพียงพอ 2) รับเข้าแบบมีเงื่อนไข/ต้องได้รับการสนับสนุน ESL เนื่องจากคาดว่าผู้เรียนจะสามารถพัฒนาทักษะภาษาอังกฤษในระหว่างที่กำลังศึกษา หากผู้เรียนได้รับการสนับสนุน ESL และ 3) รับเข้าแบบไม่มีเงื่อนไข/ไม่มีการสนับสนุน ESL เนื่องจากผู้เรียนมีทักษะด้านภาษาเพียงพอในการเรียนการสอน โดยการสอบ TOEFL Junior เป็นการประเมินความสามารถทางภาษาอังกฤษที่ใช้ในชีวิตประจำวันและในห้องเรียน สำหรับผู้เรียนอายุ 11 ปีขึ้นไป ที่เรียนภาษาอังกฤษเป็นภาษาที่สอง การกำหนดมาตรฐานเพื่อหาคะแนนจุดตัด โดยใช้วิธีการของแองกอฟที่ได้รับการปรับปรุง และมีคณะกรรมการกำหนดมาตรฐาน จำนวน 16 คน เป็นผู้เชี่ยวชาญด้านภาษาจากโรงเรียนเอกชนใน 12 รัฐของสหรัฐอเมริกา ในกระบวนการนี้ได้ทำการกำหนดคะแนนจุดตัดทั้งหมด 3 รอบ โดยในแต่ละรอบมีการคำนวณค่าสถิติ ได้แก่ ค่าเฉลี่ย ค่ามัธยฐาน ค่าต่ำสุด ค่าสูงสุด ส่วนเบี่ยงเบนมาตรฐาน และค่าความคลาดเคลื่อนมาตรฐานของการตัดสินใจ ผลการวิจัยพบว่า คะแนนจุดตัดของแบบทดสอบ TOEFL Junior ในแต่ละส่วน (part) แสดงดังตารางที่ 6

ตารางที่ 6 คะแนนจุดตัดแบบทดสอบ TOEFL Junior ในแต่ละส่วนกับ EST

Test	TOEFL Junior Comprehensive	Needs ESL Support		No ESL Support	
		Raw ^a score	Scale score	Raw score	Scale score
TOEFL Junior	Listening comprehension	9	220	25	285
Standard	Language form and meaning	9	210	26	280
	Reading comprehension	9	210	25	275
TOEFL Junior	Speaking	8	8	13	13
Comprehensive	Writing	6	6	13	13
	Reading		144 ^b		156 ^b
	Listening		142 ^b		156 ^b

Note: ^aRaw scores have been rounded up to the next achievable whole number score

^bBased on a scale alignment study

ชื่นจิตต์ อธิวรกุล และ จิรดา วุฑฒยากร (2561) ได้เทียบคะแนนของแบบทดสอบ สมितिภาพภาษาอังกฤษแห่งมหาวิทยาลัยศรีนครินทรวิโรฒ หรือ SWU-SET (Srinakharinwirot University - Standardized English Test) กับกรอบมาตรฐาน CEFR ที่ระดับ A2 B1 และ B2 โดยแบบทดสอบ SWU-SET สร้างโดยคณาจารย์ศูนย์ภาษาและบริการวิชาการ มี Test Specification ครอบคลุมบริบทสำคัญ ได้แก่ บริบททั่วไป (general setting) บริบทชีวิตมหาวิทยาลัย (campus life) และบริบทวิชาการ (academic setting) แบบทดสอบนี้มี 5 ตอน แต่ละตอนมี 20 ข้อ รวม 100 ข้อ ประกอบด้วย การฟัง (listening) คำศัพท์ (vocabulary) การใช้ภาษา (usage and functional language) โครงสร้างภาษา (structure) และการอ่าน (reading) ศูนย์ภาษาและบริการวิชาการได้กำหนดระดับช่วงชั้นของแบบทดสอบไว้ 3 ระดับ คือ A2 B1 และ B2 ซึ่งการกำหนดช่วงชั้นนี้เกิดจากประสบการณ์สอนของคณาจารย์ศูนย์ภาษาและบริการวิชาการ จำนวน 14 คน ที่สอนนิสิตมหาวิทยาลัยศรีนครินทรวิโรฒทุกชั้นปีจากทุกคณะ และคณาจารย์มีความเห็นพ้องต้องกันว่าช่วงชั้นความสามารถของนิสิต อยู่ระหว่าง A2 ถึง B2 โดยการเทียบคะแนนของแบบทดสอบ SWU-SET กับกรอบมาตรฐาน CEFR มีวิธีการกำหนดมาตรฐานตามขั้นตอนดังนี้

1. ขั้นตอนกำหนดผู้เชี่ยวชาญ (panelist) ผู้เชี่ยวชาญในการสร้างมาตรฐาน ประกอบด้วยคณาจารย์ประจำของศูนย์ภาษาและบริการวิชาการ วิทยาลัยนานาชาติเพื่อศึกษาความยั่งยืน มหาวิทยาลัยศรีนครินทรวิโรฒ ผู้ที่มีประสบการณ์การสอนและการออกแบบทดสอบภาษาอังกฤษในระดับอุดมศึกษาตั้งแต่ 5 ปี ขึ้นไป รวม 14 คน เป็นเพศหญิง 10 คน และเพศชาย 4 คน ผู้เชี่ยวชาญมีภูมิหลังทางการศึกษาในระดับปริญญาโทและปริญญาเอกในทิศทางเดียวกัน ได้แก่ การสอนภาษาอังกฤษ ภาษาศาสตร์ และการสื่อสารระหว่างวัฒนธรรม ปัจจุบันผู้เชี่ยวชาญทุกคนรับผิดชอบวิชาเดียวกัน คือ วิชาภาษาอังกฤษพื้นฐานให้นิสิตทุกคณะระหว่างชั้นปีที่ 1 และชั้นปีที่ 2 ของมหาวิทยาลัยศรีนครินทรวิโรฒ

2. ขั้นตอนเตรียมความพร้อม (orientation) ผู้ดำเนินการเปิดโอกาสให้ผู้เชี่ยวชาญแสดงความคิดเห็น แลกเปลี่ยนประสบการณ์ในการพัฒนาแบบทดสอบ การทดลองใช้แบบทดสอบ และการปรับปรุงแบบทดสอบ โดยใช้แบบทดสอบ SWU-SET

3. ขั้นการสร้างความชัดเจนเกี่ยวกับวิธีการและการเทียบคะแนน (training session) ในการเทียบคะแนนแบบทดสอบกับกรอบ CEFR ผู้ดำเนินการใช้วิธีการของแองกอฟที่ได้รับการปรับปรุงเนื่องจากเหมาะกับแบบทดสอบแบบเลือกตอบ และกำหนดจุดตัดเพียง 3 จุด ได้แก่ A2 B1 และ B2

ผู้ดำเนินการเริ่มต้นด้วยการอธิบายคำบรรยายคุณลักษณะในระดับต่าง ๆ ของกรอบ CEFR เพื่อให้ผู้เชี่ยวชาญเห็นภาพรวมของความสามารถทางภาษาของนิสิตอย่างชัดเจน พร้อมทั้งเปิดโอกาสให้ผู้เชี่ยวชาญที่มีประสบการณ์สอนในคณะต่าง ๆ ได้อภิปรายและแลกเปลี่ยนข้อมูลเกี่ยวกับระดับความสามารถของนิสิตในการใช้ภาษาอังกฤษ ซึ่งมีความหลากหลายตามบริบทการเรียน

และรายวิชาที่แตกต่างกัน กระบวนการดังกล่าวมีจุดมุ่งหมายเพื่อสร้างความเข้าใจร่วมกันเกี่ยวกับ
 ธรรมชาติและสมรรถนะของนิสิตในแต่ละระดับ และมองเห็นภาพของผู้ที่มีความสามารถต่ำสุดของ
 ระดับนั้น (just qualified candidate) ตรงกัน

ทั้งนี้ คำถามสำคัญในการดำเนินงาน เพื่อเทียบคะแนนแบบทดสอบด้วยวิธีการของ
 แอวกอฟที่ได้รับการปรับปรุงในระดับ A2 B1 และ B2 มีดังนี้

นิสิตที่มีความสามารถต่ำสุดในระดับ A2 มีโอกาสจะทำข้อสอบข้อที่ i ได้ถูกต้องกี่
 เปอร์เซ็นต์

นิสิตที่มีความสามารถต่ำสุดในระดับ B1 มีโอกาสจะทำข้อสอบข้อที่ i ได้ถูกต้องกี่
 เปอร์เซ็นต์

นิสิตที่มีความสามารถต่ำสุดในระดับ B2 มีโอกาสจะทำข้อสอบข้อที่ i ได้ถูกต้องกี่
 เปอร์เซ็นต์

ผู้เชี่ยวชาญระบุเปอร์เซ็นต์ที่นิสิตที่มีความสามารถต่ำสุดในระดับนั้น ๆ จะตอบถูก
 เพิ่มขึ้นทีละ 5 เปอร์เซ็นต์ จาก 0, 5, 10, 15, 20, 25, ไปเรื่อย ๆ จนถึง 100 เปอร์เซ็นต์ พร้อมทั้ง
 บันทึกลงในแบบฟอร์มที่กำหนดให้

4. ขั้นตอนวิธีการสร้างมาตรฐานเพื่อหาคะแนนจุดตัด รอบที่ 1

4.1 การเทียบคะแนนระดับ A2 ผู้ดำเนินการเริ่มต้นให้ผู้เชี่ยวชาญ พิจารณาว่า
 “นิสิตที่มีความสามารถต่ำสุดในระดับ A2 มีโอกาสที่จะทำข้อสอบข้อที่ i ได้ถูกต้องกี่เปอร์เซ็นต์” โดย
 พิจารณาข้อสอบทีละข้อ เริ่มจากข้อที่ 1 ไปจนถึงข้อที่ 100 เช่น ผู้เชี่ยวชาญคนที่ i พิจารณาข้อสอบ
 ในระดับ A2 จะสามารถทำข้อนี้ถูกกี่เปอร์เซ็นต์ โดยบันทึกค่าเปอร์เซ็นต์ลงในแบบฟอร์ม

4.2 การเทียบคะแนนระดับ B2 ต่อจากนั้นผู้ดำเนินการได้เข้ามาเทียบคะแนน
 ระดับ B2 เพื่อให้ผู้เชี่ยวชาญได้เห็นภาพ ความสามารถภาษาอังกฤษที่แตกต่างชัดเจนจาก A2 และ B1
 โดยในระดับ B2 นี้มีคำถามสำคัญว่า นิสิตที่มีความสามารถต่ำสุดในระดับ B2 มีโอกาสที่จะทำข้อสอบ
 ข้อที่ i ได้ถูกต้องกี่เปอร์เซ็นต์ โดยพิจารณาข้อสอบทีละข้อ เริ่มจากข้อที่ 1 ไปจนถึงข้อที่ 100

4.3 การเทียบคะแนนระดับ B1 ผู้ดำเนินการได้กลับมาเทียบคะแนนระดับ B1 โดย
 ให้คำแนะนำกับผู้เชี่ยวชาญว่า ระดับคะแนน B1 ควรจะเป็นคะแนนที่อยู่ระหว่างระดับ A2 และ B2
 ทั้งนี้ยังคงให้ยึดคำถามสำคัญว่า นิสิตที่มีความสามารถต่ำสุดในระดับ A2 มีโอกาสที่จะทำข้อสอบข้อที่
 i ได้ถูกต้องกี่เปอร์เซ็นต์ โดยพิจารณาข้อสอบทีละข้อ เริ่มจากข้อที่ 1 ไปจนถึงข้อที่ 100 เช่นเดียวกัน
 โดยบันทึกค่าเปอร์เซ็นต์ลงในแบบฟอร์ม

เมื่อเสร็จสิ้นการพิจารณาแบบทดสอบและเทียบเคียงกับกรอบมาตรฐานฯ
 ทั้งสามระดับ ในรอบที่ 1 แล้วผู้ดำเนินการแสดงค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน คะแนนจุดตัด
 ของแต่ละระดับ

5. ขั้นตอนวิธีการสร้างมาตรฐานเพื่อหาคะแนนจุดตัด รอบที่ 2 ในการดำเนินการเทียบคะแนนในรอบที่ 2 มีขั้นตอนเหมือนกับการเทียบคะแนนในรอบที่ 1 แต่ผู้ดำเนินการได้ให้ข้อมูลสารสนเทศเพิ่มเติม 2 เรื่อง ได้แก่ (1) ค่าของคะแนนจุดตัดในรอบที่ 1 และ (2) ค่าคุณภาพของข้อสอบรายข้อ ประกอบด้วย ค่าความยากง่าย และค่าอำนาจจำแนก เพื่อพิจารณาเพิ่มเติม เมื่อผู้เชี่ยวชาญได้รับข้อมูลสารสนเทศเพิ่มเติม ผู้เชี่ยวชาญสามารถตัดสินใจใหม่ได้ว่า เปอร์เซนต์ที่นิสิตที่มีความสามารถต่ำสุดในระดับนั้น ๆ จะตอบถูกเป็นเท่าไร เช่น ในรอบแรกอาจจะให้เปอร์เซนต์สูง แต่เมื่อได้ข้อมูลเพิ่มเติมในรอบนี้ว่าข้อสอบข้อนั้นมีค่าความยากมาก มีนิติตตอบถูกน้อยมาก ผู้เชี่ยวชาญอาจปรับลดระดับเปอร์เซนต์การตอบถูกลงมาได้ ในทางกลับกันถ้าข้อสอบใดมีค่าความยากไม่มากนัก มีนิติตตอบถูกจำนวนมาก ผู้เชี่ยวชาญอาจเพิ่มระดับเปอร์เซนต์การตอบถูกขึ้นได้

6. ขั้นตอนวิธีการสร้างมาตรฐานเพื่อหาคะแนนจุดตัดรอบที่ 3 ผู้ดำเนินการได้ให้ผู้เชี่ยวชาญพิจารณาถึงความเหมาะสมของคะแนนจุดตัดในระดับ A2 B1 และ B2 พร้อมเปิดโอกาสให้ผู้เชี่ยวชาญอภิปรายแลกเปลี่ยนถึงความเหมาะสมของคะแนนในแต่ละช่วง แสดงให้เห็นว่า การสร้างมาตรฐานเพื่อหาคะแนนจุดตัดในรอบที่ 1 ประเด็นสำคัญ คือ ผู้เชี่ยวชาญทั้ง 14 คน ต้องใช้วิจารณญาณของตนเองในการตัดสินผล ในรอบที่ 2 ผู้เชี่ยวชาญจะได้ข้อมูล สารสนเทศของข้อสอบรายข้อเพิ่มเพื่อช่วยในการตัดสินใจ ส่วนในรอบที่ 3 เปิดโอกาสให้ผู้เชี่ยวชาญ อภิปรายแลกเปลี่ยนถึงความเหมาะสมของคะแนนในแต่ละระดับ ซึ่งการตัดสินผลรอบที่ 3 จะเป็นรอบที่ผู้เชี่ยวชาญมีความเห็นใกล้เคียงกันมากที่สุดเกี่ยวกับจำนวนข้อที่ผู้เข้าสอบแต่ละระดับสามารถทำได้

ผลการเทียบคะแนนแบบทดสอบ SWU-SET กับกรอบ CEFR พบว่า จากจำนวนข้อสอบ 100 ข้อ นิสิตที่มีความสามารถระดับ A2 มีคะแนน 22-49 คะแนน ระดับ B1 มีคะแนน 50-77 คะแนน และระดับ B2 มีคะแนนตั้งแต่ 78 คะแนนขึ้นไป

ต่อมา Papageorgiou et al. (2019) ได้เทียบคะแนนแบบทดสอบ TOEFL iBT กับแบบทดสอบ China's Standards of English Language Ability หรือ CSE โดยการสอบ TOEFL เป็นการประเมินความสามารถทางภาษาอังกฤษที่ได้รับการยอมรับอย่างกว้างขวางมากที่สุดในโลก มีวัตถุประสงค์เพื่อใช้สมัครเข้าศึกษาต่อมากกว่า 130 ประเทศ รวมทั้งออสเตรเลีย แคนาดา นิวซีแลนด์ สหราชอาณาจักร และสหรัฐอเมริกา การสอบ TOEFL เริ่มใช้มาตั้งแต่ ค.ศ. 1964 และมีการปรับปรุงแก้ไขมาหลายครั้ง จนปัจจุบันปรับเปลี่ยนเป็นการสอบ TOEFL iBT การสอบแบบนี้จะสอบกับคอมพิวเตอร์ โดยวัดใน 4 ทักษะ ได้แก่ การอ่าน การฟัง การพูด และการเขียน โดยมีผลคะแนนอยู่ระหว่าง 0 ถึง 120 คะแนน (ทักษะละ 30 คะแนน) ส่วนการสอบ CSE เป็นการทดสอบที่ประเทศจีนพัฒนาขึ้นเพื่อพัฒนาความสามารถด้านภาษาอังกฤษทุกทักษะ มีวัตถุประสงค์ 3 ประการ ได้แก่ (1) ใช้กำหนดและอธิบายความสามารถทางภาษาอังกฤษของผู้เรียนภาษาอังกฤษในประเทศจีน (2) ใช้เป็นข้อมูลอ้างอิงและแนวทางการเรียน การสอน และการประเมินภาษาอังกฤษ และ (3)

ใช้สำหรับการจัดตำแหน่งในระดับโลก โดย CSE แบ่งความสามารถผู้เรียนภาษาอังกฤษของคนในประเทศจีน เป็น 9 ระดับ ใน 3 ชั้นกว้าง ๆ ได้แก่ ชั้นพื้นฐาน ชั้นกลาง และชั้นสูง การวิจัยครั้งนี้จึงมุ่งเน้นไปที่ CSE ระดับที่ 4 ถึงระดับที่ 8 ซึ่งในการกำหนดมาตรฐานของการศึกษาครั้งนี้มีวิธีดำเนินการ ดังนี้

1. กำหนดคะแนนจุดตัด หลังจากที่มีการตรวจสอบความสอดคล้องของแบบทดสอบ TOEFL iBT กับแบบทดสอบ CSE ขั้นตอนต่อมาจะต้องระบุคะแนนขั้นต่ำ หรือคะแนนจุดตัดในการทดสอบแต่ละระดับความสามารถ

2. เลือกข้อมูลเชิงประจักษ์จากการสอบ TOEFL iBT ขั้นตอนนี้เป็นการเตรียมข้อมูล โดยรวบรวมข้อมูลจากผู้สอบ TOEFL iBT ปี ค.ศ. 2017 ในประเทศจีน จำนวน 2,185 คน โดยกลุ่มผู้เข้าสอบจะกระจายทั่วภูมิศาสตร์ ซึ่งถือว่าเป็นตัวแทนของผู้สอบทุกคนที่ทำการสอบในประเทศจีน

3. คัดเลือกผู้เชี่ยวชาญ จากประเทศจีน จำนวน 16 คน เป็นเพศหญิง จำนวน 12 คน และเพศชาย จำนวน 4 คน ทำหน้าที่ในการกำหนดมาตรฐาน โดยผู้ร่วมอภิปรายจะมาจากหลากหลายสถาบัน และก่อนเริ่มการอภิปราย ผู้เชี่ยวชาญทุกคนจะต้องกรอกแบบสอบถามที่ผู้ดำเนินการแจกให้ เช่น ประวัติส่วนตัว ประสบการณ์สอน (ผู้เชี่ยวชาญสอนนักศึกษาระดับปริญญาตรีสาขาวิชาภาษาอังกฤษ หรือสอนนักศึกษาระดับปริญญาตรีที่ไม่ใช่สาขาวิชาภาษาอังกฤษ)

4. เตรียมผู้เชี่ยวชาญก่อนประชุมเชิงปฏิบัติการเพื่อกำหนดมาตรฐาน ก่อนมีการกำหนดมาตรฐานจะต้องจัดทำคู่มือเตรียมการต่าง ๆ และส่งให้กับผู้เชี่ยวชาญ โดยคู่มือมีรายละเอียดเกี่ยวกับการสอบ CSE และการสอบ TOEFL iBT เพื่อให้ผู้เชี่ยวชาญได้ทำความเข้าใจ และมั่นใจว่าผู้เชี่ยวชาญมีความเข้าใจเกี่ยวกับคุณลักษณะที่แยกความแตกต่างของการสอบ CSE ทั้ง 5 ระดับ (ระดับ 4 ถึงระดับ 8) เป็นอย่างดี นอกจากนี้ ยังให้โอกาสผู้เชี่ยวชาญได้ทดลองทำแบบทดสอบก่อนที่จะเริ่มกำหนดมาตรฐาน

5. กำหนดความสามารถระดับก้ำกึ่งของผู้สอบหรือผู้เรียน (borderline student)

6. กำหนดมาตรฐานเพื่อหาจุดตัดของข้อสอบแต่ละข้อ ผู้วิจัยได้ใช้วิธีการของแองกอฟที่ได้รับการปรับปรุง ก่อนที่จะหาจุดตัดก็ให้โอกาสผู้เชี่ยวชาญได้ฝึกฝนการตัดสิน โดยผู้ดำเนินการได้ให้ผู้เชี่ยวชาญตัดสินการทดสอบด้านการอ่านสามรายการแรก และให้หารือเหตุผลของการตัดสิน จากนั้นให้ผู้เชี่ยวชาญกรอกแบบฟอร์มว่าพร้อมที่จะดำเนินการต่อไปหรือไม่ ผู้เชี่ยวชาญทุกคนเห็นว่าให้สามารถดำเนินการต่อไปได้ การวิจัยครั้งนี้กำหนดให้ผู้เชี่ยวชาญดำเนินการตัดสินหรือหาจุดตัด จำนวน 3 รอบ ในแต่ละรอบมีการให้ข้อเสนอแนะ และมีการอภิปรายร่วมกันระหว่างรอบ โดยรอบที่ 1 ให้ผู้เชี่ยวชาญตัดสิน เช่น ผู้สอบที่มีความสามารถต่ำสุดในระดับ CSE4 มีโอกาสที่จะทำข้อสอบข้อที่ i ได้ถูกต้องกี่เปอร์เซ็นต์ จนครบ 100 ข้อ และคำนวณจุดตัดรอบที่ 1 รอบที่ 2 ให้ผู้เชี่ยวชาญตัดสินใหม่อีกครั้ง โดยก่อนมีการตัดสินในรอบที่ 2 จะมีการอภิปรายร่วมกันเกี่ยวกับการกำหนดจุดตัดในรอบที่ 1

รวมถึงให้ค่าความยากของข้อสอบ เพื่อให้ผู้เชี่ยวชาญนำไปใช้ประกอบการตัดสินใจ ในรอบที่ 2 เมื่อประเมินรอบที่ 2 เสร็จก็มีการนำเสนอคะแนนจุดตัดให้แต่ละคนเห็นตามแบบฟอร์ม และรอบที่ 3 ผู้ดำเนินการให้ผู้เชี่ยวชาญตัดสินแบบองค์รวม (holistic judgments) โดยผู้เชี่ยวชาญต้องกำหนดจุดตัดเพียง 1 จุด ในแบบทดสอบแต่ละส่วน เช่น การอ่านเพื่อความเข้าใจ สถิติที่ใช้ในการคำนวณหาจุดตัดในแต่ละรอบ ได้แก่ ค่าเฉลี่ย ค่ามัธยฐาน ค่าต่ำสุด ค่าสูงสุด ส่วนเบี่ยงเบนมาตรฐาน และค่าความคลาดเคลื่อนมาตรฐานของการตัดสินใจ

7. ให้ข้อมูลย้อนกลับและอภิปราย ในขั้นตอนนี้ผู้เชี่ยวชาญแสดงคะแนนจุดตัดที่ได้ครั้งสุดท้าย และมีการหารือเกี่ยวกับความสมเหตุสมผลของการตัดสินใจ ช่วงท้ายของวันสุดท้ายของการดำเนินงาน ผู้เชี่ยวชาญต้องตอบแบบสอบถามเกี่ยวกับกระบวนการดำเนินงาน ความสำคัญของปัจจัยต่าง ๆ ในกระบวนการดำเนินงาน และปัจจัยที่มีอิทธิพลต่อการตัดสินใจ รวมทั้งให้ผู้เชี่ยวชาญระบุระดับความเชื่อมั่นของคะแนนจุดตัดที่กำหนดขึ้น

ผลการวิจัย พบว่า การเทียบผลคะแนนจุดตัดของการสอบ TOEFL iBT กับระดับ CSE แสดงดังตารางที่ 7

ตารางที่ 7 การเทียบผลคะแนนจุดตัดของการสอบ TOEFL iBT กับระดับ CSE

คะแนนสอบ	คะแนนจุดตัด				
	CSE4	CSE5	CSE6	CSE7	CSE8
การอ่าน (0-30)	7	13	17	21	25
การฟัง (0-30)	4	10	16	20	23
การพูด (0-30)	13	17	20	23	26
การเขียน (0-30)	13	17	21	23	27
รวม (0-120)	37	57	74	87	101

หมายเหตุ: CSE คือ China's Standards of English Language Ability

ตัวเลขที่ระบุเป็นคะแนน TOEFL iBT ขั้นต่ำที่ผู้สอบควรได้เพื่อจัดอยู่ในระดับ CSE

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการเทียบเคียงผลคะแนนของแบบทดสอบมาตรฐานที่ใช้วัดความสามารถทางภาษาอังกฤษกับกรอบอ้างอิงต่าง ๆ พบว่า วิธีที่ได้รับคามนิยมอย่างแพร่หลายในการกำหนดคะแนนจุดตัด คือ วิธีการของแองกอฟท์ที่ได้รับการปรับปรุงเนื่องจากเป็นกระบวนการที่สามารถนำไปใช้ได้อย่างมีประสิทธิภาพกับแบบทดสอบที่มีลักษณะเป็นแบบเลือกตอบ ซึ่งการกำหนดคะแนนจุดตัดด้วยวิธีนี้มักแยกพิจารณาตามทักษะ เช่น การฟัง การพูด การอ่าน และการเขียน เพื่อสะท้อนระดับความสามารถเฉพาะด้านของผู้สอบ อย่างไรก็ตาม มีงานวิจัยจำนวนน้อยที่มุ่งเน้นการกำหนดคะแนนจุดตัดโดยรวมทุกทักษะในแบบทดสอบเดียวกัน

สำหรับการเทียบเคียงคะแนนของแบบทดสอบ STOU-EPT กับกรอบ CEFR ผู้วิจัยเลือกใช้วิธีของแองกอฟที่ได้รับการปรับปรุง ซึ่งเป็นการกำหนดจุดตัดเชิงเนื้อหาในระดับรายข้อ โดยให้ผู้เชี่ยวชาญประเมินความน่าจะเป็นที่ “ผู้เข้าสอบซึ่งมีสมรรถนะขั้นต่ำที่ยอมรับได้จะตอบถูกในแต่ละข้อ” แล้วนำค่าที่ประมาณได้มาคำนวณจุดตัดในแต่ละระดับความสามารถ วิธีนี้มีขั้นตอนโปร่งใส ตรวจสอบได้ และมีเหตุผลเชิงเนื้อหารองรับ จึงสนับสนุนการเชื่อมโยงคะแนนของแบบทดสอบ STOU-EPT กับกรอบ CEFR อย่างมีหลักฐานรองรับ นอกจากนี้ เหมาะกับลักษณะข้อสอบ STOU-EPT ที่ให้คะแนนแบบสองค่า (ถูก = 1, ผิด = 0) และสอดคล้องกับแนวปฏิบัติในการทดสอบขนาดใหญ่ เช่น TOEFL iBT ที่ให้ความสำคัญต่อความเคร่งครัดทางระเบียบวิธีและความสามารถในการตีความคะแนน (Cizek & Bunch, 2007; Hambleton & Pitoniak, 2006; Papageorgiou et al., 2019)

การกำหนดคะแนนจุดตัดในงานวิจัยนี้ดำเนินการในสองรูปแบบ ได้แก่ คะแนนจุดตัดรวม ซึ่งพิจารณาผลคะแนนจากทุกทักษะรวมกัน และคะแนนจุดตัดแยกตามทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน การกำหนดคะแนนจุดตัดแยกตามทักษะมีข้อได้เปรียบที่สำคัญ เนื่องจากช่วยให้ผู้สอบหรือผู้ใช้ผลสอบทราบจุดแข็งและจุดที่ต้องการพัฒนาในแต่ละทักษะได้อย่างชัดเจน ซึ่งเป็นข้อมูลที่เป็นประโยชน์ต่อสถาบันการศึกษาในการออกแบบหรือปรับปรุงแนวทางการเรียนการสอนภาษาอังกฤษให้สอดคล้องกับระดับความสามารถที่กำหนดในกรอบ CEFR

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยสรุปกระบวนการกำหนดคะแนนจุดตัดด้วยวิธีการของแองกอฟที่ได้รับการปรับปรุง เป็น 4 ขั้นตอน ดังนี้

1. การกำหนดผู้เชี่ยวชาญ ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดจะต้องมีความรู้ด้านเนื้อหาและประสบการณ์ที่เกี่ยวข้อง มีประสบการณ์ในการสอนหรือออกแบบข้อสอบ รวมทั้งมีความเข้าใจเกี่ยวกับระดับความสามารถที่ใช้กำหนดคะแนนจุดตัด

2. การปฐมนิเทศผู้เชี่ยวชาญก่อนกำหนดคะแนนจุดตัด เพื่ออธิบายแนวทางการกำหนดคะแนนจุดตัดเพื่อสร้างความเข้าใจร่วมกัน เช่น ลักษณะของผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบ อภิปรายเกี่ยวกับผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบ

3. การฝึกอบรมหรือฝึกซ้อมผู้เชี่ยวชาญ เพื่อฝึกอบรมเกี่ยวกับกระบวนการกำหนดจุดตัด และการใช้เกณฑ์ตัดสิน

4. กำหนดมาตรฐานเพื่อหาคะแนนจุดตัดดำเนินการ ประมาณ 3 รอบ ดังนี้

รอบที่ 1 ผู้เชี่ยวชาญอ่านและทำความเข้าใจเนื้อหา และระดับความยากของข้อสอบแต่ละข้อ แล้วจินตนาการว่า “ผู้สอบที่มีความสามารถต่ำสุดจะทำข้อสอบข้อที่ i ถูกกี่เปอร์เซ็นต์” ทำเช่นนี้จนครบทุกข้อ และบันทึกค่าที่ประมาณได้ลงในแบบฟอร์ม

รอบที่ 2 การดำเนินงานในรอบนี้ มีขั้นตอนเช่นเดียวกับการกำหนดคะแนนจุดตัดในรอบที่ 1 แต่ในรอบที่ 2 นี้ มีการให้สารสนเทศประกอบการตัดสินใจเพิ่มเติม เช่น ค่าความยากง่าย

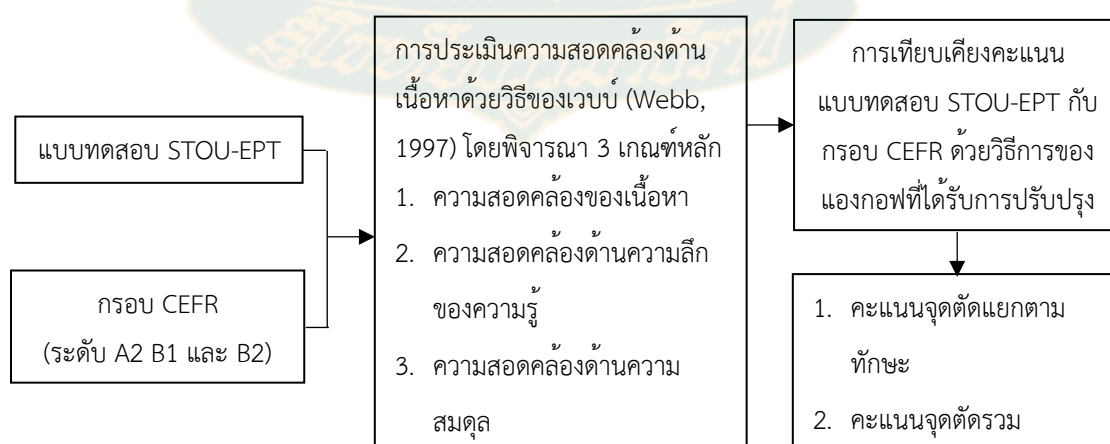
ของข้อสอบ จากนั้นผู้เชี่ยวชาญอ่านและทำความเข้าใจเนื้อหา และระดับความยากของข้อสอบแต่ละข้อ และพิจารณาสารสนเทศประกอบการตัดสินใจเพิ่มเติม แล้วจินตนาการว่า “ผู้สอบที่มีความสามารถต่ำสุดจะทำข้อสอบข้อที่ i ถูกก็เปอร์เซ็นต์” ทำเช่นนี้จนครบทุกข้อ และบันทึกค่าที่ประมาณได้ลงในแบบฟอร์ม

รอบที่ 3 ผู้เชี่ยวชาญพิจารณาความเหมาะสมของคะแนนจุดตัด และเปิดโอกาสให้มีการแลกเปลี่ยนความคิดเห็นถึงความเหมาะสมของคะแนนจุดตัดแต่ละช่วงว่ามีความเป็นไปได้และเหมาะสมหรือไม่ และให้ผู้เชี่ยวชาญได้มีโอกาสปรับค่าคะแนนจุดตัด ในรอบที่ 3 คะแนนจุดตัดสุดท้ายที่ได้จากรอบที่ 3 ถือเป็นค่าที่ใช้สำหรับการตัดสิน

ในแต่ละรอบมีการนำเสนอสถิติ เช่น ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด และค่าความคลาดเคลื่อนมาตรฐานของการตัดสิน เพื่อช่วยในการตัดสินใจของผู้เชี่ยวชาญ

2.6 กรอบแนวคิดในการวิจัย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้กำหนดกรอบแนวคิดในการวิจัยซึ่งมุ่งเน้นการเทียบเคียงคะแนนแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราช (STOU-EPT) กับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรป (CEFR) ในระดับ A2 B1 และ B2 ผู้วิจัยดำเนินการประเมินความสอดคล้องด้านเนื้อหาด้วยวิธีของเว็บ (Webb, 1997) โดยอ้างอิง 3 เกณฑ์หลัก ได้แก่ ความสอดคล้องของเนื้อหา ความสอดคล้องด้านความลึกของความรู้ และความสอดคล้องด้านความสมดุล การประเมินในส่วนนี้อาศัยความเห็นจากผู้เชี่ยวชาญ หลังจากนั้น ได้มีการเทียบเคียงคะแนนแบบทดสอบ STOU-EPT กับกรอบ CEFR ด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง ซึ่งในการกำหนดคะแนนจุดตัด แบ่งออกเป็น 2 แบบ ได้แก่ คะแนนจุดตัดแยกตามทักษะ และคะแนนจุดตัดรวม โดยกรอบแนวคิดในการวิจัย แสดงดังภาพที่ 5



ภาพที่ 5 กรอบแนวคิดการเทียบเคียงคะแนนแบบทดสอบ STOU-EPT กับกรอบ CEFR