

บทที่ 3

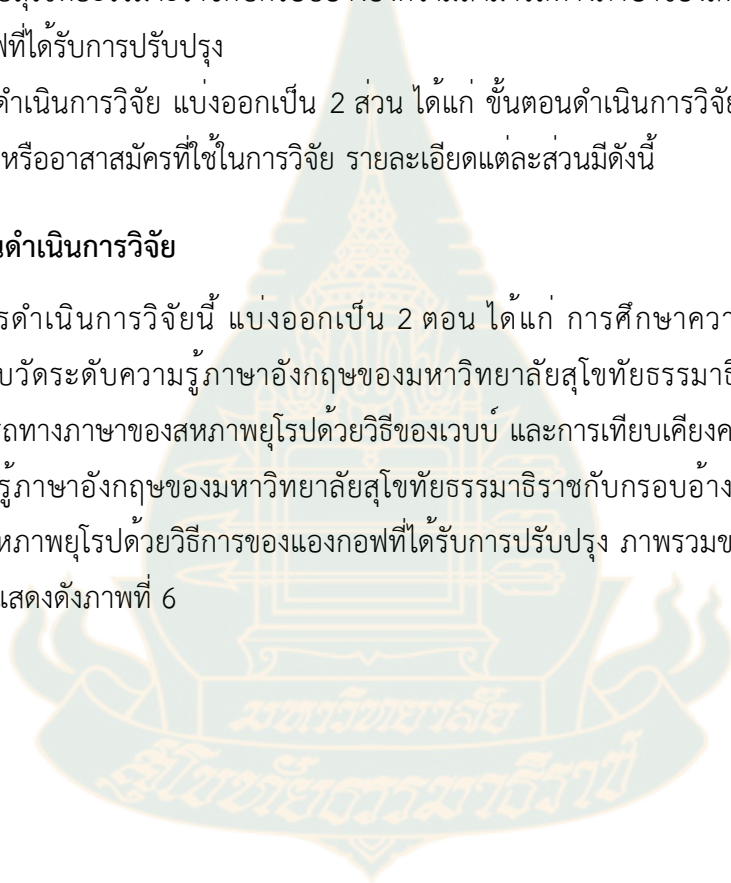
วิธีดำเนินการวิจัย

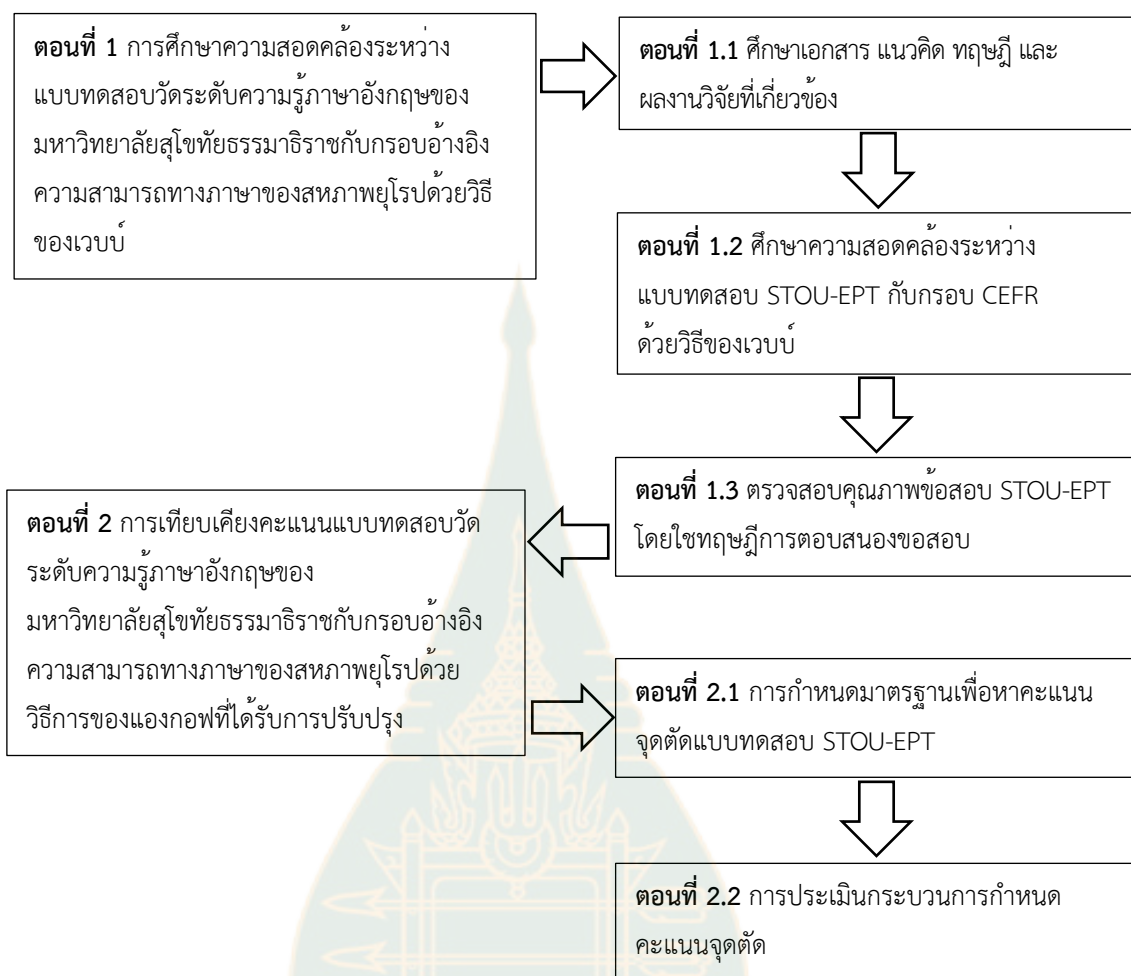
การวิจัยนี้มีวัตถุประสงค์การวิจัยเพื่อศึกษาความสอดคล้องระหว่างแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีของเว็บ และเทียบเคียงคะแนนแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง

วิธีดำเนินการวิจัย แบ่งออกเป็น 2 ส่วน ได้แก่ ขั้นตอนดำเนินการวิจัย และการพิทักษ์สิทธิ์ของตัวอย่างหรืออาสาสมัครที่ใช้ในการวิจัย รายละเอียดแต่ละส่วนมีดังนี้

3.1 ขั้นตอนดำเนินการวิจัย

การดำเนินการวิจัยนี้ แบ่งออกเป็น 2 ตอน ได้แก่ การศึกษาความสอดคล้องระหว่างแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีของเว็บ และการเทียบเคียงคะแนนแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง ภาพรวมของการดำเนินการวิจัยทั้ง 2 ตอน แสดงดังภาพที่ 6





ภาพที่ 6 ขั้นตอนดำเนินการวิจัย

รายละเอียดของการดำเนินการวิจัยแต่ละขั้นตอน มีดังนี้

ตอนที่ 1 การศึกษาความสัมพันธ์ระหว่างแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีของเว็บ

การดำเนินการวิจัย แบ่งออกเป็น 3 ตอนย่อย ได้แก่ ศึกษาเอกสาร แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง ศึกษาความสัมพันธ์ระหว่างแบบทดสอบ STOU-EPT กับกรอบ CEFR ด้วยวิธีของเว็บ และตรวจสอบคุณภาพข้อสอบ STOU-EPT โดยผู้เชี่ยวชาญการตอบสนองข้อสอบ รายละเอียดแต่ละตอนมีดังนี้

ตอนที่ 1.1 ศึกษาเอกสาร แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาเอกสารที่เกี่ยวข้องกับการเทียบเคียงคะแนนแบบทดสอบวัดระดับความรู้ภาษาอังกฤษ โดยพิจารณาจากแหล่งข้อมูลหลากหลายทั้งตำรา งานวิจัย บทความวิชาการ บทความวิจัย และเอกสารอื่น ๆ ที่เกี่ยวข้อง ทั้งในประเทศและต่างประเทศ เพื่อสร้างความเข้าใจเกี่ยวกับหลักการ แนวคิด ทฤษฎี และวิธีการที่ใช้ในการเทียบเคียงคะแนนแบบทดสอบ ตลอดจนกรอบมาตรฐานสากลที่เกี่ยวข้อง อันจะนำไปสู่การออกแบบวิธีวิจัยที่เหมาะสม

ตอนที่ 1.2 ศึกษาความสอดคล้องระหว่างแบบทดสอบ STOU-EPT กับกรอบ CEFR ด้วยวิธีของเว็บ

ในการศึกษาความสอดคล้องระหว่างแบบทดสอบ STOU-EPT กับกรอบ CEFR ผู้วิจัยได้นำวิธีการของเว็บ (Webb's Alignment Method) มาใช้เป็นกรอบในการประเมิน โดยมุ่งเน้นเกณฑ์ประเมินความสอดคล้องด้านเนื้อหา ซึ่งเป็นการพิจารณาเนื้อหาตามมาตรฐาน ประกอบด้วย 3 เกณฑ์หลัก ได้แก่ (1) ความสอดคล้องของเนื้อหา (2) ความสอดคล้องด้านความลึกของความรู้ แบ่งออกเป็น 4 ระดับ ได้แก่ ระดับ 1 ความจำ ระดับ 2 การมีทักษะและมนทัศน์ ระดับ 3 การคิดเชิงกลยุทธ์ และระดับ 4 การคิดหลายขั้นตอน และ (3) ความสอดคล้องด้านความสมดุล ทั้งสามเกณฑ์นี้เป็นหลักฐานสำคัญที่ช่วยยืนยันความน่าเชื่อถือของการตีความคะแนน และสอดคล้องกับแนวคิด Argument-Based Validity ของ Kane (2006) และมาตรฐานสากลของ AERA, APA, and NCME (2014) ซึ่งระบุว่า การมีหลักฐานที่แข็งแกร่งเป็นหัวใจสำคัญของความตรงของแบบทดสอบ อย่างไรก็ตาม มี 2 เกณฑ์ที่ไม่นำมาใช้ ได้แก่ ความกว้างของความรู้ และแหล่งที่มาของความท้าทาย เนื่องจากแบบทดสอบ STOU-EPT ได้พัฒนาโดยอิงกับกรอบ CEFR ซึ่งมีขอบเขตเนื้อหาที่กำหนดไว้อย่างชัดเจน จึงไม่จำเป็นต้องวิเคราะห์ในมิตินี้เพิ่มเติม ส่วนแหล่งที่มาของความท้าทาย ซึ่งใช้เพื่อตรวจสอบว่าความยากของข้อสอบเกิดจากปัจจัยที่ไม่เกี่ยวข้องกับความยากที่ต้องการวัด เช่น อคติทางวัฒนธรรม หรือข้อสอบที่ต้องอาศัยความรู้เฉพาะทาง ก็ไม่จำเป็นต้องนำมาพิจารณาเช่นกัน เนื่องจากแบบทดสอบ STOU-EPT พัฒนาขึ้นโดยอิงกับกรอบ CEFR โดยตรง ทำให้เชื่อมั่นได้ว่าเนื้อหา โครงสร้าง และองค์ประกอบของข้อสอบมีมาตรฐานที่เหมาะสมอยู่แล้ว

การดำเนินงานในตอนที่ 1.2 แบ่งออกเป็น 4 ส่วน ได้แก่ ผู้ให้ข้อมูลหลัก เครื่องมือที่ใช้ในการวิจัย การเก็บรวบรวมข้อมูล และการวิเคราะห์ข้อมูล รายละเอียดแต่ละส่วนมีดังนี้

1.2.1 ผู้ให้ข้อมูลหลัก

ผู้ให้ข้อมูลหลัก (key informant) เป็นผู้เชี่ยวชาญประเมินความสอดคล้อง แบ่งออกเป็น 2 กลุ่ม ได้แก่ กลุ่มที่ 1 ผู้เชี่ยวชาญประเมินความสอดคล้องระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละทักษะกับระดับความลึกของความรู้ตามวิธีของเว็บ และกลุ่มที่ 2 ผู้เชี่ยวชาญประเมินความสอดคล้องระหว่างแบบทดสอบ STOU-EPT กับกรอบ CEFR

กลุ่มที่ 1 ผู้เชี่ยวชาญประเมินความสอดคล้องระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละทักษะกับระดับความลึกของความรู้ตามวิธีของเว็บ ผู้วิจัยได้กำหนดจำนวนผู้เชี่ยวชาญไว้ 5 คน โดยใช้วิธีการเลือกแบบเจาะจง (purposive sampling) ตามเกณฑ์การคัดเลือกผู้เชี่ยวชาญที่กำหนดไว้ ผู้วิจัยติดต่อผู้เชี่ยวชาญโดยตรงผ่านทางโทรศัพท์ เพื่อเชิญชวนเข้าร่วมประเมินความสอดคล้องระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละทักษะกับระดับความลึกของความรู้ตามวิธีของเว็บ ผลการติดต่อ พบว่า ผู้เชี่ยวชาญทั้ง 5 คนตอบรับเข้าร่วม และมีคุณสมบัติตรงตามเกณฑ์การคัดเลือกที่กำหนดไว้

กลุ่มที่ 2 ผู้เชี่ยวชาญประเมินความสอดคล้องระหว่างแบบทดสอบ STOU-EPT กับกรอบ CEFR ผู้วิจัยกำหนดจำนวนผู้เชี่ยวชาญ 10 คน โดยใช้วิธีการเลือกแบบเจาะจงตามเกณฑ์การคัดเลือกผู้เชี่ยวชาญที่กำหนดไว้ แม้ สำนวน จดกระโทก (2561) แนะนำให้ใช้ผู้เชี่ยวชาญ 3-5 คนในการประเมินความสอดคล้อง แต่การวิจัยนี้เพิ่มจำนวนเป็น 10 คน เพื่อเสริมความน่าเชื่อถือและเพิ่มความหลากหลายในการประเมิน ซึ่งสอดคล้องกับแนวคิดของ Cizek and Bunch (2007) ผู้วิจัยติดต่อผู้เชี่ยวชาญโดยตรงผ่านทางโทรศัพท์เพื่อเชิญชวนเข้าร่วมประเมินความสอดคล้องระหว่างแบบทดสอบ STOU-EPT กับกรอบ CEF ผลการติดต่อ พบว่า ผู้เชี่ยวชาญทั้ง 10 คนตอบรับเข้าร่วม และมีคุณสมบัติตรงตามเกณฑ์การคัดเลือกที่กำหนดไว้

ผู้วิจัยได้กำหนดเกณฑ์การคัดเลือกผู้เชี่ยวชาญกลุ่มที่ 1 และกลุ่มที่ 2 ตลอดจนขั้นตอนและกระบวนการขอความยินยอมในการเข้าร่วมประเมินความสอดคล้อง ดังนี้

1.2.1.1 เกณฑ์การคัดเลือกผู้เชี่ยวชาญ ประกอบด้วย เกณฑ์การคัดเลือกเข้า เกณฑ์การคัดออก และเกณฑ์การหยุดโครงการวิจัย

เกณฑ์การคัดเลือกเข้า

1) เป็นผู้มีความรู้ระดับปริญญาตรี ปริญญาโท และ/หรือปริญญาเอก ในทิศทางเดียวกัน ได้แก่ การสอนภาษาอังกฤษ ภาษาศาสตร์ หรือการสื่อสารระหว่างวัฒนธรรม

2) เป็นผู้มีความรู้เกี่ยวกับการทดสอบภาษาอังกฤษตามกรอบ CEFR เช่น การออกแบบและพัฒนาแบบทดสอบ การฝึกอบรมผู้สอบและกรรมการประเมิน การตรวจข้อสอบ และการประเมินผล การวิจัยและพัฒนาเกี่ยวกับการทดสอบตามกรอบ CEFR

3) เป็นผู้มีความรู้ในการสอนภาษาอังกฤษในระดับอุดมศึกษา ตั้งแต่ 5 ปีขึ้นไป และเคยมีส่วนร่วมในการประเมินผลหรือการพัฒนาหลักสูตรตามกรอบ CEFR

เกณฑ์การคัดออก

ผู้เชี่ยวชาญประเมินความสอดคล้อง ติดภารกิจหรือไม่สามารถประเมินความสอดคล้องได้ตามวัน เวลา และสถานที่ที่กำหนด

เกณฑ์การหยุดโครงการวิจัย

ผู้เชี่ยวชาญประเมินความสอดคล้องสามารถออกจากการวิจัยเมื่อใดก็ได้ ไม่จำเป็นต้องบอกเหตุผล

1.2.1.2 ขั้นตอนและกระบวนการขอความยินยอมในการเข้าร่วมประเมินความสอดคล้อง มีดังนี้

1) ผู้วิจัยประสานงานไปยังผู้เชี่ยวชาญประเมินความสอดคล้องผ่านทางโทรศัพท์เพื่อเชิญชวนผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญรับฟังรายละเอียดของโครงการวิจัย

2) ผู้วิจัยเป็นผู้อธิบายกิจกรรมที่ให้ผู้ที่จะเป็นผู้เชี่ยวชาญทำอย่างละเอียดตามที่กำหนดไว้ในเอกสารแนะนำสำหรับผู้เชี่ยวชาญที่ประเมินความสอดคล้อง โดยไม่มีการใช้สื่อเสริมอื่น ๆ นอกเหนือจากการพูดอธิบาย ในระหว่างอธิบายกิจกรรมผู้วิจัยมีการสอบถามความเข้าใจผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญว่า มีคำถาม ข้อสงสัย หรือข้อวิตกกังวลหรือไม่ หากมีคำถามผู้วิจัยตอบคำถามต่าง ๆ ด้วยความเต็มใจ ไม่ปิดบัง ซ่อนเร้น เมื่อมีการอธิบายกิจกรรมให้กับผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญครบถ้วนจนหมดข้อสงสัย ผู้วิจัยก็สรุปใจความสำคัญเกี่ยวกับกิจกรรมที่ให้ผู้ที่จะเป็นผู้เชี่ยวชาญต้องทำ

3) ผู้วิจัยให้โอกาสและเวลาผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญ ประมาณ 1 สัปดาห์พิจารณาตัดสินใจว่าจะประเมินความสอดคล้องหรือไม่ พร้อมทั้งแจ้งว่าจะส่งเอกสารข้อ 3.1) ถึง 3.3) ไปยังผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญโดยผู้วิจัยไปส่งเอกสารให้ผู้เชี่ยวชาญทางไปรษณีย์ เพื่อให้โอกาสผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญได้คิดและ/หรือปรึกษากับผู้ที่เกี่ยวข้องก่อนตัดสินใจประเมินความสอดคล้อง โดยเอกสารที่จัดส่งมีดังนี้

3.1) เอกสารแนะนำสำหรับผู้เชี่ยวชาญที่ประเมินความสอดคล้อง เป็นเอกสารที่มีรายละเอียดเกี่ยวกับโครงการวิจัยที่ผู้วิจัยได้ชี้แจงแก่ผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญ เพื่อให้ได้อ่านทำความเข้าใจ และพิจารณาตัดสินใจประเมินความสอดคล้อง ทั้งนี้ หากผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญมีข้อซักถามเพิ่มเติม ให้สอบถามผู้วิจัยได้จนหมดข้อสงสัย โดยสามารถติดต่อสอบถามผู้วิจัยโดยตรง

3.2) หนังสือแสดงความยินยอมเข้าร่วมโครงการวิจัย หลังจากผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญอ่านทำความเข้าใจข้อมูลในเอกสารแนะนำสำหรับผู้เชี่ยวชาญอย่างรอบคอบจนหมดข้อสงสัยและตัดสินใจประเมินความสอดคล้อง ให้ผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญแสดงความยินยอมโดยการลงลายมือชื่อด้วยตนเองในหนังสือแสดงความยินยอมเข้าร่วมโครงการวิจัยที่ได้จัดส่งให้

3.3) ซองติดแสตมป์ หลังจากผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญลงลายมือชื่อด้วยตนเองในหนังสือแสดงความยินยอมเข้าร่วมโครงการวิจัยที่ได้จัดส่งให้แล้ว ให้นำหนังสือแสดงความยินยอมเข้าร่วมโครงการวิจัยใส่ซองติดแสตมป์ที่ผู้วิจัยจัดให้ส่งกลับคืนผู้วิจัย

4) ผู้วิจัยจัดส่งเอกสารประกอบการประเมินความสอดคล้อง พร้อมทั้งแจ้งกำหนดวัน เวลา และสถานที่ที่จะจัดกิจกรรมประเมินความสอดคล้องไปยังผู้เชี่ยวชาญทางไปรษณีย์

5) ผู้เชี่ยวชาญเข้าร่วมกิจกรรมประเมินความสอดคล้องตามวัน เวลา และสถานที่ที่ผู้วิจัยกำหนด โดยในการเข้าร่วมกิจกรรมประเมินความสอดคล้อง ผู้เชี่ยวชาญมีอิสระในการที่อภิปรายโต้ตอบหรือไม่อภิปรายโต้ตอบระหว่างสมาชิกในกลุ่มก็ได้

1.2.2 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัย มี 3 ชุด ได้แก่ ชุดที่ 1 แบบทดสอบ STOU-EPT ชุดที่ 2 แบบประเมินความสอดคล้องระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละทักษะกับระดับความลึกของความรู้ตามวิธีของเว็บ และชุดที่ 3 แบบประเมินความสอดคล้องระหว่างข้อสอบ ตัวชี้วัดตามกรอบ CEFR และระดับความลึกของความรู้ด้วยวิธีของเว็บ และคำอธิบายตัวชี้วัดสำหรับประเมินความสอดคล้อง รายละเอียดแต่ละชุดมีดังนี้

ชุดที่ 1 แบบทดสอบ STOU-EPT เป็นแบบทดสอบที่พัฒนาโดยคณาจารย์แขนงวิชาภาษาอังกฤษ สาขาวิชาศิลปศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช ภายใต้โครงการปรับปรุงข้อสอบ STOU-EPT มีวัตถุประสงค์เพื่อเพิ่มข้อสอบวัดความรู้ภาษาอังกฤษ STOU-EPT ในคลังข้อสอบ โดยมีตัวชี้วัดความสำเร็จของโครงการ คือ จำนวนข้อสอบ STOU-EPT ในคลังข้อสอบ เพิ่มขึ้น จำนวน 800 ข้อ จัดเป็นแบบทดสอบรวมทั้งสิ้น 8 แบบ ประกอบด้วย แบบทดสอบที่ปรับปรุงจากข้อสอบเดิม 4 แบบ และแบบทดสอบที่พัฒนาขึ้นใหม่ จำนวน 4 แบบ

การดำเนินโครงการนี้แล้วเสร็จในปี พ.ศ. 2566 โดยผู้วิจัยเป็นหัวหน้าโครงการ และดำเนินงานร่วมกับทีมพัฒนาข้อสอบ เพื่อให้ได้แบบทดสอบที่มีคุณภาพ สอดคล้องกับมาตรฐานสากล และสามารถใช้เป็นเครื่องมือวัดสมรรถนะทางภาษาอังกฤษของผู้สอบได้อย่างมีประสิทธิภาพ

แบบทดสอบ STOU-EPT ที่ใช้ในการวิจัยครั้งนี้ เป็นแบบทดสอบที่เลือกมา 1 แบบ จากทั้งหมด 4 แบบที่พัฒนาขึ้นใหม่ในปี พ.ศ. 2566 โดยมีลักษณะเป็นแบบทดสอบคู่ขนาน (parallel test) ครอบคลุม 3 ทักษะ ได้แก่ การฟัง จำนวน 25 ข้อ โครงสร้างภาษา จำนวน 35 ข้อ และการอ่าน จำนวน 40 ข้อ รวมทั้งสิ้น 100 ข้อ ข้อสอบในแต่ละทักษะออกแบบให้วัดระดับความสามารถทางภาษาระดับ A2 B1 และ B2 ตามกรอบ CEFR โดยไม่ครอบคลุมระดับ A1 C1 และ C2 ทั้งนี้แบบทดสอบทุกแบบพัฒนาตามคุณลักษณะเฉพาะของข้อสอบ (test specifications) และผ่านกระบวนการตรวจสอบคุณภาพ (quality control) ที่เข้มงวดเหมือนกัน ดังนั้น แบบทดสอบที่ใช้ในการวิจัยนี้จึงถือเป็นตัวแทน (representative sample) ของแบบทดสอบ STOU-EPT ฉบับปรับปรุงปี พ.ศ. 2566 ทั้งหมด แสดงดังตารางที่ 8

ตารางที่ 8 โครงสร้างของแบบทดสอบ STOU-EPT (ฉบับปรับปรุง พ.ศ. 2566)

ทักษะ	ลักษณะข้อสอบ		จำนวน ข้อสอบ
การฟัง	Text 1	การสนทนาสั้น ๆ ให้ผู้สอบฟังบทสนทนา และตอบข้อสอบแบบเลือกตอบ มี 3 ตัวเลือก โดยเลือกคำตอบที่ถูกต้องที่สุด	8
	Text 2	ให้ผู้สอบฟังเสียงแล้วจับคู่ให้ตรงกับสิ่งใดฟัง	6
	Text 3	ให้ผู้สอบฟังเสียงแล้วตอบข้อสอบแบบเติมคำ	6
	Text 4	ให้ผู้สอบฟังเสียงแล้วตอบข้อสอบแบบถูก-ผิด	5
	รวม (ข้อ 1-25)		25
โครงสร้าง		ข้อสอบแบบเลือกตอบ มี 4 ตัวเลือก ให้ผู้สอบเลือกคำตอบที่ถูกต้องที่สุด	35
ภาษา		รวม (ข้อ 26-60)	35
การอ่าน	Passage 1	ข้อสอบแบบเลือกตอบ มี 4 ตัวเลือก ให้ผู้สอบเลือกคำตอบที่ถูกต้องที่สุด	6
	Passage 2	ข้อสอบแบบเลือกตอบ มี 4 ตัวเลือก ให้ผู้สอบเลือกคำตอบที่ถูกต้องที่สุด	7
	Passage 3	ข้อสอบแบบเลือกตอบ มี 4 ตัวเลือก ให้ผู้สอบเลือกคำตอบที่ถูกต้องที่สุด	7
	Passage 4	ข้อสอบแบบเลือกตอบ มี 4 ตัวเลือก ให้ผู้สอบเลือกคำตอบที่ถูกต้องที่สุด	6
	Passage 5	ข้อสอบแบบเลือกตอบ มี 4 ตัวเลือก ให้ผู้สอบเลือกคำตอบที่ถูกต้องที่สุด	7
	Passage 6	ข้อสอบแบบเลือกตอบ มี 4 ตัวเลือก ให้ผู้สอบเลือกคำตอบที่ถูกต้องที่สุด	7
	รวม (ข้อ 61-100)		40
รวมข้อสอบทั้ง 3 ทักษะ			100

ชุดที่ 2 แบบประเมินความสอดคล้องระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละทักษะกับระดับความลึกของความรู้ตามวิธีของเว็บ แสดงตัวอย่างดังตารางที่ 9

ตารางที่ 9 ตัวอย่างแบบประเมินความสอดคล้องระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละทักษะ
กับระดับความลึกของความรู้ตามวิธีของเว็บ

ทักษะ (skills)	ระดับความลึกของความรู้			
	ความจำ (recall)	การมี ทักษะและ มโนทัศน์ (skills and concepts)	การคิด เชิงกลยุทธ์ (strategic thinking)	การคิด หลาย ขั้นตอน (extended thinking)
ด้าน Listening				
ตัวชี้วัด (มาตรฐานระดับ A2)				
Can understand simple information, questions, and short conversations about family, people, homes, work, hobbies, and daily life if people speak slowly and clearly and offer help if needed. Can follow changes of topic in TV news reports and understand the main points of the information. Can understand short, clear, and simple messages in public places such as airports or railway stations (e.g., “The train to London leaves at 4:30”). Can understand the main information in announcements, such as weather reports, if spoken very clearly.				

ผู้วิจัยได้นำแบบประเมินความสอดคล้องระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละทักษะกับระดับความลึกของความรู้ตามวิธีของเว็บ (รายละเอียดดังภาคผนวก จ) ให้ผู้เชี่ยวชาญ จำนวน 5 คน (รายชื่อดังภาคผนวก ก) ประเมินความสอดคล้องฯ

ผลการประเมินระดับความลึกของความรู้ระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละทักษะกับระดับความลึกของความรู้ตามวิธีของเว็บ โดยใช้ค่านิยาม (mode) พบว่า ทักษะการฟัง ระดับ A2 และ B1 วัดความลึกของความรู้ด้านการมีทักษะและมโนทัศน์ และระดับ B2 วัดความลึกของความรู้ด้านการคิดเชิงกลยุทธ์ ทักษะโครงสร้างภาษาทั้ง 3 ระดับ ได้แก่ A2 B1 และ B2 วัดความลึกของความรู้ด้านการมีทักษะและมโนทัศน์ และทักษะการอ่าน ระดับ A2 และ B1 วัดความลึกของความรู้ด้านการมีทักษะและมโนทัศน์ และระดับ B2 วัดความลึกของความรู้ด้านการคิดเชิงกลยุทธ์ แสดงดังตารางที่ 10

การตรวจสอบความเที่ยงระหว่างผู้ประเมิน (inter-rater reliability) ในการประเมินระดับความลึกของความรู้ระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละทักษะกับระดับความลึกของความรู้ตามวิธีของเว็บ ผู้วิจัยใช้การวิเคราะห์ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์แบบลำดับของสเปียร์แมนระหว่างผู้ประเมินแต่ละคู่ (mean of bivariate Spearman's rho rank correlations between raters: ρ) ซึ่งเป็นการนำค่า ρ ของผู้ประเมินแต่ละคู่มาหาค่าเฉลี่ยระดับความสอดคล้องโดยรวมของกลุ่มผู้ประเมินทั้งหมด วิธีการนี้เหมาะสมกับกรณีที่มีผู้ประเมินมากกว่า 2 คน ทั้งนี้ Evans (1996) ได้เสนอเกณฑ์การแปลผลค่าสัมประสิทธิ์สหสัมพันธ์ ดังนี้

ค่าสัมประสิทธิ์สหสัมพันธ์ .00 ถึง .19 หมายความว่า สอดคล้องกันในระดับต่ำมาก

ค่าสัมประสิทธิ์สหสัมพันธ์ .20 ถึง .39 หมายความว่า สอดคล้องกันในระดับต่ำ

ค่าสัมประสิทธิ์สหสัมพันธ์ .40 ถึง .59 หมายความว่า สอดคล้องกันในระดับปานกลาง

ค่าสัมประสิทธิ์สหสัมพันธ์ .60 ถึง .79 หมายความว่า สอดคล้องกันในระดับสูง

ค่าสัมประสิทธิ์สหสัมพันธ์ .80 ถึง 1.00 หมายความว่า สอดคล้องกันในระดับสูงมาก

ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมินในการประเมินระดับความลึกของความรู้ระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละทักษะกับระดับความลึกของความรู้ตามวิธีของเว็บ พบว่าทักษะการฟัง มีค่า ρ เท่ากับ .911 และทักษะการอ่าน มีค่า ρ เท่ากับ .911 แสดงว่า ทักษะการฟัง และทักษะการอ่าน ผู้ประเมินประเมินสอดคล้องกันในระดับสูงมาก ส่วนทักษะโครงสร้างภาษา มีค่า ρ เท่ากับ .540 แสดงว่า ผู้ประเมินประเมินสอดคล้องกันในระดับปานกลาง ผลดังกล่าวสอดคล้องกับเกณฑ์ที่ Evans (1996) กำหนดไว้ ซึ่งความเที่ยงในระดับปานกลางสำหรับทักษะโครงสร้างภาษา สะท้อนถึงความจำเป็นในการเสริมสร้างความเข้าใจร่วมกันของผู้ประเมินเกี่ยวกับตัวชี้วัด CEFR และระดับความลึกของความรู้ (DOK) ผ่านการอภิปรายอย่างเป็นระบบหรือการจัดประชุมชี้แจงก่อนดำเนินการประเมิน เพื่อให้การตัดสินมีความสอดคล้องกันและมีพื้นฐานเชิงแนวคิดที่ชัดเจน แสดงดังตารางที่ 10

ตารางที่ 10 ผลการวิเคราะห์ระดับความลึกของความรู้ตามวิธีของเว็บ (n = 5)

ทักษะ	ระดับ CEFR	ระดับความลึกของความรู้ (ความถี่)				ระดับ ความลึก (ฐาน นิยม)	p	แปล ความหมาย (p)
		(1) ความจำ	(2) การมี ทักษะและ มโนทัศน์	(3) การคิด เชิงกล ยุทธ	(4) การคิด หลาย ขั้นตอน			
การฟัง	A2	2	3			(2)	.911	สอดคล้อง
	B1		5			(2)		กันในระดับ
	B2		1	4		(3)		สูงมาก
โครงสร้าง ภาษา	A2	2	3			(2)	.540	สอดคล้อง
	B1		3	2		(2)		กันในระดับ
	B2		3		2	(2)		ปานกลาง
การอ่าน	A2	2	3			(2)	.911	สอดคล้อง
	B1		5			(2)		กันในระดับ
	B2		1	4		(3)		สูงมาก

ชุดที่ 3 แบบประเมินความสอดคล้องระหว่างข้อสอบ ตัวชี้วัดตามกรอบ CEFR และระดับความลึกของความรู้ด้วยวิธีของเว็บ และคำอธิบายตัวชี้วัดสำหรับประเมินความสอดคล้อง (รายละเอียดดังภาคผนวก ฉ)

แบบประเมินความสอดคล้องระหว่างข้อสอบ ตัวชี้วัดตามกรอบ CEFR และระดับความลึกของความรู้ด้วยวิธีของเว็บ แบ่งออกเป็น 2 เกณฑ์ ได้แก่ ความสอดคล้องด้านเนื้อหา และความสอดคล้องด้านความลึกของความรู้

ความสอดคล้องด้านเนื้อหา เป็นการประเมินว่าเนื้อหาข้อสอบ STOU-EPT แต่ละข้อสอดคล้องกับกรอบมาตรฐาน CEFR ระดับใด โดยใช้การพิจารณาของผู้เชี่ยวชาญ ซึ่งประเมินว่าข้อสอบ STOU-EPT แต่ละข้อตรงกับตัวชี้วัดของกรอบมาตรฐาน CEFR ระดับ A2 B1 หรือ B2

ความสอดคล้องด้านความลึกของความรู้ เป็นการประเมินเนื้อหาข้อสอบ STOU-EPT แต่ละข้อว่าวัดความซับซ้อนของการคิดของผู้สอบที่จะทำข้อสอบให้ถูกหรือผ่านระดับใด โดยใช้การพิจารณาของผู้เชี่ยวชาญ ซึ่งระดับความซับซ้อนของการคิด แบ่งออกเป็น 4 ระดับ ได้แก่

ระดับ 1 ความจำ เป็นระดับความรู้ความสามารถทางสมองที่ต้องการให้ผู้เรียนจำข้อมูล และบอกสิ่งที่เคยเห็นมา เช่น ข้อเท็จจริง นิยาม วิธีการ คุณสมบัติ ซึ่งความรู้ระดับนี้จะไม่ซับซ้อน เช่น การใช้สูตร การทำตามคำสั่ง คำที่แสดงถึงความจำ เช่น ระบุ จำได้ ใช้ คำนวณ

ระดับ 2 การมีทักษะและมโนทัศน์ เป็นเรื่องของการตัดสินใจเกี่ยวกับการแก้ปัญหา และการแก้ปัญหามีการดำเนินการมากกว่า 1 ขั้นตอน ซึ่งทักษะนี้เป็นการใช้ทักษะกระบวนการทางสมองสูงกว่าระดับ 1 คำที่แสดงถึงการมีทักษะและมโนทัศน์ เช่น จำแนก สร้าง ประเมิน รวบรวม นำเสนอ เปรียบเทียบ

ระดับ 3 การคิดเชิงกลยุทธ์ เป็นการคิดที่ต้องมีเหตุผล มีการวางแผน และมีการใช้หลักฐานสนับสนุนอย่างสมเหตุสมผล และเป็นการคิดที่สูงกว่าระดับ 1 และระดับ 2 ซึ่งคำที่แสดงถึงการคิดเชิงกลยุทธ์ เช่น การสรุป การวิพากษ์ด้วยหลักฐานอย่างมีเหตุผล การแก้ปัญหที่ไม่ใช่ปัญหาในชีวิตประจำวัน

ระดับ 4 การคิดหลายขั้นตอน เป็นการคิดขั้นสูงและซับซ้อน ต้องการให้ผู้สอบได้เชื่อมโยงประเด็นหลายส่วนเข้าด้วยกัน เช่น สามารถสรุปอ้างอิงผลการศึกษาและนำไปใช้แก้ปัญหาในสถานการณ์ใหม่ได้ ซึ่งคำที่แสดงถึงการคิดหลายขั้นตอน เช่น งานที่ต้องใช้การวิเคราะห์ ที่ต้องอาศัยการทำความเข้าใจต่อเนื่อง หลายขั้นตอน

สำหรับความสอดคล้องด้านความสมดุล มีวัตถุประสงค์เพื่อพิจารณาว่าข้อสอบ STOU-EPT ในแต่ละทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน มีการกระจายตัวของข้อสอบอย่างเท่าเทียมตามมาตรฐาน CEFR ระดับ A2 B1 และ B2 หรือทำให้ทราบว่ามาตรฐานใดที่ออกข้อสอบมากกว่ามาตรฐานอื่น หากพบว่าข้อสอบมีการกระจุกตัวอยู่ในระดับใดระดับหนึ่งมากกว่าระดับอื่น สะท้อนถึงความไม่สมดุลของข้อสอบ ซึ่งอาจส่งผลให้การวัดไม่ครอบคลุมเนื้อหาตามกรอบมาตรฐานที่กำหนด

ในการวิจัยครั้งนี้ การประเมินความสอดคล้องด้านความสมดุลไม่ได้ดำเนินการผ่านการพิจารณาโดยผู้เชี่ยวชาญ แต่ใช้วิธีการวิเคราะห์เชิงปริมาณ โดยคำนวณค่าดัชนีความสมดุล (balance index: Q) เพื่อวิเคราะห์การกระจายตัวของข้อสอบในแต่ละทักษะว่าครอบคลุมระดับมาตรฐาน CEFR (A2 B1 และ B2) อย่างสมดุลหรือไม่

ตัวอย่างของแบบประเมินความสอดคล้องระหว่างข้อสอบ ตัวชี้วัดตามกรอบ CEFR และระดับความลึกของความรู้ด้วยวิธีของเวบบ์ แสดงดังตารางที่ 11

ตารางที่ 11 ตัวอย่างแบบประเมินความสอดคล้องระหว่างข้อสอบ ตัวชี้วัดตามกรอบ CEFR
และระดับความลึกของความรู้ด้วยวิธีของ Webb

Item	CEFR Framework Indicators			Depth of knowledge (Webb,1997)			
	A2	B1	B2	Recall	Skills and concepts	Strategic thinking	Extended thinking
Listening = 25 Items (Items 1 – 25)							
1							
2							
3							
4							
5							
6							
.							
.							
.							
25							
Structure = 35 Items (Items 26 – 60)							
26							
27							
28							
29							
30							
31							
.							
.							
.							
60							
Reading = 40 Items (Item 61 - 100)							
61							
62							
63							
64							
65							
66							
.							
.							
.							
100							

การพัฒนาตัวชี้วัดสำหรับประเมินความสอดคล้อง

ผู้วิจัยได้พัฒนาตัวชี้วัดเพื่อใช้ควบคู่กับแบบประเมินความสอดคล้องระหว่างข้อสอบตัวชี้วัดตามกรอบ CEFR และระดับความลึกของความรู้ด้วยวิธีของเวบบ์ โดยมีจุดมุ่งหมายเพื่อให้การประเมินข้อสอบเป็นไปอย่างครอบคลุม ชัดเจน และสอดคล้องกับกรอบ CEFR

การดำเนินการพัฒนาตัวชี้วัด มี 4 ขั้นตอน ดังนี้

1. การนำคำอธิบาย (descriptors) ซึ่งมีลักษณะเป็นแบบตรวจสอบรายการ (checklists) ในส่วนของทักษะการฟัง และการอ่าน พัฒนาโดย European Association for Quality Language Services (EAQUALS) (2008) มาประยุกต์ใช้ในการพัฒนาตัวชี้วัดสำหรับการประเมินความสอดคล้องทักษะการฟัง และการอ่าน และนำ Core of English language taught at CEFR เฉพาะส่วนของ Grammar ที่พัฒนาโดย British Council and EAQUALS (2015) มาประยุกต์ใช้ในการพัฒนาตัวชี้วัดสำหรับการประเมินความสอดคล้องทักษะโครงสร้างภาษาเพื่อให้ได้แบบประเมินความสอดคล้องที่เป็นไปตามกรอบ CEFR ใน 3 ระดับความสามารถ ได้แก่ ระดับ A2 B1 และ B2

2. การพัฒนาตัวชี้วัดสำหรับประเมินความสอดคล้อง แบ่งออกเป็น 3 ทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน โดยแต่ละทักษะ จำแนกออกเป็น 3 ระดับความสามารถ ได้แก่ ระดับ A2 B1 และ B2 หลังจากพัฒนาตัวชี้วัดเสร็จแล้ว ผู้วิจัยได้นำตัวชี้วัดเหล่านี้ไปให้ที่ปรึกษาโครงการวิจัยตรวจสอบความถูกต้องและความเหมาะสมว่าตัวชี้วัดสอดคล้องกับกรอบ CEFR หรือไม่ รวมทั้งตรวจสอบความชัดเจนของภาษาที่ใช้ในแต่ละตัวชี้วัด

3. หลังจากปรับแก้ตัวชี้วัดสำหรับประเมินสอดคล้องตามคำแนะนำของที่ปรึกษาโครงการวิจัยแล้ว ผู้วิจัยได้นำตัวชี้วัดดังกล่าวไปให้ผู้เชี่ยวชาญด้านภาษาอังกฤษที่มีประสบการณ์เกี่ยวกับการทดสอบภาษาอังกฤษตามกรอบ CEFR จำนวน 3 คน ตรวจสอบความตรงเชิงเนื้อหา ความครอบคลุมโครงสร้าง ความถูกต้อง และความเหมาะสมของตัวชี้วัดว่าสอดคล้องกับกรอบ CEFR หรือไม่ รวมทั้งตรวจสอบความชัดเจนของการใช้ภาษาของแต่ละตัวชี้วัด โดยคำนวณค่า Item Objective Congruence (IOC) ซึ่งเกณฑ์การพิจารณาความสอดคล้องของประเด็นการประเมินกับจุดมุ่งหมายที่ต้องการวัด ค่าดัชนี IOC ที่คำนวณได้ต้องมีค่ามากกว่าหรือเท่ากับ .50 (Rovinelli & Hambleton, 1977) หากตัวชี้วัดใดมีค่า IOC ต่ำกว่าเกณฑ์ ต้องนำไปปรับแก้เพิ่มเติม เพื่อให้มั่นใจว่าแบบประเมินมีความตรงเชิงเนื้อหา ครอบคลุมโครงสร้าง มีความถูกต้อง และสอดคล้องกับกรอบ CEFR

ผลการตรวจสอบความตรงเชิงเนื้อหา โดยผู้เชี่ยวชาญ จำนวน 3 คน (รายชื่อดังภาคผนวก ก) พบว่า ตัวชี้วัดที่พัฒนาขึ้นมีความตรงเชิงเนื้อหา ความถูกต้อง และความเหมาะสมสามารถนำไปใช้ประเมินความสอดคล้องได้ โดยทุกตัวชี้วัดมีค่า IOC อยู่ระหว่าง .67 ถึง 1 ซึ่งสูงกว่าเกณฑ์ที่กำหนดไว้ที่ .50 ดังนั้น ตัวชี้วัดทั้งหมดจึงผ่านเกณฑ์และสามารถนำไปใช้ประเมินได้ อย่างไรก็ตาม ผู้เชี่ยวชาญมีข้อเสนอแนะให้รวมตัวชี้วัดบางตัวที่มีเนื้อหาซ้ำซ้อนกัน ผู้วิจัยจึงได้ดำเนินการปรับแก้โดยรวมตัวชี้วัดดังกล่าวเข้าด้วยกัน พร้อมทั้งปรับค่าให้มีความถูกต้องและเหมาะสมยิ่งขึ้น (รายละเอียดดังภาคผนวก ง)

4. กำหนดทักษะด้านเชิงลึกของความรู้ด้วยวิธีของเวบบ์ มี 4 ระดับ ได้แก่ ความจำ การมีทักษะและเมโนทัศน์ การคิดเชิงกลยุทธ์ และการคิดหลายขั้นตอน

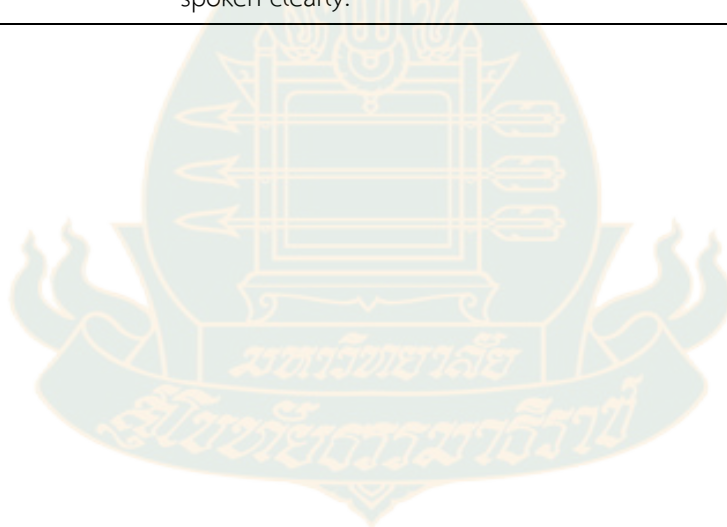
รายละเอียดแสดงดังตารางที่ 12

ตารางที่ 12 ตัวชี้วัดสำหรับประเมินความสอดคล้องตามกรอบ CEFR จำแนกตามทักษะ และระดับความสามารถ

ทักษะ	ระดับ ความสามารถ	ตัวชี้วัด
การฟัง	A2	สามารถเข้าใจข้อมูลง่าย ๆ คำถาม และบทสนทนาสั้น ๆ เกี่ยวกับครอบครัว ผู้คน บ้าน ที่ทำงาน งานอดิเรก และการใช้ชีวิตประจำวัน หากผู้พูดพูดอย่างช้าและชัดเจน และเสนอความช่วยเหลือหากจำเป็น สามารถเข้าใจการเปลี่ยนแปลงของหัวข้อในรายงานข่าวโทรทัศน์และเข้าใจประเด็นสำคัญของข้อมูลที่นำเสนอได้ สามารถเข้าใจข้อความสั้น ๆ ชัดเจน และง่าย ๆ ในสถานที่สาธารณะ เช่น สนามบินหรือสถานีรถไฟ เช่น รถไฟไปลอนดอนออกเวลา 4:30 สามารถเข้าใจข้อมูลสำคัญในประกาศ หากผู้พูดพูดอย่างชัดเจนมาก ๆ เช่น รายงานสภาพอากาศ Can understand simple information, questions, and short conversations about family, people, homes, work, hobbies, and daily life if people speak slowly and clearly and offer help if needed. Can follow changes of topic in TV news reports and understand the main points of the information. Can understand short, clear, and simple messages in public places such as airports or railway stations (e.g., “The train to London leaves at 4 : 3 0 ”). Can understand the main information in announcements, such as weather reports, if spoken very clearly.

ตารางที่ 12 ตัวชี้วัดสำหรับประเมินความสอดคล้องตามกรอบ CEFR จำแนกตามทักษะ
และระดับความสามารถ (ต่อ)

ทักษะ	ระดับ	ตัวชี้วัด
ความสามารถ		
การฟัง	B1	สามารถเข้าใจประเด็นสำคัญในเรื่องที่ฟัง เมื่อผู้พูดสื่อสารด้วยภาษามาตรฐานและชัดเจน ในประเด็นที่คุ้นเคย ใช้ในการสนทนาในชีวิตประจำวัน การอภิปราย หรือรายการโทรทัศน์ โดยอาจมีบางโอกาสที่ต้องมีการพูดซ้ำหรือขยายความเพิ่มเติม สามารถเข้าใจการพูดคุยนัย ๆ หรือการอภิปรายที่ชัดเจนและตรงไปตรงมาในหัวข้อที่คุ้นเคยในสถานการณ์ชีวิตประจำวันได้ สามารถเข้าใจข้อมูลทางเทคนิคง่าย ๆ เช่น คู่มือการใช้งานสำหรับอุปกรณ์ประเภทที่คุ้นเคย หากผู้พูดพูดอย่างชัดเจน Can understand the main points of clear standard speech on familiar, everyday topics in conversations, discussions, or TV programs, though I may sometimes need repetition or clarification for details. Can follow clearly spoken, straightforward short talks or discussions on familiar topics in everyday situations. Can understand simple technical information, such as operating instructions for familiar equipment if spoken clearly.



ตารางที่ 12 ตัวชี้วัดสำหรับประเมินความสอดคล้องตามกรอบ CEFR จำแนกตามทักษะ
และระดับความสามารถ (ต่อ)

ทักษะ	ระดับ	ตัวชี้วัด
ความสามารถ		
การฟัง	B2	สามารถเข้าใจแนวคิดสำคัญ และรายละเอียดของการพูดที่ซับซ้อนในหัวข้อที่เป็นรูปธรรมและนามธรรม รวมถึงการอภิปรายทางเทคนิคที่ใช้ศัพท์เฉพาะ เมื่อมีการสื่อสารด้วยภาษาที่เป็นมาตรฐาน สามารถเข้าใจสาระสำคัญในการบรรยาย การพูดคุย รายงาน และการนำเสนอทางวิชาการหรือวิชาชีพที่ซับซ้อนโดยใช้ศัพท์เฉพาะ สามารถเข้าใจละครโทรทัศน์ ภาพยนตร์ ข่าว สถานการณ์ปัจจุบัน สารคดี บทสัมภาษณ์ และรายการทอล์กโชว์ เมื่อสื่อสารด้วยภาษามาตรฐาน สามารถเข้าใจประกาศและข้อความที่ผู้พูดพูดด้วยภาษามาตรฐานที่ความเร็วปกติ สามารถจับใจความในเรื่องที่ผู้พูดพูดได้มาก โดยเป็นเรื่องใกล้ ๆ ตัว ซึ่งต้องใช้ความพยายามมาก แต่อาจเป็นการยากที่จะเข้าใจการอภิปรายระหว่างผู้พูดหลายคนที่ไม่มีการปรับระดับภาษา Can understand the main ideas and details of complex speech on concrete and abstract topics, including technical discussions in my field, when delivered in a standard dialect. Can follow the essentials of lectures, talks, reports, and other complex academic or professional presentations in my field. Can follow TV dramas, films, news, current affairs, documentaries, interviews, and talk shows in standard dialect. Can understand announcements and messages spoken in standard spoken language at normal speed. Can with some effort, catch much of what is said around me, but may struggle with discussions involving multiple speakers who do not adjust their language.

ตารางที่ 12 ตัวชี้วัดสำหรับประเมินความสอดคล้องตามกรอบ CEFR จำแนกตามทักษะ
และระดับความสามารถ (ต่อ)

ทักษะ	ระดับ	ตัวชี้วัด
ความสามารถ		
โครงสร้าง ภาษา	A2	สามารถเข้าใจไวยากรณ์หรือโครงสร้างประโยคง่ายๆ เกี่ยวกับการใช้: Can understand simple grammar or sentence structures regarding the use of: - Adjectives – comparative, – utilizing 'than' and the definite article - Adjectives – superlative – incorporating the definite article - Adverbial phrases of time, place, and frequency – including considerations of word order - Adverbs of frequency - Articles – with countable and uncountable nouns - Countables and uncountables: much/many in everyday life - Future Time (will and going to) - Gerunds - Going to - Imperatives - Modals (can/could) - Modals (have to) - Modals (should) in everyday life - Past continuous - Past simple - Phrasal verbs (common) - Possessives (use of 's, s') - Prepositional phrases (place, time and movement) - Prepositions of time (on/in/at) in everyday life - Present continuous - Present continuous for future - Present perfect - Questions - Verb + ing/infinitive: like/want-would like - Wh-questions in the past - Zero and 1st conditional in everyday life

ตารางที่ 12 ตัวชี้วัดสำหรับประเมินความสอดคล้องตามกรอบ CEFR จำแนกตามทักษะ
และระดับความสามารถ (ต่อ)

ทักษะ	ระดับ	ตัวชี้วัด
ความสามารถ		
โครงสร้าง ภาษา	B1	สามารถเข้าใจไวยากรณ์หรือโครงสร้างประโยคของข้อความทั่วไป เกี่ยวกับการใช้: Can understand the grammar or sentence structure of texts in general regarding the use of - Adverbs - A broader range of intensifiers such as "too" and "enough" - Comparatives and superlatives - Complex question tags - Conditionals (2nd and 3rd) - Connecting words expressing cause and effect, contrast, etc. - Dealing with familiar or engaging subjects, such as in the workplace, school, or leisure activities - Future continuous - Modals (must/can't deduction) - Modals (might, may, will, probably) - Modals (should have/might have/etc.) - Modals (must/have to). Dealing with familiar or engaging subjects, such as in the workplace, school, or leisure activities - Past continuous - Past perfect - Past simple - Past tense responses - Phrasal verbs, extended - Present perfect continuous - Present perfect/past simple. Dealing with familiar or engaging subjects, such as in the workplace, school, or leisure activities - Reported speech (range of tenses) - Simple passive - Wh-questions in the past - Will and going to, for prediction. Dealing with familiar or engaging subjects, such as in the workplace, school, or leisure activities

ตารางที่ 12 ตัวชี้วัดสำหรับประเมินความสอดคล้องตามกรอบ CEFR จำแนกตามทักษะ
และระดับความสามารถ (ต่อ)

ทักษะ	ระดับ	ตัวชี้วัด
ความสามารถ		
โครงสร้าง ภาษา	B2	สามารถเข้าใจไวยากรณ์หรือโครงสร้างประโยคของข้อความที่ซับซ้อนเกี่ยวกับการใช้: Can understand more complex grammar or sentence structures regarding the use of - Adjectives and adverbs - Future continuous - Future perfect - Future perfect continuous - Mixed conditionals correctly and fluently in various subjects and topics. - Modals (can't have, needn't have) - Modals of deduction and speculation - Narrative tenses - Passives - Past perfect - Past perfect continuous correctly and fluently in various subjects and topics - Phrasal verbs, extended - Relative clauses - Reported speech - Will and going to, for prediction - Wish - Would expressing habits, in the past correctly and fluently in various subjects and topics

ตารางที่ 12 ตัวชี้วัดสำหรับประเมินความสอดคล้องตามกรอบ CEFR จำแนกตามทักษะ
และระดับความสามารถ (ต่อ)

ทักษะ	ระดับ	ตัวชี้วัด
ความสามารถ		
การอ่าน	A2	สามารถเข้าใจข้อความสั้น ๆ ง่าย ๆ ที่มีคำศัพท์ที่คุ้นเคย รวมถึงคำศัพท์ที่เป็นภาษาต่างประเทศ สามารถหาข้อมูลสำคัญที่ปรากฏในโฆษณา แผ่นพับ หน้าเว็บ แคตตาล็อก ตารางเวลา เป็นต้น สามารถเข้าใจสาระสำคัญในรายการข่าวสั้น ๆ และง่าย ๆ ถ้ามีความรู้บางอย่างเกี่ยวกับเรื่องนั้น เช่น ข่าวเกี่ยวกับกีฬา หรือบุคคลที่มีชื่อเสียง สามารถเข้าใจวิธีการใช้หรือคำแนะนำที่ชัดเจน เช่น วิธีใช้โทรศัพท์ เครื่องกดเงินสด หรือการทำเครื่องดื่ม สามารถเข้าใจสาระสำคัญในเรื่องราวสั้น ๆ ง่าย ๆ ในชีวิตประจำวันโดยเฉพาะอย่างยิ่งหากมีการนำเสนอด้วยภาพ สามารถเข้าใจข้อความสั้น ๆ ง่าย ๆ จากเพื่อน เช่น อีเมล แชนบนเว็บ ไปสการ์ด หรือจดหมายสั้น ๆ Can understand short, simple texts containing familiar vocabulary including international words. Can find the most important information in advertisements, information leaflets, webpages, catalogues, timetables etc. 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. Can understand clear instructions. For example: how to use a telephone, a cash machine or a drinks machine. Can understand the main points in short, simple, everyday stories, especially if there is visual support. Can understand short simple messages from friends. For example: e-mails, web chats, postcards or short letters.

ตารางที่ 12 ตัวชี้วัดสำหรับประเมินความสอดคล้องตามกรอบ CEFR จำแนกตามทักษะ
และระดับความสามารถ (ต่อ)

ทักษะ	ระดับ	ตัวชี้วัด
ความสามารถ		
การอ่าน	B1	สามารถเข้าใจสาระสำคัญในข้อความที่เป็นข้อเท็จจริง ตรงไปตรงมา ในเรื่องที่เป็ ความสนใจส่วนบุคคลหรืออาชีพได้ดีพอที่จะพูดถึงในภายหลัง สามารถหาและ ทำความเข้าใจข้อมูลที่ต้องการในโบรชัวร์ แผ่นพับ และข้อความสั้น ๆ อื่น ๆ ที่เกี่ยวข้องกับความสนใจของบุคคล สามารถเข้าใจสาระสำคัญในบทความสั้น ๆ ที่อยู่ในหนังสือพิมพ์ และนิตยสาร โดยเป็นเรื่องเกี่ยวกับปัจจุบันและเรื่องที่คุ้นเคย สามารถทำตามคำสั่งที่ง่าย ๆ เช่น การเล่นเกม การทำอาหารโดยใช้อุปกรณ์ที่ คุ้นเคย สามารถเข้าใจเรื่องราวในรูปแบบที่ง่าย ๆ ของนวนิยาย และสามารถ ติดตามเนื้อเรื่องในเรื่องสั้นที่มีโครงสร้างชัดเจน โดยต้องใช้ความพยายาม และ ใช้พจนานุกรมเป็นประจำ สามารถเข้าใจจดหมายส่วนตัวเกี่ยวกับเหตุการณ์ ความรู้สึก และความปรารถนาได้ดี พอที่จะเขียนตอบกลับ Can understand the main points in straightforward factual texts on subjects of personal or professional interest well enough to talk about them afterwards. Can find and understand the information I need in brochures, leaflets and other short texts relating to my interests. Can understand the main points in short newspaper and magazine articles about current and familiar topics. Can follow simple instructions, for example for a game, using familiar types of equipment or cooking a meal. 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. Can understand private letters about events, feelings and wishes well enough to write back.

ตารางที่ 12 ตัวชี้วัดสำหรับประเมินความสอดคล้องตามกรอบ CEFR จำแนกตามทักษะ
และระดับความสามารถ (ต่อ)

ทักษะ	ระดับ	ตัวชี้วัด
ความสามารถ		
การอ่าน	B2	สามารถอ่านเรื่องได้ด้วยตนเอง โดยใช้พจนานุกรมและแหล่งข้อมูลอ้างอิงอื่น ๆ เมื่อจำเป็น สามารถเข้าใจเนื้อหาและสาระสำคัญของข่าว บทความ และรายงาน ในหัวข้อที่เกี่ยวข้องกับความสนใจหรืองานได้อย่างรวดเร็ว และตัดสินใจว่าเรื่องที่อ่านคุ้มค่าหรือไม่ สามารถเข้าใจบทความ รายงาน และบทวิจารณ์ ที่ผู้เขียนแสดงมุมมองเฉพาะ เช่น การวิจารณ์ทางการเมือง การวิพากษ์วิจารณ์ทฤษฎีการปกครอง ภาพยนตร์ เป็นต้น สามารถเข้าใจคำแนะนำที่ยาว ๆ เช่น อ่านคู่มือการใช้งานทีวีหรือกล้องดิจิทัลสำหรับติดตั้งซอฟต์แวร์ トラバได้ที่สามารถอ่านส่วนที่ยากเข้าใจ สามารถอ่านเรื่องสั้นและนวนิยายที่เขียนด้วยภาษาและลีลาที่ตรงไปตรงมา โดยใช้พจนานุกรม หากคุ้นเคยกับเรื่องราว และ/หรือผู้เขียน สามารถเข้าใจสาระสำคัญในจดหมายที่เป็นทางการ และไม่เป็นการที่เกี่ยวกับความสนใจส่วนตัวและอาชีพ โดยใช้พจนานุกรมเป็นบางครั้ง Can read with a large degree of independence, using dictionaries and other reference sources selectively when necessary. 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. 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). 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. 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. Can understand the main points in formal and informal letters relating to my personal and professional interests, with occasional use of a dictionary.

หลังจากตรวจสอบความตรงเชิงเนื้อหาของตัวชี้วัดสำหรับประเมินความสอดคล้องในแต่ละทักษะจากผู้เชี่ยวชาญ ผู้วิจัยได้นำตัวชี้วัดที่ผ่านการปรับปรุงแล้ว มาให้ผู้เชี่ยวชาญ จำนวน 3 คน ทดลองประเมินความสอดคล้องฯ มีคุณสมบัติดังนี้

1. เป็นผู้มีความรู้ในการศึกษาระดับปริญญาตรี ปริญญาโท และ/หรือปริญญาเอกในทิศทางเดียวกัน ได้แก่ การสอนภาษาอังกฤษ ภาษาศาสตร์ หรือการสื่อสารระหว่างวัฒนธรรม
2. เป็นผู้มีความรู้ประสบการณ์เกี่ยวกับการทดสอบภาษาอังกฤษตามกรอบ CEFR เช่น การออกแบบและพัฒนาแบบทดสอบ การฝึกอบรมผู้สอบและกรรมการประเมิน การตรวจข้อสอบและการประเมินผล การวิจัยและพัฒนาเกี่ยวกับการทดสอบตามกรอบ CEFR
3. เป็นผู้มีความรู้ประสบการณ์การสอนภาษาอังกฤษในระดับอุดมศึกษา ตั้งแต่ 5 ปีขึ้นไป และเคยมีส่วนร่วมในการประเมินผลหรือการพัฒนาหลักสูตรตามกรอบ CEFR

การทดลองประเมินความสอดคล้องของเนื้อหาของแบบทดสอบ STOU-EPT โดยผู้เชี่ยวชาญแต่ละคนประเมินความสอดคล้องของเนื้อหา และความสอดคล้องด้านความลึกของความรู้ของแบบทดสอบ STOU-EPT เพื่อตรวจสอบความเที่ยงระหว่างผู้ประเมิน (inter-rater reliability) โดยใช้การวิเคราะห์ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์แบบลำดับของสเปียร์แมนระหว่างผู้ประเมินแต่ละคู่ (ρ)

ผลการทดลองประเมินความสอดคล้องของเนื้อหาของแบบทดสอบ STOU-EPT แต่ละทักษะ พบว่า ทักษะการฟัง มีค่า ρ เท่ากับ .644 และทักษะโครงสร้างภาษา มีค่า ρ เท่ากับ .655 แสดงว่า ทักษะการฟัง และทักษะโครงสร้างภาษา ผู้ประเมินประเมินสอดคล้องกันในระดับสูง ส่วนทักษะการอ่าน มีค่า ρ เท่ากับ .479 แสดงว่า ผู้ประเมินประเมินสอดคล้องกันในระดับปานกลาง ผลดังกล่าวสอดคล้องกับเกณฑ์ที่ Evans (1996) กำหนด แสดงดังตารางที่ 13

ผลการทดลองประเมินความสอดคล้องด้านความลึกของความรู้ของแบบทดสอบ STOU-EPT แต่ละทักษะ พบว่า ทักษะการฟัง มีค่า ρ เท่ากับ .882 แสดงว่า ผู้ประเมินประเมินสอดคล้องกันในระดับสูงมาก ทักษะโครงสร้างภาษา มีค่า ρ เท่ากับ .554 แสดงว่า ผู้ประเมินประเมินสอดคล้องกันในระดับปานกลาง และทักษะการอ่าน มีค่า ρ เท่ากับ .200 แสดงว่า ผู้ประเมินประเมินสอดคล้องกันในระดับต่ำ ผลดังกล่าวสอดคล้องกับเกณฑ์ที่ Evans (1996) กำหนดไว้ แสดงดังตารางที่ 13

เมื่อพิจารณาผลการประเมินทั้งสองด้านร่วมกัน พบว่า ทักษะโครงสร้างภาษาและทักษะการอ่านมีความเที่ยงเพียงระดับปานกลางถึงระดับต่ำ ซึ่งอาจเกิดจากความแตกต่างในการตีความตัวชี้วัดของกรอบ CEFR และระดับ DOK เพื่อปรับปรุงการให้คะแนนสอดคล้องกันมากขึ้นในอนาคต ควรมีการจัดประชุมหรือฝึกอบรมให้ผู้ประเมินได้ทำความเข้าใจเกณฑ์ร่วมกัน พร้อมฝึก

ประเมินจากข้อสอบตัวอย่างที่มีมาตรฐาน (benchmark items) โดยเฉพาะในทักษะที่ซับซ้อน เพื่อให้การประเมินมีความสอดคล้อง และมีพื้นฐานทางแนวคิดที่ชัดเจนมากยิ่งขึ้น

ตารางที่ 13 ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมินที่ทดลองประเมินความสอดคล้องของเนื้อหา และความลึกของความรู้

ทักษะ	ความสอดคล้อง	จำนวนข้อสอบ	จำนวนผู้ประเมิน	p	แปลความหมาย
การฟัง	เนื้อหา	25	3	.644	สอดคล้องกันในระดับสูง
	ความลึกของความรู้	25	3	.882	สอดคล้องกันในระดับสูงมาก
โครงสร้างภาษา	เนื้อหา	35	3	.655	สอดคล้องกันในระดับสูง
	ความลึกของความรู้	35	3	.554	สอดคล้องกันในระดับปานกลาง
การอ่าน	เนื้อหา	40	3	.479	สอดคล้องกันในระดับปานกลาง
	ความลึกของความรู้	40	3	.200	สอดคล้องกันในระดับต่ำ

1.2.3 การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลในการวิจัย ผู้วิจัยดำเนินการโดยให้ผู้เชี่ยวชาญ จำนวน 10 คน ประเมินความสอดคล้อง เพื่อตรวจสอบว่าข้อสอบ STOU-EPT เป็นไปตามกรอบ CEFR และวัดระดับความลึกของความรู้ด้วยวิธีของเว็บด้านใดบ้าง

การเก็บรวบรวมข้อมูล มี 4 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การปฐมนิเทศผู้เชี่ยวชาญ ผู้วิจัยได้ปฐมนิเทศผู้เชี่ยวชาญทั้ง 10 คน (รายชื่อดังภาคผนวก ข) เพื่ออธิบายรายละเอียดเกี่ยวกับตัวชี้วัดใน 3 ทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน รวมทั้งอธิบายแนวคิดเกี่ยวกับระดับความลึกของความรู้ตามวิธีของเว็บ ประกอบด้วย ความจำ การมีทักษะและมโนทัศน์ การคิดเชิงกลยุทธ์ และการคิดหลายขั้นตอน เพื่อให้ผู้เชี่ยวชาญเข้าใจตัวชี้วัด และระดับความลึกของความรู้อย่างชัดเจน พร้อมทั้งแนะนำวิธีการประเมินความสอดคล้อง

ขั้นตอนที่ 2 การฝึกซ้อมประเมินความสอดคล้อง หลังจากปฐมนิเทศ ผู้วิจัยให้ผู้เชี่ยวชาญได้ฝึกซ้อมประเมินความสอดคล้อง โดยใช้แบบประเมินที่กำหนดไว้ เพื่อให้คุ้นเคยกับกระบวนการประเมิน

ขั้นตอนที่ 3 การซักถามและแลกเปลี่ยนความคิดเห็น หลังจากฝึกซ้อมประเมินความสอดคล้องเสร็จสิ้น ผู้วิจัยเปิดโอกาสให้ผู้เชี่ยวชาญได้ซักถาม แลกเปลี่ยนความคิดเห็น และปรับความเข้าใจร่วมกัน จนมั่นใจว่าผู้เชี่ยวชาญทุกคนมีความเข้าใจตัวชี้วัด และระดับความลึกของความรู้ตรงกัน

ทั้งนี้ การดำเนินงานในขั้นตอนที่ 1 ถึง 3 ดำเนินการแล้วเสร็จในระยะเวลา 1 วัน

ขั้นตอนที่ 4 การประเมินความสอดคล้อง ผู้เชี่ยวชาญประเมินความสอดคล้องของข้อสอบ STOU-EPT ครอบคลุม 3 ทักษะ ได้แก่ การฟัง จำนวน 25 ข้อ โครงสร้างภาษา จำนวน 35 ข้อ และการอ่าน จำนวน 40 ข้อ โดยพิจารณาข้อสอบทีละข้อ พร้อมประเมินความสอดคล้องกับตัวชี้วัดตามกรอบ CEFR และระดับความลึกของความรู้ใน 4 ระดับ ได้แก่ ความจำ การมีทักษะและมนทัศน์ การคิดเชิงกลยุทธ์ และการคิดหลายขั้นตอน ทั้งนี้ ผู้วิจัยกำหนดเวลาในการประเมินความสอดคล้องไว้ 14 วัน เพื่อให้ผู้เชี่ยวชาญมีเวลามากพอในการพิจารณาข้อสอบแต่ละข้ออย่างละเอียด และลดความเหนื่อยล้าจากการประเมิน

ผู้วิจัยได้กำหนดตารางดำเนินงานให้ผู้เชี่ยวชาญ เพื่อให้การประเมินเป็นไปอย่างถูกต้องและเป็นระบบ รวมระยะเวลา 14 วัน แบ่งออกเป็น 4 ช่วง ดังนี้

วันที่ 1-2 ผู้เชี่ยวชาญทำความเข้าใจกับเอกสารประกอบการประเมิน ประกอบด้วย รายละเอียดข้อสอบ STOU-EPT ตัวชี้วัดสำหรับประเมินความสอดคล้องตามกรอบ CEFR และ คำจำกัดความของระดับความลึกของความรู้ (DOK) ตามวิธีของเว็บ ขั้นตอนนี้มีวัตถุประสงค์เพื่อสร้างความเข้าใจที่ถูกต้องตรงกันในแนวคิดและเกณฑ์การประเมินก่อนเริ่มดำเนินการ

วันที่ 3-10 ผู้เชี่ยวชาญประเมินข้อสอบทีละข้อโดยอิสระ ครอบคลุมทั้งสามทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน โดยระบุความสอดคล้องของข้อสอบกับตัวชี้วัดตามกรอบ CEFR และระดับความลึกของความรู้ตามเกณฑ์ของเว็บ พร้อมทำเครื่องหมาย (✓) ในแบบประเมินความสอดคล้องฯ

วันที่ 11-12 ผู้เชี่ยวชาญทบทวนผลการประเมินของตนเอง โดยเน้นข้อสอบที่ยังไม่มั่นใจในรอบแรก เพื่อพิจารณาอย่างรอบคอบอีกครั้ง และสามารถขอคำชี้แจงจากผู้วิจัยหากจำเป็น ก่อนตัดสินใจหรือปรับแก้ผลการประเมิน เพื่อให้ได้ผลการประเมินที่ถูกต้องที่สุด ทั้งนี้ยังคงรักษาความเป็นอิสระในการตัดสินใจของแต่ละบุคคล

วันที่ 13-14 ผู้เชี่ยวชาญดำเนินการประเมินรอบสุดท้าย (final ratings) หลังจากได้ ทบทวนและรับคำชี้แจงเพิ่มเติมแล้ว พร้อมส่งผลการประเมินทั้งหมดให้แก่ผู้วิจัย

ตลอดระยะเวลา 14 วัน ผู้เชี่ยวชาญสามารถส่งคำถามหรือขอคำชี้แจงเพิ่มเติมได้ทุกเมื่อ ผ่านทางอีเมลหรือการประชุมออนไลน์ที่นัดหมายไว้ล่วงหน้า ผู้วิจัยรวบรวมคำชี้แจงทั้งหมดเป็นลายลักษณ์อักษรและส่งต่อให้ผู้เชี่ยวชาญทุกคน เพื่อให้เกิดความเข้าใจและการตีความเกณฑ์การประเมินที่สอดคล้องกัน ขณะเดียวกันก็ยังคงความเป็นอิสระในการตัดสินใจของแต่ละบุคคล

1.2.4 การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูล ผู้วิจัยวิเคราะห์ค่าดัชนีความสอดคล้องตามวัตถุประสงค์ข้อที่ 1 เพื่อศึกษาความสอดคล้องระหว่างแบบทดสอบ STOU-EPT กับกรอบ CEFR ด้วยวิธีของเว็บ โดยใช้โปรแกรม R การวิเคราะห์แบ่งออกเป็น 2 ส่วน ดังนี้

ส่วนที่ 1 การวิเคราะห์ความสอดคล้องของเนื้อหา และด้านความลึกของความรู้

1) การวิเคราะห์ข้อมูลในส่วนนี้ ผู้วิจัยนำผลประเมินของผู้เชี่ยวชาญทั้ง 10 คน มาคำนวณค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน สำหรับการวิเคราะห์ของเนื้อหา ส่วนการวิเคราะห์ด้านความลึกของความรู้คำนวณร้อยละ และค่าเฉลี่ย

2) การตรวจสอบความเที่ยงระหว่างผู้ประเมิน (inter-rater reliability) โดยใช้การวิเคราะห์ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์แบบลำดับของสเปียร์แมนระหว่างผู้ประเมินแต่ละคู่ (p)

ส่วนที่ 2 การวิเคราะห์ความสอดคล้องด้านความสมดุล

ความสมดุลของการกระจาย (balance of representation หรือ Q) เป็นเกณฑ์สำคัญในการประเมินความสอดคล้องของแบบทดสอบกับระดับความสามารถแต่ละทักษะ เพื่อให้มั่นใจว่าข้อสอบ STOU-EPT มีการกระจายในแต่ละระดับความสามารถของแต่ละทักษะ (การฟัง โครงสร้างภาษา และการอ่าน) อย่างเท่าเทียม โดยค่าดัชนีความสมดุล หรือ Q ที่มีค่าเข้าใกล้ 1 แสดงถึงความสมดุลของแต่ละระดับความสามารถของแต่ละทักษะที่ให้น้ำหนักความสำคัญอย่างเท่าเทียมกัน ในขณะที่ค่า Q เข้าใกล้ 0 แสดงถึงความไม่สมดุล โดยอาจมีการเน้นน้ำหนักในแต่ละระดับความสามารถ บางตัวมากเกินไปหรือน้อยเกินไป (Webb, 2007)

การวิเคราะห์ความสอดคล้องด้านความสมดุล ใช้สูตรในการคำนวณ ดังนี้

$$\text{ดัชนีความสมดุล} = 1 - \left(\sum_{k=1}^O \left| \frac{I(k)}{O} - \frac{I(k)}{H} \right| \right) \div 2$$

โดยที่

O = จำนวนตัวชี้วัดที่ตรงกับมาตรฐาน

I (k) = จำนวนข้อสอบที่ตรงกับตัวชี้วัดตัวที่ (k)

H = จำนวนข้อสอบทั้งหมดที่ตรงกับมาตรฐาน

ดัชนีความสมดุลที่คำนวณได้ ควรมีค่าตั้งแต่ .70 ขึ้นไป เพื่อแสดงถึงความสมดุลในระดับที่ยอมรับได้ (Webb, 2007)

ตอนที่ 1.3 ตรวจสอบคุณภาพข้อสอบ STOU-EPT โดยใช้ทฤษฎีการตอบสนองข้อสอบ

การตรวจสอบคุณภาพแบบทดสอบ STOU-EPT โดยใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อวิเคราะห์คุณภาพข้อสอบรายข้อ โดยผลการวิเคราะห์คุณภาพข้อสอบจากขั้นตอนนี้ ได้ถูกนำไปใช้เป็นข้อมูลประกอบการพิจารณาของผู้เชี่ยวชาญในการกำหนดมาตรฐานเพื่อหาคะแนนจุดตัด ในรอบที่ 2

การดำเนินงานในตอนนี้ แบ่งออกเป็น 4 ส่วน ได้แก่ ตัวอย่างที่ใช้ในการวิจัย เครื่องมือที่ใช้ในการวิจัย การเก็บรวบรวมข้อมูล และการวิเคราะห์ข้อมูล รายละเอียดแต่ละส่วนมีดังนี้

1.3.1 ตัวอย่างที่ใช้ในการวิจัย

ผู้วิจัยได้นำแบบทดสอบ STOU-EPT ที่ผ่านการประเมินความสอดคล้องจากตอนที่ 1.2 ไปทดลองสอบกับผู้เข้าสอบอาสาสมัครที่กำหนดไว้เบื้องต้น จำนวน 600 คน ซึ่งประกอบด้วย นักศึกษาของมหาวิทยาลัยสุโขทัยธรรมมาธิราช และผู้สมัครจากภายนอกที่ลงทะเบียนสอบด้วยความสมัครใจ ระหว่างเดือนพฤษภาคมถึงกันยายน พ.ศ. 2567 อย่างไรก็ตาม ระหว่างการดำเนินการพบว่า มีผู้เข้าสอบมากกว่าเป้าหมายที่กำหนดไว้ และระบบสอบอิเล็กทรอนิกส์ของมหาวิทยาลัยมีความพร้อมรองรับผู้เข้าสอบจำนวนมาก จึงสามารถเก็บข้อมูลจากผู้เข้าสอบได้รวมทั้งสิ้น 1,272 คน ซึ่งมากกว่าจำนวนที่กำหนดไว้เดิม

การได้กลุ่มตัวอย่างขนาดใหญ่เช่นนี้เป็นประโยชน์ต่อการวิเคราะห์คุณภาพข้อสอบด้วย ทฤษฎีการตอบสนองข้อสอบ เนื่องจากการประมาณค่าพารามิเตอร์ ได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าความเดา (c) จำเป็นต้องใช้ข้อมูลจากกลุ่มตัวอย่างขนาดใหญ่เพื่อเพิ่มความแม่นยำ และความน่าเชื่อถือในการตีความผลการวิเคราะห์ (de Ayala, 2009; Hambleton, Swaminathan, & Rogers, 1991) ทั้งนี้ จำนวนผู้เข้าสอบที่ได้จริงสอดคล้องกับข้อแนะนำทางจิตมิติ (psychometric guidelines) สำหรับการปรับเทียบโมเดล IRT มักกำหนดว่าควรมีผู้เข้าร่วมประมาณ 500–1,000 คน (DeMars, 2010)

1.3.2 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัย เป็นแบบทดสอบ STOU-EPT ประกอบด้วย 3 ทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน

เมื่อพิจารณาคุณภาพเครื่องมือที่ใช้ในการวิจัยด้านความเป็นเอกมิติของแบบทดสอบ ด้านความเป็นอิสระของข้อสอบ ด้านความเป็นฟังก์ชันเพิ่ม และด้านพารามิเตอร์ความยากและขนาดอิทธิพลของเทสต์เล็ด ผลมีดังนี้

ด้านความเป็นเอกมิติของแบบทดสอบ

ผลการตรวจสอบความเป็นเอกมิติ (unidimensionality) ของแบบทดสอบ STOU-EPT ด้วยวิธีการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) พบว่า องค์ประกอบที่ 1 อธิบายความแปรปรวนของข้อมูลได้ ร้อยละ 9.10 องค์ประกอบที่ 2 และองค์ประกอบที่ 3 อธิบายความแปรปรวนของข้อมูลได้ ร้อยละ 4.10 และร้อยละ 2.28 ตามลำดับ เมื่อพิจารณาสัดส่วนของความแปรปรวนที่องค์ประกอบทั้งสามสามารถอธิบายร่วมกัน พบว่า องค์ประกอบที่ 1 อธิบายได้ร้อยละ 59 ของความแปรปรวนที่สามารถอธิบายได้ทั้งหมด ผลการวิเคราะห์ดังกล่าว สะท้อนให้เห็นว่า ข้อมูลมีโครงสร้างมากกว่าหนึ่งมิติ ซึ่งเป็นหลักฐานว่ามีการละเมิดสมมติฐานความเป็นเอกมิติ แสดงดังตารางที่ 14

ตารางที่ 14 ผลการวิเคราะห์องค์ประกอบเชิงสำรวจ

ค่าสถิติ	องค์ประกอบที่ 1	องค์ประกอบที่ 2	องค์ประกอบที่ 3
SS loadings	9.10	4.10	2.28
Proportion Var	.09	.04	.02
Cumulative Var	.09	.13	.15
Proportion Explained	.59	.26	.15
Cumulative Proportion	.59	.85	1.00

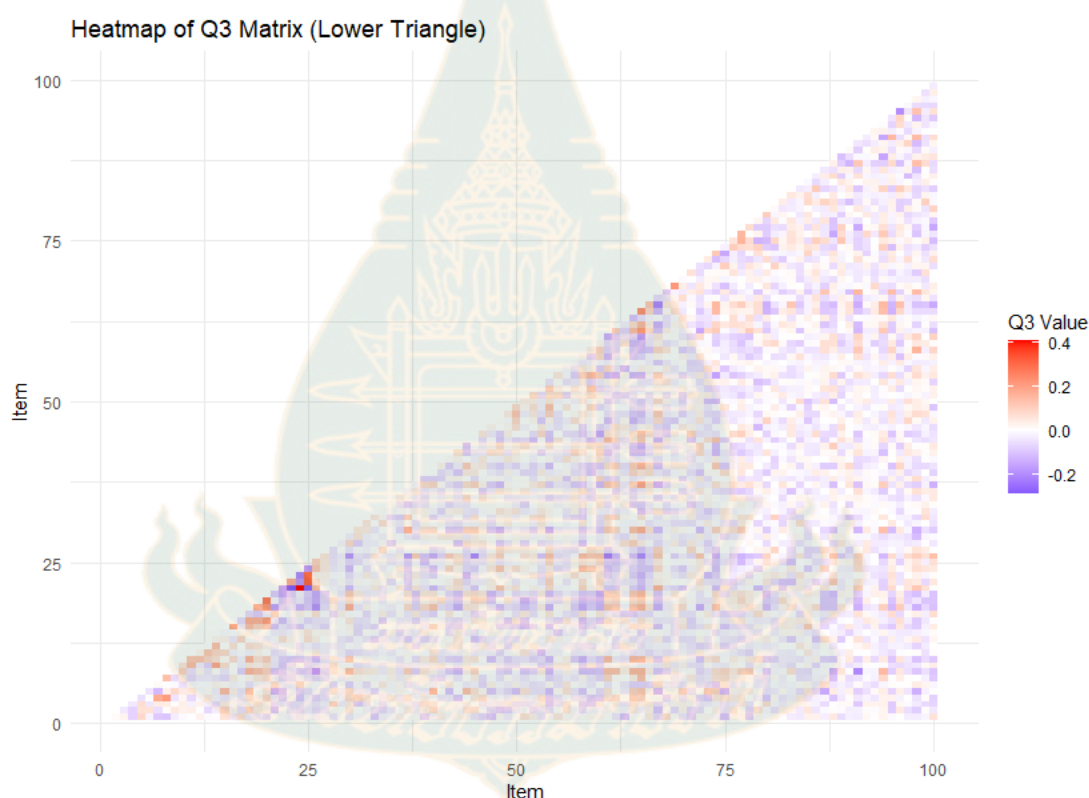
ด้านความเป็นอิสระของข้อสอบ

ผลการตรวจสอบความเป็นอิสระของข้อสอบ STOU-EPT รายข้อ โดยใช้การวิเคราะห์ค่าสัมบูรณ์ของสัมประสิทธิ์สหสัมพันธ์ส่วนเหลือ หรือค่าสถิติ Q3 ที่ได้จากโมเดล Rasch พบว่าค่าสถิติ Q3 ส่วนใหญ่มีค่าต่ำ ซึ่งสอดคล้องกับสมมติฐานความเป็นอิสระ (local independence) ที่ข้อสอบแต่ละข้อควรมีความเป็นอิสระต่อกัน เมื่อควบคุมผลกระทบจากความสามารถแฝงของผู้สอบ เมื่อพิจารณาค่าสถิติ Q3 พบว่า มีจำนวนคู่ของข้อสอบทั้งสิ้น 4,633 คู่ (ร้อยละ 93.60) ที่มีค่าสถิติ Q3 ต่ำกว่า .10 ซึ่งถือว่าอยู่ในระดับต่ำ สะท้อนให้เห็นว่า ข้อสอบส่วนใหญ่มีความเป็นอิสระต่อกันตามสมมติฐานของความเป็นอิสระของข้อสอบ (local independence) อย่างไรก็ตาม พบว่า มีข้อสอบจำนวน 298 คู่ (ร้อยละ 6.02) ที่มีค่าสถิติ Q3 อยู่ในช่วงตั้งแต่ .10 แต่ไม่ถึง .20 โดยค่าช่วงนี้ถือว่าอยู่ในระดับที่ยอมรับได้ในทางปฏิบัติ นอกจากนี้ พบว่า มีคู่ข้อสอบ จำนวน 16 คู่ (ร้อยละ .32) ที่มีค่าสถิติ Q3 อยู่ในช่วงตั้งแต่ .20 แต่ไม่ถึง .30 และ 3 คู่ (ร้อยละ .06) ที่มีค่าสถิติ Q3 ตั้งแต่ .30 ขึ้นไป ซึ่งค่าดังกล่าวถือเป็นสัดส่วนที่น้อยมากเมื่อเทียบกับจำนวนคู่ทั้งหมด แต่ยังสะท้อนถึงความเป็นไปได้ในการเกิดความสัมพันธ์ระหว่างข้อสอบบางคู่ โดยเฉพาะข้อสอบที่ใช้สิ่งเร้าร่วมกัน แสดงดังตารางที่ 15

ตารางที่ 15 ผลการแจกแจงความถี่ของค่าสัมประสิทธิ์สหสัมพันธ์ส่วนเหลือ

ค่าสถิติ Q3	จำนวนคู่	ร้อยละ
ต่ำกว่า .10	4,633	93.60
ตั้งแต่ .10 แต่ไม่ถึง .20	298	6.02
ตั้งแต่ .20 แต่ไม่ถึง .30	16	.32
ตั้งแต่ .30 ขึ้นไป	3	.06
รวม	4,950	100

เมื่อพิจารณาแผนที่ความร้อน (heatmap) แสดงดังภาพที่ 7 ในสามเหลี่ยมล่างของเมทริกซ์ ค่าสถิติ Q3 แสดงให้เห็นว่า การแจกแจงของค่าสถิติ Q3 โดยพื้นที่ส่วนใหญ่ของแผนที่ความร้อนแสดงสีที่บ่งชี้ถึงค่าความสัมพันธ์ที่ต่ำ ขณะที่ค่าความสัมพันธ์ที่สูงจะปรากฏเป็นจุดเล็ก ๆ ที่กระจายตัวอย่างประปราย สะท้อนให้เห็นว่า โดยภาพรวมข้อสอบ STOU-EPT มีคุณภาพที่ดี และสอดคล้องกับสมมติฐานความเป็นอิสระของข้อสอบ ซึ่งเป็นเงื่อนไขสำคัญของทฤษฎีการตอบสนองข้อสอบ อย่างไรก็ตาม มีข้อสอบบางส่วนที่มีค่าสถิติ Q3 สูงกว่าเกณฑ์ $|Q3| > .20$ ดังนั้น เพื่อลดปัญหาความไม่เป็นอิสระกันของข้อสอบ ซึ่งอาจกระทบต่อความแม่นยำในการวัด ผู้วิจัยจึงเลือกใช้ Rasch Testlet Model ในการวิเคราะห์คุณภาพข้อสอบ โดยใช้โปรแกรม R เป็นเครื่องมือในการวิเคราะห์ข้อมูล



ภาพที่ 7 แผนที่ความร้อนของค่าสถิติ Q3 ของข้อสอบ STOU-EPT

ด้านความเป็นฟังก์ชันเพิ่ม

ความเป็นฟังก์ชันเพิ่ม (monotonicity หรือ monotonic increasing function) เป็นความน่าจะเป็นที่ผู้เข้าสอบจะตอบคำถามได้ถูกต้อง ต้องไม่ลดลง เมื่อระดับความสามารถ (θ) เพิ่มขึ้น กล่าวอีกนัยหนึ่ง คือ ผู้มีความสามารถสูงกว่าควรมีโอกาสตอบถูกมากกว่าหรือเท่ากับผู้มีความสามารถต่ำกว่าเสมอ

เพื่อให้การนำเสนอผลการวิเคราะห์มีความชัดเจนและเข้าใจตรงกัน ผู้วิจัยจึงได้กำหนดสัญลักษณ์และคำอธิบายของค่าต่าง ๆ ที่ปรากฏในตารางที่ 16 ดังนี้

สัญลักษณ์ที่ใช้แทนค่าต่าง ๆ

สัญลักษณ์ ความหมาย

itemH ค่า Loevinger's H ของแต่ละข้อ เป็นดัชนีวัดความสามารถในการจำแนกระดับความสามารถของผู้สอบ (ค่ายิ่งสูงยิ่งดี: $H \geq 0.3$ ถือว่าพอใช้, $H \geq 0.4$ ดี)

#ac จำนวน active pairs หรือคู่เปรียบเทียบที่ใช้ในการตรวจสอบ monotonicity ของข้อนั้น

#vi จำนวนคู่ที่ละเมิดความเป็นฟังก์ชันเพิ่ม (violation)

#vi/#ac สัดส่วนของคู่ที่ละเมิดจากคู่ทั้งหมด $= \#vi \div \#ac$

maxvi ขนาดของการละเมิดที่มากที่สุดจากคู่เปรียบเทียบทั้งหมด

sum ผลรวมของขนาดการละเมิดทั้งหมด

sum/#ac ค่าเฉลี่ยของการละเมิดต่อคู่ $(sum \div \#ac)$

zmax ค่าสถิติ z ที่สูงที่สุดที่สัมพันธ์กับการละเมิด monotonicity สำหรับไอเทมนั้น

#zsig จำนวนคู่เปรียบเทียบที่มีการละเมิดที่มีนัยสำคัญทางสถิติ (z-test)

Crit ค่าดัชนีวิกฤต (critical value index) สำหรับแต่ละข้อ ใช้ประเมินระดับความรุนแรงของการละเมิดโดยรวม โดย van der Ark (2007) และ Sijtsma and Molenaar (2002) ได้เสนอเกณฑ์พิจารณา ดังนี้

ค่า crit ระหว่าง 0 – 39 หมายถึง ไม่มีปัญหาข้อสอบแสดงลักษณะ monotonic ได้ดี

ค่า crit ระหว่าง 40 – 79 หมายถึง อาจมีปัญหาเล็กน้อย มีบางช่วงที่ละเมิด อาจพิจารณาไว้เป็นข้อสังเกต

ค่า crit ตั้งแต่ 80 ขึ้นไป หมายถึง มีปัญหารุนแรง ข้อนี้ละเมิดสมมติฐานอย่างชัดเจน ควรพิจารณาตัดออกหรือปรับปรุง

ผลการตรวจสอบสมมติฐานความเป็นฟังก์ชันเพิ่ม พบว่า ค่า ItemH เป็นดัชนีวัดความเป็นความเป็นฟังก์ชันเพิ่มของข้อสอบแต่ละข้อมีความหลากหลาย โดยบางข้อ เช่น ข้อ 24, 26, 58, 67 และ 95 มีค่า $\#vi/\#ac$ สูงเกิน .50 บ่งชี้ว่า มีการละเมิดความเป็นฟังก์ชันเพิ่มอย่างรุนแรง นอกจากนี้ ค่าสูงสุดของ z_{max} ในข้อ 26 เท่ากับ 6.15 บ่งชี้ว่า ข้อสอบข้อ 26 ไม่สอดคล้องกับสมมติฐานที่กำหนดไว้ ส่งผลให้เกิดความไม่แน่นอนในการประมาณค่าระดับความสามารถของผู้เข้าสอบ นอกจากนั้น ข้อสอบบางข้อ เช่น ข้อ 6, 13, 17, 57, 60, 80 และ 91 มีค่า ItemH ต่ำกว่า .10 แสดงว่า ความเป็นฟังก์ชันเพิ่มของข้อสอบในกลุ่มดังกล่าวอยู่ในระดับต่ำมาก ทำให้ข้อสอบไม่สามารถแยกความสามารถของผู้สอบในระดับที่แตกต่างกันได้อย่างเหมาะสม การละเมิดสมมติฐานความเป็นฟังก์ชันเพิ่ม อาจเกิดจากการออกแบบข้อสอบที่ไม่เหมาะสม เช่น ความซับซ้อนของเนื้อหา คำสั่งที่คลุมเครือ หรือการเชื่อมโยงข้อคำถามที่ไม่ชัดเจน ดังนั้น ผู้วิจัยจึงเชิญกรรมการผู้ออกข้อสอบมาพิจารณาปรับเนื้อหาข้อสอบหรือการตรวจสอบคุณภาพเพิ่มเติม โดยเฉพาะข้อสอบที่มีการละเมิดสมมติฐานหรือมีค่าดัชนีวิกฤต (crit) เพื่อให้ข้อสอบสามารถทำหน้าที่เป็นเครื่องมือวัดที่มีความน่าเชื่อถือ และแม่นยำตามมาตรฐานของการวัดและประเมินผลก่อนนำไปใช้ในการกำหนดมาตรฐานเพื่อหาคะแนนจุดตัด แสดงดังตารางที่ 16



ตารางที่ 16 ผลการตรวจสอบสมมติฐานความเป็นฟังก์ชันเพิ่มของข้อสอบ STOU-EPT

item	itemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
1	.07	28	1	.04	.15	.15	.01	2.28	1	64
2	.15	28	0	.00	.00	.00	.00	.00	0	0
3	.16	28	0	.00	.00	.00	.00	.00	0	0
4	.14	28	7	.25	.10	.48	.02	2.25	3	109
5	.22	28	0	.00	.00	.00	.00	.00	0	0
6	.06	28	9	.32	.10	.60	.02	1.53	0	93
7	.12	28	7	.25	.17	.66	.02	2.90	3	128
8	.19	28	1	.04	.05	.05	.00	1.07	0	23
9	.13	28	0	.00	.00	.00	.00	.00	0	0
10	.17	28	1	.04	.04	.04	.00	.74	0	22
11	.09	28	0	.00	.00	.00	.00	.00	0	0
12	.15	28	1	.04	.03	.03	.00	.64	0	21
13	.06	28	4	.14	.10	.23	.01	2.16	1	75
14	.12	28	4	.14	.06	.17	.01	1.27	0	48
15	.16	28	1	.04	.04	.04	.00	.52	0	20
16	.14	28	2	.07	.04	.08	.00	.75	0	29
17	.03	28	4	.14	.10	.29	.01	2.15	2	87
18	.20	28	0	.00	.00	.00	.00	.00	0	0
19	.20	36	0	.00	.00	.00	.00	.00	0	0
20	.19	28	1	.04	.05	.05	.00	1.09	0	23
21	-.02	28	16	.57	.20	1.54	.05	3.79	7	234
22	.02	28	6	.21	.08	.36	.01	1.20	0	70
23	.04	36	7	.19	.08	.38	.01	1.14	0	65
24	.00	36	21	.58	.26	2.29	.06	4.17	11	267
25	.01	28	3	.11	.07	.17	.01	1.11	0	50
26	-.08	36	30	.83	.35	5.19	.14	6.15	21	444

ตารางที่ 16 ผลการตรวจสอบสมมติฐานความเป็นฟังก์ชันเพิ่มของข้อสอบ STOU-EPT (ต่อ)

item	itemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
27	.03	28	8	.29	.07	.34	.01	2.22	2	102
28	.13	28	0	.00	.00	.00	.00	.00	0	0
29	.09	28	2	.07	.04	.07	.00	.57	0	31
30	.19	28	0	.00	.00	.00	.00	.00	0	0
31	.10	28	1	.04	.04	.04	.00	.66	0	25
32	.14	28	0	.00	.00	.00	.00	.00	0	0
33	.09	36	0	.00	.00	.00	.00	.00	0	0
34	.08	28	2	.07	.06	.09	.00	.87	0	35
35	.11	28	0	.00	.00	.00	.00	.00	0	0
36	.07	28	0	.00	.00	.00	.00	.00	0	0
37	.17	28	0	.00	.00	.00	.00	.00	0	0
38	.04	28	0	.00	.00	.00	.00	.00	0	0
39	.14	36	2	.06	.05	.09	.00	.70	0	28
40	.18	28	1	.04	.07	.07	.00	1.06	0	28
41	.09	28	2	.07	.09	.15	.01	1.55	0	45
42	.03	28	2	.07	.06	.09	.00	.89	0	38
43	.03	28	5	.18	.06	.23	.01	1.15	0	58
44	.15	36	0	.00	.00	.00	.00	.00	0	0
45	.16	28	2	.07	.07	.13	.00	1.25	0	36
46	.12	28	5	.18	.08	.24	.01	1.43	0	57
47	.04	36	9	.25	.12	.57	.02	1.86	1	98
48	.02	28	12	.43	.13	.82	.03	3.01	5	166
49	.10	28	5	.18	.06	.24	.01	1.36	0	56
50	.09	28	15	.54	.11	.92	.03	3.70	8	196
51	.05	28	2	.07	.04	.08	.00	.83	0	34
52	.07	28	2	.07	.05	.08	.00	1.09	0	36
53	.11	28	2	.07	.05	.09	.00	.80	0	33
54	-.02	28	11	.39	.09	.80	.03	2.10	3	143
55	.08	28	0	.00	.00	.00	.00	.00	0	0

ตารางที่ 16 ผลการตรวจสอบสมมติฐานความเป็นฟังก์ชันเพิ่มของข้อสอบ STOU-EPT (ต่อ)

item	itemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
56	.11	28	0	.00	.00	.00	.00	.00	0	0
57	.06	28	6	.21	.10	.43	.02	1.95	2	99
58	-.05	28	21	.75	.16	1.59	.06	3.86	9	263
59	.01	28	7	.25	.09	.35	.01	1.70	1	91
60	-.01	28	9	.32	.11	.62	.02	2.01	2	122
61	.20	28	2	.07	.11	.20	.01	1.84	1	59
62	.16	28	1	.04	.04	.04	.00	.80	0	23
63	.09	28	3	.11	.08	.19	.01	1.26	0	48
64	.18	28	1	.04	.04	.04	.00	.61	0	20
65	.20	28	0	.00	.00	.00	.00	.00	0	0
66	.14	28	0	.00	.00	.00	.00	.00	0	0
67	.00	36	22	.61	.24	2.64	.07	3.93	11	277
68	.10	28	0	.00	.00	.00	.00	.00	0	0
69	.12	28	1	.04	.03	.03	.00	.55	0	22
70	.14	28	1	.04	.04	.04	.00	.69	0	24
71	.10	36	4	.11	.08	.24	.01	1.25	0	48
72	.02	36	9	.25	.11	.65	.02	1.70	1	100
73	.08	28	1	.04	.08	.08	.00	1.23	0	35
74	.18	28	2	.07	.15	.18	.01	2.42	1	65
75	.12	36	12	.33	.09	.73	.02	1.50	0	90
76	.15	28	1	.04	.06	.06	.00	1.27	0	29
77	.16	28	2	.07	.03	.07	.00	.70	0	27
78	.14	28	0	.00	.00	.00	.00	.00	0	0
79	.12	28	0	.00	.00	.00	.00	.00	0	0
80	.01	28	5	.18	.12	.37	.01	1.95	1	89
81	.09	28	3	.11	.11	.23	.01	1.84	1	69
82	.12	36	1	.03	.04	.04	.00	.52	0	22
83	.12	28	3	.11	.05	.12	.00	.97	0	39
84	.09	28	5	.18	.09	.39	.01	1.77	2	89

ตารางที่ 16 ผลการตรวจสอบสมมติฐานความเป็นฟังก์ชันเพิ่มของข้อสอบ STOU-EPT (ต่อ)

item	itemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
85	.17	28	2	.07	.06	.10	.00	.93	0	32
86	.12	28	0	.00	.00	.00	.00	.00	0	0
87	.03	36	10	.28	.14	.81	.02	2.24	1	113
88	.01	28	4	.14	.15	.53	.02	2.74	3	114
89	.12	28	5	.18	.08	.25	.01	1.23	0	56
90	.11	28	0	.00	.00	.00	.00	.00	0	0
91	-.01	28	13	.46	.12	.91	.03	2.17	3	158
92	.11	28	3	.11	.05	.13	.00	.97	0	40
93	.09	28	1	.04	.09	.09	.00	1.55	0	37
94	.18	28	1	.04	.07	.07	.00	1.81	1	45
95	.00	36	20	.56	.26	2.58	.07	4.36	13	282
96	.14	28	0	.00	.00	.00	.00	.00	0	0
97	.04	28	10	.36	.14	.88	.03	3.14	6	168
98	-.05	28	16	.57	.13	1.18	.04	3.33	5	201
99	.07	28	3	.11	.08	.19	.01	1.26	0	49
100	.00	28	6	.21	.12	.45	.02	2.37	2	107

ด้านพารามิเตอร์ความยากและขนาดอิทธิพลของเทสต์เล็ต

การวิเคราะห์คุณภาพข้อสอบที่มีโครงสร้างแบบ Testlet ซึ่งมีปัญหาความเป็นอิสระ (local dependence) ของข้อสอบ การใช้โมเดล Testlet Response Theory (TRT) เป็นเครื่องมือที่เหมาะสมในการควบคุมผลกระทบจากความสัมพันธ์ของกลุ่มข้อสอบที่ใช้สิ่งเร้าร่วมกันดังกล่าว โดยเฉพาะการประมาณค่าพารามิเตอร์ผ่านวิธีการ Markov Chain Monte Carlo (MCMC) ซึ่งเป็นวิธีการสุ่มตัวอย่างพารามิเตอร์ซ้ำ ๆ หลายครั้งจากการกระจายความน่าจะเป็นร่วมของข้อมูลทำให้สามารถประมาณค่าพารามิเตอร์ได้อย่างแม่นยำมากขึ้น ในกระบวนการนี้ผู้วิจัยได้นำข้อมูลการตอบข้อสอบถูกมาวิเคราะห์โดยใช้ Rasch Testlet Model ซึ่งมีการกำหนดพารามิเตอร์ที่สำคัญ เช่น ค่าความยาก (difficulty) ขนาดอิทธิพลของเทสต์เล็ต (testlet effect) หรือความแปรปรวนของเทสต์เล็ต (variance of testlet) โดยไม่รวมการประมาณค่าอื่น เช่น ความชัน (slope) การเดา (guessing) ทั้งนี้ กระบวนการประมาณค่าพารามิเตอร์เริ่มต้นด้วยการกำหนดช่วง Burn-in จำนวน 2,000 รอบ เพื่อลดอิทธิพลจากค่าตั้งต้นของโซ่ของมาร์คอฟ (Markov chain) และขจัดความเอนเอียง

ที่อาจเกิดขึ้นในช่วงต้นของการรันโซ่ จากนั้นดำเนินการสุ่มตัวอย่างต่อเนื่องอีก 10,000 รอบ ภายหลังจากที่โซ่ได้ลู่เข้าสู่การแจกแจงภายหลัง (posterior distribution) อย่างมีเสถียรภาพ เพื่อใช้ในการประมาณค่าพารามิเตอร์ ผลการวิเคราะห์นำเสนอในรูปแบบตาราง ซึ่งแสดงค่าพารามิเตอร์ร่วมกับตัวชี้วัดความน่าเชื่อถือ ได้แก่ ส่วนเบี่ยงเบนมาตรฐาน (standard deviation: SD) ของค่าประมาณที่ได้จากการแจกแจง posterior ซึ่งสะท้อนระดับความไม่แน่นอนของค่าประมาณ และดัชนีการลู่เข้า ($\hat{R}$ หรือ Rhat) ซึ่งค่าที่ใกล้เคียง 1.00 บ่งชี้ว่า โซ่ได้ลู่เข้าสู่การแจกแจงเป้าหมายและมีความน่าเชื่อถือในการประมาณค่า

ผลการประมาณค่าพารามิเตอร์ความยาก (b-parameter) ของข้อสอบแต่ละข้อ ภายใต้ Rasch Testlet Model พร้อมทั้งขนาดอิทธิพลของเทสต์เล็ต พบว่า ค่าพารามิเตอร์ความยาก มีการกระจายตัวอยู่ในช่วงที่หลากหลาย สะท้อนให้เห็นว่า ข้อสอบสามารถวัดความสามารถของผู้สอบได้หลายระดับ โดยค่าพารามิเตอร์ที่เป็นลบ เช่น ข้อ 48 ($b[48] = -1.11$) และข้อ 50 ($b[50] = -1.50$) บ่งชี้ว่า เป็นข้อสอบที่ง่าย ผู้เข้าสอบส่วนใหญ่สามารถตอบข้อสอบถูกต้อง ขณะที่ข้อที่มีค่าบวก เช่น ข้อ 3 ($b[3] = .78$) และข้อ 5 ($b[5] = .89$) บ่งชี้ว่า เป็นข้อสอบที่มีความยากในระดับปานกลาง ซึ่งต้องอาศัยผู้สอบที่มีความสามารถสูงกว่าค่าเฉลี่ยจึงจะสามารถตอบได้ถูกต้อง นอกจากนี้ ส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าพารามิเตอร์โดยรวมอยู่ในช่วงระหว่าง .02 ถึง .05 แสดงให้เห็นว่าค่าประมาณของพารามิเตอร์มีความเสถียรและน่าเชื่อถือ อีกทั้ง ดัชนีการลู่เข้า ($\hat{R}$ หรือ Rhat) มีค่าใกล้เคียง 1 ซึ่งสะท้อนว่า โซ่มาร์คอฟได้ลู่เข้าสู่การแจกแจงภายหลังได้อย่างสมบูรณ์

เมื่อพิจารณาขนาดอิทธิพลของเทสต์เล็ต พบว่า ข้อสอบที่อยู่ในกลุ่มเดียวกันมีความสัมพันธ์ภายในที่แตกต่างกัน โดยกลุ่ม testlet L2 (ข้อสอบทักษะการฟัง ข้อที่ 16-20) มีความแปรปรวนของอิทธิพลเทสต์เล็ตสูงสุด เท่ากับ .48 สะท้อนถึงความเชื่อมโยงภายในกลุ่มที่ชัดเจน ขณะที่กลุ่ม testlet R6 (ข้อสอบทักษะการอ่าน ข้อที่ 94-100) มีความแปรปรวนต่ำสุด เท่ากับ .13 ซึ่งแสดงถึงอิทธิพลของเทสต์เล็ตที่น้อยกว่า

ความแตกต่างของขนาดอิทธิพลดังกล่าว สะท้อนให้เห็นว่า ข้อสอบบางกลุ่มอาจมีความเชื่อมโยงกันสูงเนื่องจากการใช้สิ่งเร้าร่วมกันหรืออยู่ภายใต้บริบทเดียวกัน ส่งผลให้ข้อสอบภายในกลุ่มนั้นมีความไม่เป็นอิสระต่อกัน (local item dependence) อย่างไรก็ตาม การประยุกต์ใช้ Rasch Testlet Model สามารถควบคุมและจัดการอิทธิพลของเทสต์เล็ตได้อย่างมีประสิทธิภาพ ส่งผลให้ค่าพารามิเตอร์ที่ประมาณได้มีความน่าเชื่อถือ และสนับสนุนการวิเคราะห์คุณภาพข้อสอบให้มีความถูกต้องแม่นยำยิ่งขึ้น แสดงดังตารางที่ 17

ตารางที่ 17 ผลการประมาณค่าพารามิเตอร์ของข้อสอบและเทสต์เลท

Item	Testlet	Parameter	Mean of parameter	SD	Rhat	แปลค่า b
1	NA	b[1]	-.18	.04	1.00	ปานกลาง
2	NA	b[2]	.35	.04	1.00	ปานกลาง
3	NA	b[3]	.78	.04	1.00	ปานกลาง
4	NA	b[4]	-.66	.04	1.00	ปานกลาง
5	NA	b[5]	.89	.04	1.00	ปานกลาง
6	NA	b[6]	.03	.04	1.00	ปานกลาง
7	NA	b[7]	-.02	.04	1.00	ปานกลาง
8	NA	b[8]	-.54	.04	1.00	ปานกลาง
9	L1_1	b[9]	.10	.04	1.00	ปานกลาง
10	L1_2	b[10]	-.38	.04	1.01	ปานกลาง
11	L1_3	b[11]	-.65	.04	1.01	ปานกลาง
12	L1_4	b[12]	-1.24	.05	1.00	ง่าย
13	L1_5	b[13]	-.62	.04	1.00	ปานกลาง
14	L1_6	b[14]	-.90	.04	1.01	ปานกลาง
15	L1_7	b[15]	.30	.04	1.00	ปานกลาง
16	L2_1	b[16]	-.62	.04	1.00	ปานกลาง
17	L2_2	b[17]	-1.18	.05	1.00	ง่าย
18	L2_3	b[18]	-.51	.04	1.00	ปานกลาง
19	L2_4	b[19]	-.82	.04	1.00	ปานกลาง
20	L2_5	b[20]	-.66	.04	1.00	ปานกลาง
21	L3_1	b[21]	.41	.04	1.00	ปานกลาง
22	L3_2	b[22]	.13	.03	1.00	ปานกลาง
23	L3_3	b[23]	-.03	.04	1.00	ปานกลาง
24	L3_4	b[24]	-.01	.04	1.00	ปานกลาง
25	L3_5	b[25]	.16	.04	1.00	ปานกลาง
26	NA	b[26]	-.40	.03	1.00	ปานกลาง
27	NA	b[27]	-1.47	.05	1.00	ง่าย
28	NA	b[28]	.11	.04	1.00	ปานกลาง
29	NA	b[29]	-.10	.04	1.00	ปานกลาง
30	NA	b[30]	-.53	.04	1.00	ปานกลาง
31	NA	b[31]	-.26	.04	1.01	ปานกลาง
32	NA	b[32]	-.67	.04	1.00	ปานกลาง

ตารางที่ 17 ผลการประมาณค่าพารามิเตอร์ของข้อสอบและเทสต์เลท (ต่อ)

Item	Testlet	Parameter	Mean of parameter	SD	Rhat	แปลค่า b
33	NA	b[33]	.04	.04	1.00	ปานกลาง
34	NA	b[34]	-.25	.04	1.00	ปานกลาง
35	NA	b[35]	-.65	.04	1.00	ปานกลาง
36	NA	b[36]	-.54	.04	1.00	ปานกลาง
37	NA	b[37]	-.48	.04	1.00	ปานกลาง
38	NA	b[38]	-.55	.04	1.00	ปานกลาง
39	NA	b[39]	.11	.04	1.00	ปานกลาง
40	NA	b[40]	.38	.04	1.00	ปานกลาง
41	NA	b[41]	-.41	.04	1.00	ปานกลาง
42	NA	b[42]	-.32	.04	1.00	ปานกลาง
43	NA	b[43]	-.67	.04	1.00	ปานกลาง
44	NA	b[44]	.22	.04	1.00	ปานกลาง
45	NA	b[45]	-.32	.04	1.00	ปานกลาง
46	NA	b[46]	-.53	.04	1.00	ปานกลาง
47	NA	b[47]	-.17	.04	1.00	ปานกลาง
48	NA	b[48]	-1.11	.04	1.01	ง่าย
49	NA	b[49]	-1.03	.04	1.00	ง่าย
50	NA	b[50]	-1.50	.05	1.00	ง่าย
51	NA	b[51]	-1.00	.04	1.00	ปานกลาง
52	NA	b[52]	-.99	.04	1.00	ปานกลาง
53	NA	b[53]	-.30	.04	1.00	ปานกลาง
54	NA	b[54]	-.89	.04	1.00	ปานกลาง
55	NA	b[55]	-.34	.04	1.00	ปานกลาง
56	NA	b[56]	-.28	.04	1.00	ปานกลาง
57	NA	b[57]	-.55	.04	1.00	ปานกลาง
58	NA	b[58]	-.93	.04	1.00	ปานกลาง
59	NA	b[59]	-.97	.04	1.01	ปานกลาง
60	NA	b[60]	-.57	.04	1.00	ปานกลาง
61	R1_1	b[61]	.20	.04	1.00	ปานกลาง
62	R1_2	b[62]	-.25	.04	1.00	ปานกลาง
63	R1_3	b[63]	-.16	.04	1.00	ปานกลาง
64	R1_4	b[64]	-.11	.04	1.00	ปานกลาง

ตารางที่ 17 ผลการประมาณค่าพารามิเตอร์ของข้อสอบและเทสต์เลท (ต่อ)

Item	Testlet	Parameter	Mean of parameter	SD	Rhat	แปลค่า b
65	R1_5	b[65]	-.16	.04	1.00	ปานกลาง
66	R1_6	b[66]	.15	.04	1.00	ปานกลาง
67	R2_1	b[67]	-.43	.04	1.00	ปานกลาง
68	R2_2	b[68]	-.56	.04	1.00	ปานกลาง
69	R2_3	b[69]	-.53	.04	1.00	ปานกลาง
70	R2_4	b[70]	.31	.04	1.00	ปานกลาง
71	R2_5	b[71]	-.17	.04	1.00	ปานกลาง
72	R2_6	b[72]	-.29	.04	1.00	ปานกลาง
73	R2_7	b[73]	.11	.04	1.00	ปานกลาง
74	R3_1	b[74]	.13	.04	1.00	ปานกลาง
75	R3_2	b[75]	-.29	.04	1.01	ปานกลาง
76	R3_3	b[76]	-.52	.04	1.00	ปานกลาง
77	R3_4	b[77]	-.62	.04	1.00	ปานกลาง
78	R3_5	b[78]	-.06	.04	1.01	ปานกลาง
79	R3_6	b[79]	-.11	.04	1.00	ปานกลาง
80	R3_7	b[80]	.07	.04	1.00	ปานกลาง
81	R4_1	b[81]	-.61	.04	1.00	ปานกลาง
82	R4_2	b[82]	.09	.04	1.00	ปานกลาง
83	R4_3	b[83]	-.41	.04	1.00	ปานกลาง
84	R4_4	b[84]	-.87	.04	1.00	ปานกลาง
85	R4_5	b[85]	.80	.04	1.00	ปานกลาง
86	R4_6	b[86]	-.38	.04	1.00	ปานกลาง
87	R5_1	b[87]	-.33	.04	1.00	ปานกลาง
88	R5_2	b[88]	-.18	.04	1.00	ปานกลาง
89	R5_3	b[89]	-.21	.04	1.00	ปานกลาง
90	R5_4	b[90]	-.25	.03	1.00	ปานกลาง
91	R5_5	b[91]	-.50	.04	1.00	ปานกลาง
92	R5_6	b[92]	-.63	.04	1.00	ปานกลาง
93	R5_7	b[93]	-.40	.04	1.00	ปานกลาง
94	R6_1	b[94]	-.77	.04	1.00	ปานกลาง
95	R6_2	b[95]	-.45	.04	1.00	ปานกลาง
96	R6_3	b[96]	-.09	.04	1.00	ปานกลาง

ตารางที่ 17 ผลการประมาณค่าพารามิเตอร์ของข้อสอบและเทสต์เลต (ต่อ)

Item	Testlet	Parameter	Mean of parameter	SD	Rhat	แปลค่า b
97	R6_4	b[97]	-.95	.04	1.01	ปานกลาง
98	R6_5	b[98]	-.97	.04	1.00	ปานกลาง
99	R6_6	b[99]	-.39	.04	1.01	ปานกลาง
100	R6_7	b[100]	-.62	.04	1.00	ปานกลาง
9-15	testlet L1	sigma.testlet[1]	.39	.03	1.05	อิทธิพลสูง
16-20	testlet L2	sigma.testlet[2]	.48	.04	1.00	อิทธิพลสูง
21-25	testlet L3	sigma.testlet[3]	.18	.03	1.05	อิทธิพลต่ำ
61-66	testlet R1	sigma.testlet[4]	.40	.03	1.00	อิทธิพลสูง
67-73	testlet R2	sigma.testlet[5]	.18	.03	1.00	อิทธิพลต่ำ
74-80	testlet R3	sigma.testlet[6]	.22	.03	1.03	อิทธิพลปานกลาง
81-86	testlet R4	sigma.testlet[7]	.22	.03	1.01	อิทธิพลปานกลาง
87-93	testlet R5	sigma.testlet[8]	.13	.02	1.01	อิทธิพลต่ำ
94-100	testlet R6	sigma.testlet[9]	.13	.02	1.03	อิทธิพลต่ำ
MIN b			-1.50			
MAX b			.89			
MEAN b			-.35			

หมายเหตุ: b = ค่าพารามิเตอร์ความยาก sigma.testlet = ความแปรปรวนของอิทธิพลของเทสต์เลต

1.3.3 การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลในครั้งนี้ ผู้วิจัยจัดเตรียมผลการตอบข้อสอบของผู้เข้าสอบ STOU-EPT ระหว่างเดือนพฤษภาคมถึงกันยายน พ.ศ. 2567 โดยประสานงานขอผลการตอบจาก สำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช

1.3.4 การวิเคราะห์ข้อมูล

ผู้วิจัยนำผลการตอบข้อสอบ STOU-EPT ของผู้เข้าสอบ จำนวน 1,272 คน มาวิเคราะห์คุณภาพข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) โดยทำการตรวจสอบข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ เริ่มจากการตรวจสอบความเป็นเอกมิติ (unidimensionality) ด้วยการวิเคราะห์องค์ประกอบ (factor analysis) เพื่อประเมินว่าข้อสอบ STOU-EPT วัดความสามารถเดียวกันหรือไม่ และมีการตรวจสอบความเป็นอิสระของข้อสอบ (local independence) โดยใช้สถิติ Q3 ซึ่งเป็นตัวชี้วัดระดับความสัมพันธ์ที่ไม่พึงประสงค์ระหว่างข้อสอบใน Testlet หากพบว่า ข้อสอบละเมิดสมมติฐานความเป็นเอกมิติหรือความเป็นอิสระของข้อสอบ ผู้วิจัยจะดำเนินการแก้ไขโดยเลือกใช้

โมเดลที่เหมาะสมเพื่อลดอิทธิพลของความสัมพันธ์ระหว่างข้อสอบ และทำให้ผลการวิเคราะห์ IRT มีความแม่นยำมากขึ้น

แบบทดสอบ STOU-EPT ได้รับการออกแบบมาเพื่อวัดทักษะภาษาอังกฤษใน 3 ด้าน ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน โดยในส่วนของ การฟังและการอ่าน ข้อสอบมีลักษณะการใช้สิ่งเร้าร่วมกัน เช่น ทักษะการฟัง ผู้สอบจะได้ฟังเสียงและทำข้อสอบในรูปแบบต่างๆ ได้แก่ การจับคู่ข้อมูลกับสิ่งที่ได้ยิน จำนวน 6 ข้อ การตอบคำถามแบบเติมคำจากเรื่องที่ได้ยิน จำนวน 6 ข้อ และการตอบคำถามในรูปแบบถูก-ผิด จำนวน 5 ข้อ สำหรับทักษะการอ่าน ข้อสอบจะมีเนื้อเรื่องให้ผู้สอบอ่านประมาณ 6 เรื่อง แต่ละเรื่องมีคำถาม 6-7 ข้อ ข้อสอบในลักษณะนี้ เรียกว่า Testlet หมายถึง กลุ่มข้อสอบที่ใช้สิ่งเร้าร่วมกัน เช่น เสียงหรือข้อความ ส่งผลให้ข้อสอบภายในกลุ่มมีความสัมพันธ์กัน (local dependence) ซึ่งอาจส่งผลต่อความแม่นยำในการประเมินคุณภาพข้อสอบ เพื่อจัดการกับปัญหาดังกล่าว ผู้วิจัยจึงเลือกใช้ Rasch Testlet Model ซึ่งเป็นโมเดลที่เหมาะสมสำหรับการจัดการปัญหาความเชื่อมโยงกันของข้อสอบภายใน Testlet ส่งผลให้ผลการวิเคราะห์สะท้อนความสามารถของผู้สอบ และพารามิเตอร์ของข้อสอบได้อย่างถูกต้องและมีความน่าเชื่อถือสูง

การวิเคราะห์มี 4 ขั้นตอน ดังนี้

- 1) ตรวจสอบความเป็นเอกมิติของข้อสอบ STOU-EPT ด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA)
- 2) ตรวจสอบความเป็นอิสระของข้อสอบรายข้อ โดยใช้การวิเคราะห์ค่าสัมบูรณ์ของสัมประสิทธิ์สหสัมพันธ์ส่วนเหลือหรือค่าสถิติ Q3
- 3) ตรวจสอบความเป็นฟังก์ชันเพิ่ม (monotonicity) โดยวิเคราะห์ค่า ItemH ของข้อสอบแต่ละข้อ
- 4) ประเมินค่าพารามิเตอร์ความยาก (b-parameter) และขนาดอิทธิพลการจัดกลุ่มข้อสอบ (testlet effect) โดยประมาณค่าพารามิเตอร์ของข้อสอบด้วย Rasch Testlet Mode

ตอนที่ 2 การเทียบเคียงคะแนนแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีการของแองกอฟที่ได้รับการปรับปรุง

การดำเนินการวิจัยในตอนนี้ แบ่งออกเป็น 2 ตอนย่อย ได้แก่ ตอนที่ 2.1 การกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดแบบทดสอบ STOU-EPT และตอนที่ 2.2 การประเมินกระบวนการกำหนดคะแนนจุดตัด รายละเอียดแต่ละตอนมีดังนี้

ตอนที่ 2.1 การกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดแบบทดสอบ STOU-EPT

การดำเนินงานในตอนนี้ แบ่งออกเป็น 4 ส่วน ได้แก่ ผู้ให้ข้อมูลหลัก เครื่องมือที่ใช้ในการวิจัย การเก็บรวบรวมข้อมูล และการวิเคราะห์ข้อมูล รายละเอียดแต่ละส่วนมีดังนี้

2.1.1 ผู้ให้ข้อมูลหลัก

ผู้ให้ข้อมูลหลักเป็นผู้เชี่ยวชาญในการกำหนดมาตรฐานเพื่อหาคะแนนจุดตัด โดยผู้วิจัยกำหนดจำนวนผู้เชี่ยวชาญไว้ทั้งหมด 12 คน ซึ่งคัดเลือกแบบเจาะจงตามเกณฑ์การคัดเลือกที่กำหนดไว้ล่วงหน้า ผู้วิจัยติดต่อผู้เชี่ยวชาญโดยตรงผ่านทางโทรศัพท์เพื่อเชิญชวนเข้าร่วมกำหนดมาตรฐานจากจำนวนนี้มีผู้ตอบรับเข้าร่วม 10 คน ซึ่งมีคุณสมบัติตรงตามเกณฑ์ที่กำหนดไว้ ทั้งนี้ เนื่องจากการกำหนดมาตรฐานในบริบทที่มีความสำคัญสูง (high-stakes) มักใช้ผู้เชี่ยวชาญจำนวน 15–20 คน เพื่อให้มั่นใจในความสอดคล้องของผลการตัดสิน (Cizek & Bunch, 2007) ขนาดของผู้เชี่ยวชาญที่น้อยกว่ามาตรฐานดังกล่าวจึงจำเป็นต้องมีการตรวจสอบความถูกต้องเพิ่มเติม ผู้วิจัยจึงใช้การวิเคราะห์ความแม่นยำของการจำแนก (classification accuracy: CA) เพื่อเปรียบเทียบผลการจำแนกระดับความสามารถที่ได้จากคะแนนจุดตัดของผู้เชี่ยวชาญกับระดับความสามารถที่ประมาณโดยทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งเป็นการตรวจสอบจากข้อมูลจริงว่าผลการกำหนดคะแนนจุดตัดสุดท้ายของผู้เชี่ยวชาญสอดคล้องกับผลการจำแนกจากโมเดล IRT หรือไม่

ผู้วิจัยได้กำหนดเกณฑ์การคัดเลือกผู้เชี่ยวชาญ ตลอดจนขั้นตอนและกระบวนการขอความยินยอมในการเข้าร่วมประเมินความสอดคล้อง ดังนี้

2.1.1.1 เกณฑ์การคัดเลือกผู้เชี่ยวชาญ ประกอบด้วย เกณฑ์การคัดเลือกเข้า เกณฑ์การคัดออก และเกณฑ์การหยุดโครงการวิจัย

เกณฑ์การคัดเลือกเข้า

1) เป็นอาจารย์ประจำแขนงวิชาภาษาอังกฤษหรือหลักสูตรและการสอนภาษาอังกฤษ มหาวิทยาลัยสุโขทัยธรรมมาธิราชที่ไม่มีส่วนร่วมในการออกข้อสอบ STOU-EPT หรืออาจารย์ประจำศูนย์ภาษา สถาบันภาษา สาขาวิชาภาษาอังกฤษจากสถาบันอุดมศึกษาอื่น รวมถึงผู้เชี่ยวชาญด้านภาษาอังกฤษในกระทรวงศึกษาธิการ

2) เป็นผู้มีความรู้ในการศึกษาระดับปริญญาตรี ปริญญาโท และ/หรือปริญญาเอก ในทิศทางเดียวกัน ได้แก่ การสอนภาษาอังกฤษ ภาษาศาสตร์ หรือการสื่อสารระหว่างวัฒนธรรม

3) เป็นผู้ที่มีประสบการณ์เกี่ยวกับการทดสอบภาษาอังกฤษตามกรอบ CEFR เช่น การออกแบบและพัฒนาแบบทดสอบ การฝึกอบรมผู้สอบและกรรมการประเมิน การตรวจข้อสอบและการประเมินผล การวิจัยและพัฒนาเกี่ยวกับการทดสอบตามกรอบ CEFR

4) เป็นผู้ที่มีประสบการณ์ในการสอนภาษาอังกฤษในระดับอุดมศึกษา ตั้งแต่ 5 ปีขึ้นไป และเคยมีส่วนร่วมในการประเมินผลหรือการพัฒนาหลักสูตรตามกรอบ CEFR

เกณฑ์การคัดออก

ผู้เชี่ยวชาญที่กำหนดมาตรฐานติดภารกิจหรือไม่สามารถเข้าร่วมกิจกรรม กำหนดมาตรฐานตามวัน เวลา และสถานที่ที่กำหนด

เกณฑ์การหยุดโครงการวิจัย

ผู้เชี่ยวชาญที่กำหนดมาตรฐานสามารถออกจากการวิจัยเมื่อใดก็ได้ ไม่จำเป็นต้อง บอกเหตุผล

2.1.1.2 ขั้นตอนและกระบวนการขอความยินยอมในการเข้าร่วมกิจกรรมกำหนด มาตรฐานมีดังนี้

1) ผู้วิจัยประสานงานไปยังผู้เชี่ยวชาญที่กำหนดมาตรฐาน ผ่านทางโทรศัพท์ เพื่อเชิญชวนผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญรับฟังรายละเอียดของโครงการวิจัย

2) ผู้วิจัยเป็นผู้อธิบายกิจกรรมที่ให้ผู้ที่จะเป็นผู้เชี่ยวชาญทำอย่างละเอียดตามที่กำหนดไว้ในเอกสารแนะนำสำหรับผู้เชี่ยวชาญที่เข้าร่วมกำหนดมาตรฐานโดยไม่มีการใช้ สื่อเสริมอื่น ๆ นอกเหนือจากการพูดอธิบาย ในระหว่างอธิบายกิจกรรมผู้วิจัยมีการสอบถามความ เข้าใจผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญว่า มีคำถาม ข้อสงสัย หรือข้อวิตกกังวลหรือไม่ หากมีคำถามผู้วิจัยจะ ตอบคำถามต่าง ๆ ด้วยความเต็มใจ ไม่ปิดบัง ซ่อนเร้น เมื่อมีการอธิบายกิจกรรมให้กับผู้ที่คาดว่าจะ เป็นผู้เชี่ยวชาญครบถ้วนจนหมดข้อสงสัย ผู้วิจัยก็สรุปใจความสำคัญเกี่ยวกับกิจกรรมที่ให้ผู้ที่จะเป็นผู้ ข้าราชการต้องทำ

3) ผู้วิจัยให้โอกาสและเวลาผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญ ประมาณ 1 สัปดาห์ พิจารณาตัดสินใจว่าจะเข้าหรือไม่เข้าร่วมกำหนดมาตรฐาน พร้อมทั้งแจ้งว่าจะส่งเอกสารข้อ 3.1) ถึง 3.3) ไปยังผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญทางไปรษณีย์ เพื่อให้โอกาสผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญได้คิด และ/หรือปรึกษากับผู้ที่เกี่ยวข้องก่อนตัดสินใจเข้าร่วมกำหนดมาตรฐาน โดยเอกสารที่จัดส่งมีดังนี้

3.1) เอกสารแนะนำสำหรับผู้เชี่ยวชาญที่เข้าร่วมกำหนดมาตรฐาน เป็นเอกสารที่มีรายละเอียดเกี่ยวกับโครงการวิจัยที่ผู้วิจัยได้ชี้แจงแก่ผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญ เพื่อให้ได้อ่านทำความเข้าใจ และพิจารณาตัดสินใจเข้าร่วมกำหนดมาตรฐาน ทั้งนี้หากผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญมีข้อซักถามเพิ่มเติม ให้สอบถามผู้วิจัยได้จนหมดข้อสงสัย โดยติดต่อสอบถามผู้วิจัยโดยตรง

3.2) หนังสือแสดงความยินยอมเข้าร่วมโครงการวิจัย หลังจากผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญอ่านทำความเข้าใจข้อมูลในเอกสารแนะนำสำหรับผู้เชี่ยวชาญอย่างรอบคอบ จนหมดข้อสงสัย และตัดสินใจเข้าร่วมกำหนดมาตรฐาน ให้ผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญแสดงความยินยอม โดยการลงลายมือชื่อด้วยตนเองในหนังสือแสดงความยินยอมเข้าร่วมโครงการวิจัยที่ได้จัดส่งให้

3.3) ชองติดแสตมป์ หลังจากผู้ที่คาดว่าจะเป็นผู้เชี่ยวชาญลงลายมือชื่อด้วยตนเองในหนังสือแสดงความยินยอมเข้าร่วมโครงการวิจัยที่ได้จัดส่งให้แล้ว ให้นำหนังสือแสดงความยินยอมเข้าร่วมโครงการวิจัยใส่ซองติดแสตมป์ที่ผู้วิจัยจัดให้ ส่งกลับคืนผู้วิจัย

4) ผู้วิจัยจัดส่งเอกสารประกอบการกำหนดมาตรฐาน พร้อมทั้งแจ้งกำหนดวัน เวลา และสถานที่ที่จัดกิจกรรมกำหนดมาตรฐานไปยังผู้เชี่ยวชาญทางไปรษณีย์

5) ผู้เชี่ยวชาญเข้าร่วมกิจกรรมกำหนดมาตรฐานตามวัน เวลา และสถานที่ที่ผู้วิจัยกำหนด โดยในการเข้าร่วมกิจกรรมกำหนดมาตรฐาน อาสาสมัครมีอิสระในการที่จะกำหนดหรือไม่กำหนดมาตรฐานในบางกิจกรรม หรืออภิปรายหรือไม่อภิปราย ระหว่างสมาชิกในกลุ่มก็ได้

2.1.2 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยมี 4 ชุด ได้แก่

ชุดที่ 1 แบบทดสอบ STOU-EPT พัฒนาโดยคณาจารย์แขนงวิชาภาษาอังกฤษ สาขาวิชาศิลปศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช ภายใต้โครงการปรับปรุงข้อสอบวัดความรู้ ภาษาอังกฤษด้วยระบบคอมพิวเตอร์ STOU-EPT ในปี พ.ศ. 2566

ชุดที่ 2 คู่มือการกำหนดคะแนนจุดตัดด้วยวิธีเองกอฟที่ได้รับการปรับปรุง (รายละเอียด ดังภาคผนวก ข)

ชุดที่ 3 แบบประมาณโอกาสการทำข้อสอบถูกสำหรับผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถระดับต่ำสุด ในระดับ B1 (รายละเอียดดังภาคผนวก ข)

ชุดที่ 4 แบบประมาณโอกาสการทำข้อสอบถูกสำหรับผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถระดับต่ำสุด ในระดับ B2 (รายละเอียดดังภาคผนวก ข)

2.1.3 การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูล ผู้วิจัยดำเนินการโดยจัดประชุมผู้เชี่ยวชาญ จำนวน 10 คน เพื่อกำหนดคะแนนจุดตัดของแบบทดสอบ STOU-EPT ครอบคลุม 3 ทักษะ ได้แก่ การฟัง โครงสร้าง ภาษา และการอ่าน ทั้งนี้ การดำเนินการจัดประชุมผู้เชี่ยวชาญเพื่อกำหนดคะแนนจุดตัดดำเนินการ 3 รอบในแต่ละทักษะ

การดำเนินงานในส่วนนี้ แบ่งออกเป็น 4 ขั้นตอน ได้แก่ กำหนดผู้เชี่ยวชาญ ปฐมนิเทศผู้เชี่ยวชาญก่อนกำหนดคะแนนจุดตัด ฝึกซ้อมทำการกำหนดจุดตัดให้กับผู้เชี่ยวชาญ และกำหนดมาตรฐานเพื่อหาคะแนนจุดตัด รายละเอียดแต่ละขั้นตอนมีดังนี้

ขั้นตอนที่ 1 กำหนดผู้เชี่ยวชาญ

ผู้เชี่ยวชาญที่ได้รับเชิญให้เข้าร่วมกำหนดคะแนนจุดตัด มีจำนวนทั้งสิ้น 10 คน โดยใช้วิธีการเลือกแบบเจาะจง (purposive sampling) ซึ่งเป็นผู้ที่มีคุณสมบัติตรงตามเกณฑ์การคัดเลือกที่กำหนดไว้

ขั้นตอนที่ 2 ปฐมนิเทศผู้เชี่ยวชาญก่อนกำหนดคะแนนจุดตัด

ก่อนดำเนินการกำหนดคะแนนจุดตัด ผู้วิจัยได้ปฐมนิเทศผู้เชี่ยวชาญ (รายชื่อดังภาคผนวก ค) เพื่ออธิบายแนวทางการกำหนดคะแนนจุดตัดด้วยวิธีเองกอฟฟ์ที่ได้รับการปรับปรุง โดยเน้นการสร้างความเข้าใจร่วมกันในประเด็นสำคัญดังนี้

2.1 แนะนำคู่มือและแบบฟอร์มที่เกี่ยวข้อง ผู้วิจัยแนะนำวิธีการใช้คู่มือการกำหนดคะแนนจุดตัด และอธิบายวิธีการบันทึกข้อมูลลงในแบบฟอร์มที่จัดเตรียมไว้ เพื่อให้ผู้เชี่ยวชาญเข้าใจกระบวนการ และสามารถบันทึกผลการประเมินได้อย่างถูกต้อง

2.2 คำบรรยายคุณลักษณะในระดับความสามารถต่าง ๆ ตามกรอบ CEFR ผู้วิจัยอธิบายรายละเอียดของคุณลักษณะผู้สอบในแต่ละระดับความสามารถ ได้แก่ ระดับ A2 B1 และ B2 จำแนกตามทักษะการฟัง โครงสร้างภาษา และการอ่าน เพื่อช่วยให้ผู้เชี่ยวชาญสามารถเชื่อมโยงคุณลักษณะเหล่านี้กับข้อสอบได้อย่างถูกต้อง

2.3 ลักษณะของผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบ (borderline examinees) ผู้วิจัยเน้นอธิบายลักษณะของผู้สอบที่อยู่ในระดับก้ำกึ่งระหว่างคะแนนจุดตัด 2 ระดับ ได้แก่ ระดับ B1 และ B2 จำแนกตามทักษะการฟัง โครงสร้างภาษา และการอ่าน เพื่อช่วยให้ผู้เชี่ยวชาญสามารถพิจารณาและตัดสินใจในการกำหนดคะแนนจุดตัดได้อย่างแม่นยำ แสดงดังตารางที่ 18

เพื่อให้ผู้เชี่ยวชาญทุกคนเข้าใจแนวทางการทำงานได้อย่างชัดเจน ผู้วิจัยได้เปิดโอกาสให้ซักถาม แสดงความคิดเห็น และปรับความเข้าใจร่วมกันให้ตรงกันก่อนเริ่มกระบวนการกำหนดคะแนนจุดตัด

ตารางที่ 18 คุณลักษณะของผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบหรือระดับก้ำกึ่ง จำแนกตามทักษะ

STOU's CEFR Proficiency Borderlines	STOU's CEFR Proficiency Borderlines
B1 Borderline	B2 Borderline
Listening Borderlines	
A learner at the A2-B1 borderline has a good understanding of simple, clear information on familiar topics and is beginning to comprehend more standard-paced, everyday conversations. Familiar Topics: The learner can understand straightforward questions and short conversations on familiar topics (family, work, daily life) but may need some repetition or clarification to grasp details, especially when the speech is at a standard pace. Public Announcements: They can understand key information in simple announcements, such as transportation or weather updates, and may be able to follow more complex public messages, though still benefiting from slower delivery or additional support.	At this level, the learner can understand the main ideas and some details in standard-paced conversations and discussions on familiar topics, but they may struggle with more nuanced language or complex topics, as typical at B2. Main Points in Conversations: They can follow the main points of clear speech on familiar topics (e.g., everyday discussions, news reports, TV programs) without needing as much repetition. However, they may still need support or clarification for less straightforward details or abstract ideas. Short Talks and Discussions: The learner can handle straightforward talks or presentations on familiar topics and may begin to follow more in-depth discussions but often misses subtleties or complex arguments, especially if the topic is unfamiliar.

ตารางที่ 18 คุณลักษณะของผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบหรือระดับก้ำกึ่ง จำแนกตามทักษะ (ต่อ)

STOU's CEFR Proficiency Borderlines	STOU's CEFR Proficiency Borderlines
B1 Borderline	B2 Borderline
Listening Borderlines	
TV News and Programs: They can follow changes in topics in simple news reports and understand the main points, though they may struggle with details and complex topics. They can follow short, simple discussions on familiar topics but may not fully grasp content in longer or less familiar conversations. Main Challenge: The learner is still working toward a more consistent understanding of standard speech on everyday topics and may <u>rely on repetition or clarification for accuracy</u>	Understanding Details: They can catch key information in simple technical discussions (like instructions for familiar equipment), though complex technical discussions or academic presentations remain challenging. TV Programs and Films: They can follow simple dramas, talk shows, or news programs where language is straightforward, though they may still miss idiomatic expressions, emotional undertones, or quick exchanges that require a more nuanced understanding. Main Challenge: The learner is on the verge of understanding complex speech but may <u>lack the consistency needed to fully handle abstract topics, nuanced expressions, or rapid speech typical of B2-level listening.</u>

ตารางที่ 18 คุณลักษณะของผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบหรือระดับก้ำกึ่ง จำแนกตามทักษะ (ต่อ)

STOU's CEFR Proficiency Borderlines	STOU's CEFR Proficiency Borderlines
B1 Borderline	B2 Borderline
Structure Borderlines	
A learner at this borderline has a strong foundation in basic grammar structures and is beginning to use more complex structures associated with the B1 level. Present, Past, and Future Tenses: The learner can use common tenses like the present continuous, past simple, and future forms (e.g., “will” and “going to”). They are beginning to use and understand the present perfect but may struggle with distinguishing it from the past simple, especially in more nuanced contexts typical of B1. Modals and Conditionals: They can handle basic modals (e.g., <i>can</i>, <i>could</i>, <i>should</i>), especially in everyday scenarios, and have some familiarity with the first conditional. However, they may be inconsistent with B1-level modals (like <i>must</i>, <i>might</i>, and <i>may</i>) and have limited use of second and third conditionals.	At the B1-B2 borderline, a learner has a good grasp of B1-level structures and is beginning to work with more advanced B2-level grammar, although they may not consistently use them accurately. Complex Tenses: The learner can use the past perfect and present perfect continuous with some accuracy and may start to understand the distinction between narrative tenses (past simple, past continuous, past perfect) but may not yet use these fluently or in more extended contexts typical of B2. Modals and Conditionals: They can manage a wider range of B1 modals (e.g., <i>might</i>, <i>may</i>, <i>must</i>) and are beginning to use modals for deduction and speculation (e.g., <i>must have</i>, <i>might have</i>) but may need support with B2-level modals (<i>can't have</i>, <i>needn't have</i>) and mixed conditionals, especially in unfamiliar contexts.

ตารางที่ 18 คุณลักษณะของผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบหรือระดับก้ำกึ่ง จำแนกตามทักษะ (ต่อ)

STOU's CEFR Proficiency Borderlines	STOU's CEFR Proficiency Borderlines
B1 Borderline	B2 Borderline
Structure Borderlines	
Questions and Sentence Structure: They can form basic wh-questions in the past and present and are starting to understand more complex question forms (such as question tags), but they may need more accuracy to reach B1 fully. Adjectives and Comparatives: They understand basic comparative and superlative structures but may not yet handle a broader range of intensifiers (e.g., <i>too, enough</i>) typical of B1. Main Challenge: They are progressing toward B1-level grammar but may <u>lack consistency and control over more complex structures like perfect tenses, question tags, and varied conditionals, often relying on simpler sentence forms in practice.</u>	Modals and Conditionals: They can manage a wider range of B1 modals (e.g., <i>might, may, must</i>) and are beginning to use modals for deduction and speculation (e.g., <i>must have, might have</i>) but may need support with B2-level modals (<i>can't have, needn't have</i>) and mixed conditionals, especially in unfamiliar contexts. Passive and Reported Speech: They can form simple passive sentences and use basic reported speech with some accuracy, although they may struggle with more complex or embedded forms of these structures that are typical at B2. Connecting Words and Relative Clauses: The learner can use basic linking words for cause and effect, contrast, etc., and is beginning to form more varied sentence structures with relative clauses, but they may lack full fluency and may overuse simpler structures. Main Challenge: They are approaching B2 but often <u>lack control and fluency with more advanced structures, like mixed conditionals, extended passive forms, and nuanced modal usage, relying instead on simpler B1 grammar when under pressure.</u>

ตารางที่ 18 คุณลักษณะของผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบหรือระดับก้ำกึ่ง จำแนกตามทักษะ (ต่อ)

STOU's CEFR Proficiency Borderlines	STOU's CEFR Proficiency Borderlines
B1 Borderline	B2 Borderline
Reading Borderlines	
A learner at the A2-B1 borderline can read and understand straightforward texts on familiar topics but may struggle with more detailed or less familiar content. Understanding Familiar Texts: They can follow short, clear texts containing familiar vocabulary and topics of personal interest (e.g., emails, social media posts, and basic instructions) and are beginning to handle B1-level texts on familiar topics, such as short newspaper articles and simple brochures, though they may miss finer details. Advertisements and Information Texts: They can locate essential information in ads, brochures, and timetables but may find it challenging to fully understand content that includes more complex vocabulary or unfamiliar ideas.	A learner at the B1-B2 borderline is increasingly comfortable with standard B1 texts and is beginning to manage more complex B2-level reading, though they may need frequent support with challenging vocabulary and ideas. Factual Texts and Articles: They can read factual articles, reports, and short texts in magazines or online on familiar topics, often well enough to discuss them afterward. However, they may struggle with texts that contain more detailed arguments, specialized vocabulary, or subtle viewpoints that are typical of B2-level material.

ตารางที่ 18 คุณลักษณะของผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบหรือระดับก้ำกึ่ง จำแนกตามทักษะ (ต่อ)

STOU's CEFR Proficiency Borderlines	STOU's CEFR Proficiency Borderlines
B1 Borderline	B2 Borderline
Reading Borderlines	
Short Stories and Messages: The learner can understand everyday stories and short descriptions when visual support is available, and they may attempt to read simplified short stories or novels. However, they may rely heavily on a dictionary to understand unknown words and may miss subtler story details or emotions. Main Challenge: The learner is progressing toward B1 but still <u>needs support with vocabulary and comprehension for texts that are longer or contain unfamiliar vocabulary.</u> They may also find it <u>difficult to follow a storyline consistently without visuals or additional context.</u>	Personal Correspondence and Instructions: The learner can read and respond to personal letters discussing feelings and events, and they are becoming comfortable with more structured written instructions (e.g., for using appliances or following a recipe). However, lengthy or complex instructions, especially if they involve new terminology or detailed technical processes, remain challenging. Simplified Literature and Structured Stories: They can follow storylines in simplified novels or short stories with clear structures, though they may still require a dictionary for more difficult vocabulary. They can also understand basic reviews or opinion pieces on familiar topics, though they may miss subtle viewpoints or advanced expressions. Main Challenge: The learner is working towards B2 but may <u>lack the vocabulary range and fluency needed to independently handle complex or opinionated texts, longer narratives, or technical reading without relying on reference tools for support.</u>

ขั้นตอนที่ 3 ฝึกซ้อมทำการกำหนดจุดตัดให้กับผู้เชี่ยวชาญ

หลังจากผู้เชี่ยวชาญได้รับการอธิบายแนวทางการกำหนดคะแนนจุดตัดด้วยวิธี แอวกอฟที่ได้รับการปรับปรุง และได้ทำความเข้าใจแนวทางดังกล่าวร่วมกันจนมีความชัดเจนตรงกัน แล้ว ผู้วิจัยได้ให้ผู้เชี่ยวชาญฝึกซ้อมการกำหนดคะแนนจุดตัด การวิจัยครั้งนี้มีการกำหนด คะแนนจุดตัด 2 ระดับความสามารถ ได้แก่ ระดับ B1 และระดับ B2

ขั้นตอนการฝึกซ้อมทำการกำหนดคะแนนจุดตัด มีดังนี้

- 1) ให้ผู้เชี่ยวชาญอ่านและทำความเข้าใจเนื้อหา และระดับความยากของข้อสอบแต่ละข้อ โดยอิงจากประสบการณ์การสอนภาษาอังกฤษของคน
- 2) ให้จินตนาการว่า

ระดับ B1

“ผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถต่ำสุดในระดับ B1 จะทำข้อสอบข้อที่ i ถูกก็เปอร์เซ็นต์” แล้วให้ผู้เชี่ยวชาญบันทึกค่าที่ประมาณได้ลงในช่องโอกาสการทำข้อสอบถูกในแบบประมาณโอกาสการทำข้อสอบถูกสำหรับผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถระดับต่ำสุดในระดับ B1

ระดับ B2

“ผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถต่ำสุดในระดับ B2 จะทำข้อสอบข้อที่ i ถูกก็เปอร์เซ็นต์” แล้วให้ผู้เชี่ยวชาญบันทึกค่าที่ประมาณได้ลงในช่องโอกาสการทำข้อสอบถูกลงในแบบประมาณโอกาสการทำข้อสอบถูกสำหรับผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถระดับต่ำสุดในระดับ B2

โดยเปอร์เซ็นต์การทำข้อสอบถูกแต่ละข้อ มีค่าตั้งแต่ 0 ถึง 100 เช่น

0% หมายถึง ไม่มีโอกาสทำข้อสอบข้อนั้นถูกเลย

50% หมายถึง มีโอกาสทำข้อสอบข้อนั้นถูกเพียงครึ่งหนึ่ง

100% หมายถึง มีโอกาสทำข้อสอบข้อนั้นถูกแน่นอน

หลังจากผู้เชี่ยวชาญทำการฝึกซ้อมการกำหนดคะแนนจุดตัดเสร็จเรียบร้อยแล้ว ผู้วิจัยนำผลการตอบของผู้เชี่ยวชาญมาคำนวณค่าสถิติ ได้แก่ ค่าเฉลี่ย (mean: M) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation: SD) ค่าต่ำสุด (minimum: Min) ค่าสูงสุด (maximum: Max) และค่าความคลาดเคลื่อนมาตรฐานของการตัดสิน (standard error of judgment: SEJ) เพื่อประเมินว่าผู้เชี่ยวชาญมีความคิดเห็นสอดคล้องกันเพียงใด และผู้วิจัยเปิดโอกาสให้ผู้เชี่ยวชาญได้ซักถาม และแลกเปลี่ยนความคิดเห็นเกี่ยวกับผลการตอบ จนมั่นใจว่าผู้เชี่ยวชาญมีความเข้าใจและมีความคิดเห็นสอดคล้องกัน

ทั้งนี้ การดำเนินงานในขั้นตอนที่ 2 และขั้นตอนที่ 3 แล้วเสร็จภายใน 1 วัน

ขั้นตอนที่ 4 กำหนดมาตรฐานเพื่อหาคะแนนจุดตัด

การกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดของแบบทดสอบ STOU-EPT ผู้วิจัยใช้วิธีการของแองกอฟที่ได้รับการปรับปรุง โดยกำหนดคะแนนจุดตัดเป็น 2 แบบ คือ คะแนนจุดตัดรวม (ข้อ 1-100) และคะแนนจุดตัดแยกตามทักษะ ได้แก่ การฟัง (ข้อ 1-25 รวม 25 ข้อ) โครงสร้างภาษา (ข้อ 26-60 รวม 35 ข้อ) และการอ่าน (ข้อ 61-100 รวม 40 ข้อ) โดยจุดตัดแต่ละแบบ แบ่งเป็น 2 จุด ได้แก่ ระดับ B1 (อยู่ระหว่างระดับ A2 กับ B1) และระดับ B2 (อยู่ระหว่างระดับ B1 กับ B2) ซึ่งการกำหนดมาตรฐานเพื่อหาคะแนนจุดตัด มีการดำเนินการ จำนวน 3 รอบ ดังนี้

รอบที่ 1

การกำหนดคะแนนจุดตัด รอบที่ 1 ผู้วิจัยให้ผู้เชี่ยวชาญแต่ละคนกำหนดคะแนนจุดตัดข้อสอบ STOU-EPT จำนวน 100 ข้อ ซึ่งครอบคลุม 3 ทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน โดยแต่ละทักษะกำหนดคะแนนจุดตัดใน 2 ระดับ ดังนี้

ระดับ B1 ให้ผู้เชี่ยวชาญอ่านและทำความเข้าใจเนื้อหา และระดับความยากของข้อสอบแต่ละข้อ โดยอิงจากประสบการณ์การสอนภาษาอังกฤษของคน แล้วจินตนาการว่า “ผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถต่ำสุดในระดับ B1 จะทำข้อสอบข้อที่ i ถูกกี่เปอร์เซ็นต์” ทำเช่นนี้ตั้งแต่ข้อ 1-100 พร้อมบันทึกค่าที่ประมาณได้ลงในช่องโอกาสการทำข้อสอบถูกในแบบประมาณโอกาสการทำข้อสอบถูกสำหรับผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถระดับต่ำสุดในระดับ B1

ระดับ B2 ให้ผู้เชี่ยวชาญอ่านและทำความเข้าใจเนื้อหา และระดับความยากของข้อสอบแต่ละข้อ โดยอิงจากประสบการณ์การสอนภาษาอังกฤษของคน แล้วจินตนาการว่า “ผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถต่ำสุดในระดับ B2 จะทำข้อสอบข้อที่ i ถูกกี่เปอร์เซ็นต์” ทำเช่นนี้ตั้งแต่ข้อ 1-100 พร้อมบันทึกค่าที่ประมาณได้ลงในช่องโอกาสการทำข้อสอบถูกในแบบประมาณโอกาสการทำข้อสอบถูกสำหรับผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถระดับต่ำสุดในระดับ B2

การรวบรวมและการวิเคราะห์ข้อมูล ผู้วิจัยรวบรวมผลการตอบของผู้เชี่ยวชาญทุกคนที่ได้จากการดำเนินงานในรอบที่ 1 มาคำนวณค่าสถิติ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด และค่าความคลาดเคลื่อนมาตรฐานของการตัดสินใจ

การนำเสนอผลและให้ข้อมูลย้อนกลับ หลังจากผู้วิจัยได้คำนวณค่าสถิติจากการดำเนินงานในรอบที่ 1 ผู้วิจัยได้นำเสนอผลการวิเคราะห์ให้ผู้เชี่ยวชาญทราบ โดยผู้เชี่ยวชาญแต่ละคนจะทราบผลการตอบของตนเอง และผลการตอบในภาพรวมของกลุ่ม จากนั้นเปิดโอกาสให้ผู้เชี่ยวชาญแสดงความคิดเห็นและแลกเปลี่ยนความรู้ซึ่งกันและกันเกี่ยวกับผลการตอบ โดยเฉพาะผู้เชี่ยวชาญบางคนที่มีผลการตอบแตกต่างจากกลุ่ม ซึ่งผู้เชี่ยวชาญสามารถอธิบายเหตุผลและมุมมองเกี่ยวกับผลการตอบของตนเอง ซึ่งผู้เชี่ยวชาญคนอื่น ๆ สามารถแลกเปลี่ยนประสบการณ์กันได้ นอกจากนี้ ผู้วิจัยยังเปิดโอกาสให้ผู้เชี่ยวชาญซักถามประเด็นที่ยังไม่เข้าใจหรือมีข้อสงสัย ทั้งนี้เพื่อปรับความเข้าใจของ

ผู้เชี่ยวชาญทุกคนให้ตรงกัน ก่อนดำเนินการกำหนดคะแนนจุดตัดในรอบที่ 2 ทั้งนี้ การดำเนินงานในรอบที่ 1 เสร็จสิ้นภายใน 1 วัน

รอบที่ 2

การดำเนินงานในรอบนี้ มีขั้นตอนเช่นเดียวกับการกำหนดคะแนนจุดตัดในรอบที่ 1 โดยผู้เชี่ยวชาญทำการประเมินข้อสอบ จำนวน 100 ข้อ ครอบคลุม 3 ทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน พร้อมกำหนดคะแนนจุดตัดในระดับ B1 และ B2 อย่างไรก็ตามในรอบที่ 2 นี้ ผู้วิจัยได้ให้สารสนเทศประกอบการตัดสินใจเพิ่มเติม ได้แก่

1. ผลการตอบของตนเอง และผลการตอบในภาพรวมของกลุ่ม ในรอบที่ 1 ระดับ B1 และ B2 จำแนกตามทักษะการฟัง โครงสร้างภาษา และการอ่าน

2. ค่าความยากง่ายของข้อสอบที่วิเคราะห์ด้วยแบบราสช์ เทสต์เลท (Rasch testlet model) (ดังตารางที่ 17 ผลการประมาณค่าพารามิเตอร์ของข้อสอบและเทสต์เลท)

การกำหนดคะแนนจุดตัด รอบที่ 2 ผู้เชี่ยวชาญสามารถปรับเปลี่ยนผลการตอบข้อสอบแต่ละข้อใหม่ โดยสามารถปรับลดหรือเพิ่มเปอร์เซ็นต์โอกาสการทำข้อสอบถูกแต่ละข้อได้

ระดับ B1 ให้ผู้เชี่ยวชาญอ่านและทำความเข้าใจเนื้อหา และระดับความยากของข้อสอบแต่ละข้อ โดยอิงจากประสบการณ์การสอนภาษาอังกฤษของคน และพิจารณาข้อมูลสารสนเทศประกอบการตัดสินใจเพิ่มเติม จากนั้นจินตนาการว่า “ผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถต่ำสุดในระดับ B1 จะทำข้อสอบข้อที่ i ถูกกี่เปอร์เซ็นต์” ทำเช่นนี้ตั้งแต่ข้อ 1-100 พร้อมบันทึกค่าที่ประมาณได้ลงในช่องโอกาสการทำข้อสอบถูกในแบบประมาณโอกาสการทำข้อสอบถูกสำหรับผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถระดับต่ำสุดในระดับ B1

ระดับ B2 ให้ผู้เชี่ยวชาญอ่านและทำความเข้าใจเนื้อหา และระดับความยากของข้อสอบแต่ละข้อ โดยอิงจากประสบการณ์การสอนภาษาอังกฤษของคน และพิจารณาข้อมูลสารสนเทศประกอบการตัดสินใจเพิ่มเติม จากนั้นจินตนาการว่า “ผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถต่ำสุดในระดับ B2 จะทำข้อสอบข้อที่ i ถูกกี่เปอร์เซ็นต์” ทำเช่นนี้ตั้งแต่ข้อ 1-100 พร้อมบันทึกค่าที่ประมาณได้ลงในช่องโอกาสการทำข้อสอบถูกในแบบประมาณโอกาสการทำข้อสอบถูกสำหรับผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถระดับต่ำสุดในระดับ B2

การรวบรวมและการวิเคราะห์ข้อมูล ผู้วิจัยรวบรวมผลการตอบของผู้เชี่ยวชาญทุกคนที่ได้จากการดำเนินงานในรอบที่ 2 มาคำนวณค่าสถิติ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด และค่าความคลาดเคลื่อนมาตรฐานของการตัดสินใจ ทั้งนี้ การดำเนินงานในรอบที่ 2 เสร็จสิ้นภายใน 1 วัน

รอบที่ 3

ผู้วิจัยให้ผู้เชี่ยวชาญพิจารณาความเหมาะสมของคะแนนจุดตัดระดับ B1 (อยู่ระหว่างระดับ A2 กับ B1) และจุดตัดระดับ B2 (อยู่ระหว่างระดับ B1 กับ B2) และเปิดโอกาสให้ผู้เชี่ยวชาญได้แลกเปลี่ยนความคิดเห็นถึงความเหมาะสมของคะแนนจุดตัดแต่ละช่วงว่าแต่ละระดับมีความเป็นไปได้และเหมาะสมหรือไม่ และให้ผู้เชี่ยวชาญได้มีโอกาสปรับค่าคะแนนจุดตัด ในรอบที่ 3

จากนั้นผู้วิจัยรวบรวมผลการตอบของผู้เชี่ยวชาญทุกคนที่ได้จากการดำเนินงานในรอบที่ 3 มาคำนวณค่าสถิติ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด และค่าความคลาดเคลื่อนมาตรฐานของการตัดสินใจ ซึ่งค่าสถิติของความคิดเห็นของผู้เชี่ยวชาญ ในรอบที่ 3 นี้ได้นำไปใช้เป็นคะแนนจุดตัด ทั้งนี้ การดำเนินงานในรอบที่ 3 เสร็จสิ้นภายใน 1 วัน

2.1.4 การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูล ผู้วิจัยนำผลการตอบของผู้เชี่ยวชาญทุกคนที่ได้ดำเนินงานในแต่ละรอบมาคำนวณค่าสถิติ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด และค่าความคลาดเคลื่อนมาตรฐานของการตัดสินใจ จำแนกตามทักษะการฟัง โครงสร้างภาษา และการอ่าน โดยใช้โปรแกรม Excel

หลังจากได้คะแนนจุดตัดในแต่ละทักษะแล้ว ผู้วิจัยดำเนินการหาคะแนนจุดตัดรวมโดยการคำนวณคะแนนรวมแบบถ่วงน้ำหนัก (weighted overall cut score) เพื่อกำหนดคะแนนจุดตัด รวมจากแต่ละทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน (Cizek & Bunch, 2007) โดยมีขั้นตอนการคำนวณดังนี้

1. คำนวณคะแนนถ่วงน้ำหนักของแต่ละทักษะ

คะแนนถ่วงน้ำหนักของแต่ละทักษะ คำนวณโดยใช้สูตรดังนี้

$$\text{Weighted Score} = \frac{(\text{Weight}_{\text{Skill}} \times \text{Cut Score}_{\text{Skill}})}{\text{Max Score}_{\text{Skill}}}$$

เมื่อกำหนดให้

$\text{Weight}_{\text{Skill}}$ หมายถึง สัดส่วนจำนวนข้อสอบของทักษะนั้น ๆ จากจำนวนข้อทั้งหมดในแบบทดสอบ

$\text{Cut Score}_{\text{Skill}}$ หมายถึง คะแนนจุดตัดของทักษะนั้น ๆ ที่ได้มาจากการใช้วิธีการของแองกอฟที่ได้รับการปรับปรุง

$\text{Max Score}_{\text{Skill}}$ หมายถึง คะแนนเต็มของทักษะนั้น ๆ

2. รวมคะแนนถ่วงน้ำหนัก

คะแนนรวมถ่วงน้ำหนักคำนวณโดยการรวมคะแนนถ่วงน้ำหนักของแต่ละทักษะ

ดังนี้

$$\text{Overall Cut Score} = \text{Weighted Listening} + \text{Weighted Structure} + \text{Weighted Reading}$$

3. แปลงคะแนนรวม (Standardized Score) เพื่อให้ผลลัพธ์เข้าใจง่าย

คะแนนรวมถูกแปลงให้อยู่ในช่วง 100 คะแนน โดยใช้สูตรดังนี้

$$\text{Transformed Overall Score} = \text{Overall Cut Score} \times 100$$

ตอนที่ 2.2 การประเมินกระบวนการกำหนดคะแนนจุดตัด

หลังจากทำการกำหนดคะแนนจุดตัดเสร็จแล้ว ผู้วิจัยให้ผู้เชี่ยวชาญทุกคนที่อยู่ในกระบวนการกำหนดคะแนนจุดตัดประเมินขั้นตอนการทำงานที่ได้ดำเนินการไป และประเมินตนเองว่าเข้าใจกระบวนการกำหนดคะแนนจุดตัดหรือไม่

การดำเนินงานในตอนนี้ แบ่งออกเป็น 4 ส่วน ได้แก่ ผู้ให้ข้อมูลหลัก เครื่องมือที่ใช้ในการวิจัย การเก็บรวบรวมข้อมูล และการวิเคราะห์ข้อมูล รายละเอียดแต่ละส่วนมีดังนี้

2.2.1 ผู้ให้ข้อมูลหลัก

ภายหลังจากสิ้นสุดกระบวนการกำหนดคะแนนจุดตัด ในทั้ง 3 รอบเรียบร้อยแล้ว ผู้วิจัยได้ดำเนินการให้ผู้ให้ข้อมูลหลัก ซึ่งเป็นผู้เชี่ยวชาญ จำนวน 10 คน ที่มีส่วนร่วมในกระบวนการกำหนดคะแนนจุดตัดทั้งหมด ประเมินตนเองและสะท้อนความคิดเห็นต่อกระบวนการที่ได้เข้าร่วม โดยมุ่งให้ผู้เชี่ยวชาญแสดงความคิดเห็นต่อประสบการณ์ที่ได้รับ รวมถึงสถานการณ์ต่าง ๆ ที่สะท้อนถึงความเข้าใจ ความชัดเจน ความเหมาะสมของขั้นตอน และความเชื่อมั่นที่มีต่อผลลัพธ์ของการกำหนดคะแนนจุดตัด

2.2.2 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัย เป็นแบบประเมินกระบวนการกำหนดคะแนนจุดตัด แบ่งออกเป็น 2 ตอน ตอนที่ 1 ข้อมูลทั่วไปของผู้เชี่ยวชาญกำหนดคะแนนจุดตัด ประกอบด้วย คำถาม 3 ข้อ เป็นแบบตรวจสอบรายการ (check list) ได้แก่ ระดับการศึกษาสูงสุด ประสบการณ์ด้านการสอนภาษาอังกฤษระดับอุดมศึกษา และประสบการณ์ด้านการจัดสอบหรือการสอนภาษาอังกฤษตามกรอบอ้างอิง CEFR และตอนที่ 2 ความคิดเห็นเกี่ยวกับกระบวนการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญ ประกอบด้วย คำถาม 12 ข้อ เป็นแบบมาตราประมาณค่า 4 ระดับ ได้แก่ 4 หมายถึง เห็นด้วยมากที่สุด 3 หมายถึง เห็นด้วยมาก 2 หมายถึง เห็นด้วยน้อย และ 1 หมายถึง เห็นด้วยน้อยที่สุด (รายละเอียดดังภาคผนวก ฉ)

ผู้วิจัยนำแบบประเมินกระบวนการกำหนดคะแนนจุดตัดไปให้ผู้เชี่ยวชาญ จำนวน 3 ท่าน ซึ่งเป็นผู้มีประสบการณ์ด้านการวัดและประเมินผล ตั้งแต่ 3 ปีขึ้นไป ตรวจสอบความตรงเชิงเนื้อหา ความครอบคลุมโครงสร้าง ความเหมาะสมของข้อคำถาม ความเป็นปรนัย ความชัดเจนของการใช้ ภาษา และตรวจสอบข้อคำถามแต่ละข้อว่าวัดได้ตรงตามจุดมุ่งหมายที่ต้องการวัดหรือไม่ โดยคำนวณ ค่า IOC มีเกณฑ์การพิจารณาค่าดัชนี IOC ต้องมากกว่าหรือเท่ากับ .50 ขึ้นไป จึงถือว่าข้อคำถามนั้น สอดคล้องกับวัตถุประสงค์ของการวัด (Rovinelli & Hambleton, 1977)

ผลการตรวจสอบความตรงเชิงเนื้อหา พบว่า แบบประเมินกระบวนการกำหนดคะแนน จุดตัด มีความตรงเชิงเนื้อหา ครอบคลุมตามเนื้อหาสาระ โดยมีค่า IOC เฉลี่ยรวมทั้งฉบับ เท่ากับ .95 สำหรับคุณภาพรายข้อ ค่า IOC อยู่ระหว่าง .67 ถึง 1 ซึ่งทุกข้อมีค่าเกินเกณฑ์ที่กำหนดไว้ รวมทั้ง ผู้วิจัยได้นำข้อเสนอแนะของผู้เชี่ยวชาญมาปรับปรุงข้อคำถามให้ชัดเจนมากขึ้น

2.2.3 การเก็บรวบรวมข้อมูล

ภายหลังจากผู้เชี่ยวชาญทั้ง 10 คนเสร็จสิ้นการกำหนดคะแนนจุดตัดในรอบที่ 3 ผู้วิจัย ได้ดำเนินการเก็บรวบรวมข้อมูลทันที โดยเชิญผู้เชี่ยวชาญประเมินกระบวนการดำเนินงานทั้งหมด และประเมินตนเองเกี่ยวกับความเข้าใจในกระบวนการกำหนดคะแนนจุดตัด ผ่านแบบประเมิน ออนไลน์ที่จัดทำในระบบ Google Forms เพื่อความสะดวกในการตอบแบบประเมิน และลดความ คลาดเคลื่อนจากความล่าช้าในการให้ข้อมูล

2.2.4 การวิเคราะห์ข้อมูล

ผู้วิจัยนำผลการตอบของผู้เชี่ยวชาญทั้ง 10 คน มาคำนวณค่าสถิติ ได้แก่ ร้อยละ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน โดยใช้โปรแกรม R

หลักเกณฑ์การแปลความหมายของค่าคะแนน ผู้วิจัยใช้เกณฑ์สัมบูรณ์ (absolute criteria) แบบการใช้ขอบเขตที่แท้จริง (exact limits) ได้แก่ ค่าที่อยู่ระหว่างขอบเขตต่ำ และขอบเขต สูง ดังนี้

คะแนนเฉลี่ย	ระดับความคิดเห็น
3.51-4.00	มากที่สุด
2.51-3.50	มาก
1.51-2.50	น้อย
1.00-1.50	น้อยที่สุด

3.2 การพิทักษ์สิทธิ์ของตัวอย่างหรืออาสาสมัครที่ใช้ในการวิจัย

การพิทักษ์สิทธิ์ของตัวอย่างหรืออาสาสมัครที่ใช้ในการวิจัยนี้ ผู้วิจัยเริ่มดำเนินการรวบรวมข้อมูลเพื่อการเทียบเคียงคะแนนแบบทดสอบ STOU-EPT กับกรอบ CEFR จำนวน 3 กลุ่ม ได้แก่ ตัวอย่างหรืออาสาสมัครที่ประเมินความสอดคล้อง ตัวอย่างหรืออาสาสมัครที่กำหนดคะแนนจุดตัด และตัวอย่างหรืออาสาสมัครที่เป็นผู้สอบ STOU-EPT หลังจากได้รับความเห็นชอบจากคณะกรรมการจริยธรรมการวิจัยในคนมหาวิทยาลัยสุโขทัยธรรมาธิราช เมื่อวันที่ 10 พฤษภาคม พ.ศ. 2566 และหมดอายุใบรับรองวันที่ 9 พฤษภาคม พ.ศ. 2567 และวันทำการรับรองการต่ออายุวันที่ 10 พฤษภาคม พ.ศ. 2567 และหมดอายุใบรับรองการต่ออายุ วันที่ 9 พฤษภาคม พ.ศ. 2568 รหัสโครงการวิจัย STOU-IRB 2566/011.2501 (รายละเอียดดังภาคผนวก ก) ก่อนเริ่มดำเนินการเก็บรวบรวมข้อมูลกับตัวอย่างหรืออาสาสมัครแต่ละกลุ่ม ผู้วิจัยได้ทำการชี้แจงข้อมูลสำคัญแก่ตัวอย่างหรืออาสาสมัคร โดยให้สิทธิ์ในการตัดสินใจด้วยตนเอง พร้อมทั้งเปิดโอกาสให้ตัวอย่างหรืออาสาสมัครซักถามข้อข้องใจในการวิจัยอย่างเต็มที่ในด้านความพร้อมในการให้ข้อมูล ตัวอย่างหรืออาสาสมัครสามารถถอนตัวหรือถอนตัวจากการประเมินความสอดคล้อง และกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดแบบทดสอบ STOU-EPT ได้ตลอดเวลา เมื่อตัวอย่างหรืออาสาสมัครยินดียินยอมเข้าร่วมการวิจัย ผู้วิจัยให้ตัวอย่างหรืออาสาสมัครลงลายมือชื่อเป็นลายลักษณ์อักษรในหนังสือแสดงความยินยอมเข้าร่วมโครงการวิจัย ทั้งนี้ ผู้วิจัยจะรักษาความลับของผู้ให้ข้อมูล และตัวอย่างหรืออาสาสมัคร โดยการหลีกเลี่ยงการใช้นามจริงแต่จะใช้รหัสในการจัดเก็บข้อมูล