

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยเรื่อง การเทียบเคียงคะแนนแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรป มีวัตถุประสงค์ 2 ข้อ ได้แก่ 1) เพื่อศึกษาความสอดคล้องระหว่างแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีของเว็บ และ 2) เพื่อเทียบเคียงคะแนนแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง ผู้ให้ข้อมูลหลัก ประกอบด้วย ผู้เชี่ยวชาญประเมินระดับความลึกของความรู้ จำนวน 5 คน ผู้เชี่ยวชาญประเมินความสอดคล้องด้านเนื้อหา จำนวน 10 คน และผู้เชี่ยวชาญกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดแบบทดสอบ STOU-EPT จำนวน 10 คน โดยใช้วิธีการเลือกแบบเจาะจง และตัวอย่างที่ใช้เป็นผู้ที่ลงทะเบียนสอบด้วยความสมัครใจ ระหว่างเดือนพฤษภาคมถึงกันยายน พ.ศ. 2567 จำนวน 1,272 คน

เครื่องมือที่ใช้ในการวิจัย มี 6 ชุด ได้แก่ ชุดที่ 1 แบบทดสอบ STOU-EPT เป็นแบบทดสอบที่เลือกมา 1 แบบ จากทั้งหมด 4 แบบที่พัฒนาขึ้นใหม่ในปี พ.ศ. 2566 โดยมีลักษณะเป็นแบบทดสอบคู่ขนาน (parallel test) ครอบคลุม 3 ทักษะ ได้แก่ การฟัง (25 ข้อ) โครงสร้างภาษา (35 ข้อ) และการอ่าน (40 ข้อ) รวมทั้งสิ้น 100 ข้อ ข้อสอบแต่ละทักษะออกแบบมาเพื่อวัดความสามารถทางภาษาในระดับ A2 B1 และ B2 ตามกรอบ CEFR พัฒนาโดยคณาจารย์แขนงวิชาภาษาอังกฤษ สาขาวิชาศิลปศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช ชุดที่ 2 แบบประเมินความสอดคล้องระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละระดับความสามารถกับระดับความลึกของความรู้ตามเกณฑ์ของเว็บ ผลประเมินระดับความลึกของความรู้ระหว่างตัวชี้วัดตามกรอบ CEFR แต่ละทักษะกับระดับความลึกของความรู้ตามเกณฑ์ของเว็บ โดยใช้ค่านิยาม พบว่า การฟัง ระดับ A2 และ B1 วัดความลึกของความรู้ด้านการมีทักษะและมโนทัศน์ และระดับ B2 วัดความลึกของความรู้ด้านการคิดเชิงกลยุทธ์ โครงสร้างภาษาทั้ง 3 ระดับ ได้แก่ A2 B1 และ B2 วัดความลึกของความรู้ด้านการมีทักษะและมโนทัศน์ และการอ่าน ระดับ A2 และ B1 วัดความลึกของความรู้ด้านการมีทักษะและมโนทัศน์ และระดับ B2 วัดความลึกของความรู้ด้านการคิดเชิงกลยุทธ์ เมื่อตรวจสอบความเที่ยงระหว่างผู้ประเมินทั้ง 5 คน พบว่า การฟัง มีค่า $p = .911$ และการอ่าน มีค่า $p = .911$ แสดงว่า การฟัง และการอ่าน ผู้ประเมินประเมินสอดคล้องกันในระดับสูงมาก ส่วนโครงสร้างภาษา มีค่า $p = .540$ แสดงว่า ผู้ประเมินประเมิน

สอดคล้องกันในระดับปานกลาง **ชุดที่ 3** แบบประเมินความสอดคล้องระหว่างข้อสอบ ตัวชี้วัดตามกรอบ CEFR และระดับความลึกของความรู้ด้วยวิธีของเวบบ์ และคำอธิบายตัวชี้วัดสำหรับประเมินความสอดคล้อง ผลการตรวจสอบความตรงเชิงเนื้อหา โดยคำนวณค่า IOC พบว่า ทุกตัวชี้วัดมีค่า IOC อยู่ระหว่าง .67 ถึง 1 แสดงให้เห็นว่า ทุกตัวชี้วัดมีค่า IOC มากกว่าเกณฑ์ที่กำหนด ดังนั้น ตัวชี้วัดทุกตัวสามารถนำไปใช้ในการประเมินความสอดคล้องได้ และหลังจากตรวจสอบความตรงเชิงเนื้อหาของตัวชี้วัดสำหรับประเมินความสอดคล้องในแต่ละทักษะโดยผู้เชี่ยวชาญแล้ว ผู้วิจัยได้นำตัวชี้วัดที่ผ่านการปรับปรุงมาให้ผู้เชี่ยวชาญ จำนวน 3 คน ทดลองประเมินความสอดคล้องของเนื้อหา พบว่า การฟัง มีค่า $p = .644$ และโครงสร้างภาษา มีค่า $p = .655$ แสดงว่า ความเที่ยงระหว่างผู้ประเมินอยู่ในระดับดี ส่วนการอ่าน มีค่า $p = .479$ แสดงว่า ผู้ประเมินประเมินสอดคล้องกันในระดับปานกลาง ผลดังกล่าวสอดคล้องกับเกณฑ์ที่ Evans (1996) กำหนด และทดลองประเมินความสอดคล้องด้านความลึกของความรู้ พบว่า การฟัง มีค่า $p = .882$ แสดงว่า ผู้ประเมินประเมินสอดคล้องกันในระดับสูงมาก โครงสร้างภาษา มีค่า $p = .554$ แสดงว่า ผู้ประเมินประเมินสอดคล้องกันในระดับปานกลาง และการอ่าน มีค่า $p = .200$ แสดงว่า ผู้ประเมินประเมินสอดคล้องกันในระดับต่ำ ผลดังกล่าวสอดคล้องกับเกณฑ์ที่ Evans (1996) กำหนด **ชุดที่ 4** คู่มือการกำหนดคะแนนจุดตัดด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง **ชุดที่ 5** แบบประเมินโอกาสการทำข้อสอบถูกสำหรับผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถระดับต่ำสุด ในระดับ B1 และระดับ B2 และ**ชุดที่ 6** แบบประเมินกระบวนการกำหนดคะแนนจุดตัด

การวิเคราะห์ข้อมูล ผู้วิจัยดำเนินการวิเคราะห์ข้อมูลโดยใช้โปรแกรม R และโปรแกรม Excel ดังนี้ (1) การวิเคราะห์ความสอดคล้องของเนื้อหา ความลึกของความรู้ และความสอดคล้องด้านความสมดุล ผู้วิจัยวิเคราะห์ความสอดคล้องของเนื้อหาโดยคำนวณค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน การวิเคราะห์ด้านความลึกของความรู้คำนวณร้อยละ และค่าเฉลี่ย การตรวจสอบความเที่ยงระหว่างผู้ประเมินโดยใช้การวิเคราะห์ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์แบบลำดับของสเปียร์แมนระหว่างผู้ประเมินแต่ละคู่ (ρ) และความสอดคล้องด้านความสมดุล ใช้การคำนวณดัชนีความสมดุล (Q) (2) การวิเคราะห์คุณภาพข้อสอบ STOU-EPT ตามทฤษฎีการตอบสนองข้อสอบด้วย Rasch testlet model (3) การกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดแบบทดสอบ STOU-EPT ในแต่ละรอบโดยคำนวณค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด และค่าความคลาดเคลื่อนมาตรฐานของการตัดสิน (4) การหาคะแนนจุดตัดในภาพรวม โดยคำนวณคะแนนรวมแบบถ่วงน้ำหนักตามวิธีการของ Cizek and Bunch (2007) และ (5) การประเมินกระบวนการกำหนดคะแนนจุดตัด โดยคำนวณร้อยละ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

5.1 สรุปผลการวิจัย

การนำเสนอผลการวิจัย แบ่งออกเป็น 2 ตอน ได้แก่ ผลการศึกษาความสอดคล้องระหว่างแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีของเว็บ และผลการเทียบเคียงคะแนนแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีการของแองกอฟที่ได้รับการปรับปรุง

ตอนที่ 1 ผลการศึกษาความสอดคล้องระหว่างแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีของเว็บ พิจารณาจาก 3 เกณฑ์ ได้แก่ ความสอดคล้องของเนื้อหา ความสอดคล้องด้านความลึกของความรู้ และความสอดคล้องด้านความสมดุล สรุปได้ดังนี้

1.1 ความสอดคล้องของเนื้อหา พบว่า การฟัง มีข้อสอบ STOU-EPT ที่สอดคล้องกับมาตรฐานกรอบ CEFR ในระดับ A2 จำนวน 8 ข้อ ระดับ B1 จำนวน 8 ข้อ และระดับ B2 จำนวน 9 ข้อ โครงสร้างภาษา มีข้อสอบ STOU-EPT ที่สอดคล้องกับมาตรฐานกรอบ CEFR ในระดับ A2 จำนวน 12 ข้อ ระดับ B1 จำนวน 12 ข้อ และระดับ B2 จำนวน 10 ข้อ และการอ่าน มีข้อสอบ STOU-EPT ที่สอดคล้องกับมาตรฐานกรอบ CEFR ในระดับ A2 จำนวน 6 ข้อ ระดับ B1 จำนวน 19 ข้อ และระดับ B2 จำนวน 15 ข้อ

ข้อสอบ STOU-EPT ทั้งสามทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน สอดคล้องกับมาตรฐานกรอบ CEFR ในทุกระดับความสามารถ ซึ่งผ่านเกณฑ์ความสอดคล้องของเนื้อหาตามวิธีของ Webb (1997) กำหนดว่า ข้อสอบควรมีอย่างน้อย 6 ข้อ สำหรับวัดความรู้ในแต่ละมาตรฐาน นอกจากนั้น ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมินทั้ง 10 คน พบว่า การฟัง ผู้ประเมินประเมินสอดคล้องกันในระดับสูงมาก ($p = .902$) ส่วนโครงสร้างภาษา และการอ่าน ผู้ประเมินประเมินสอดคล้องกันในระดับปานกลาง ($p = .588$ และ $.432$ ตามลำดับ) ซึ่งสอดคล้องกับเกณฑ์ของ Evans (1996)

1.2 ความสอดคล้องด้านความลึกของความรู้ พบว่า การฟัง ข้อสอบ STOU-EPT ระดับ B1 (% hit = 100) และระดับ B2 (% hit = 100) สอดคล้องกับความลึกของความรู้ ส่วนระดับ A2 ไม่ผ่านเกณฑ์ (% hit = 0, ต่ำกว่า 50%) โครงสร้างภาษา ข้อสอบ STOU-EPT ระดับ B1 (% hit = 100) และระดับ B2 (% hit = 92.31) สอดคล้องกับความลึกของความรู้ ส่วนระดับ A2 ไม่ผ่านเกณฑ์ (% hit = 0, ต่ำกว่า 50%) และการอ่าน ข้อสอบ STOU-EPT ทั้ง 3 ระดับ ได้แก่ ระดับ A2 (% hit = 100) ระดับ B2 (% hit = 100) และระดับ B2 (% above = 100) สอดคล้องกับความลึกของความรู้

ข้อสอบ STOU-EPT การอ่าน สอดคล้องกับความลึกของความรู้ในทุกระดับความสามารถ (A2 B1 และ B2) ส่วนการฟัง และโครงสร้างภาษา สอดคล้องกับความลึกของความรู้ในระดับความสามารถ B1 และ B2 ส่วน A2 ไม่ผ่านเกณฑ์ตามวิธีของ Webb (1997) กำหนดว่า มาตรฐานและเครื่องมือประเมินที่ถือว่ามีความลึกของความรู้เมื่อข้อสอบอย่างน้อย 50% ของข้อสอบทั้งหมดในมาตรฐานนั้น ๆ มีความลึกเท่ากับหรือมากกว่าความลึกที่มีในตัวชี้วัด นอกจากนั้น ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมินทั้ง 10 คน พบว่า การฟัง และโครงสร้างภาษา ผู้ประเมินประเมินสอดคล้องกันในระดับปานกลาง ($p = .525$ และ $.442$ ตามลำดับ) ส่วนการอ่าน ผู้ประเมินประเมินสอดคล้องกันในระดับต่ำ ($p = .249$) ซึ่งสอดคล้องกับเกณฑ์ของ Evans (1996)

1.3 ความสอดคล้องด้านความสมดุล พบว่า ข้อสอบ STOU-EPT ทั้งสามทักษะ ได้แก่ การฟัง ($Q = .853$) โครงสร้างภาษา ($Q = .952$) และการอ่าน ($Q = .817$) มีการกระจายในแต่ละระดับความสามารถตามกรอบ CEFR ได้อย่างสมดุล สอดคล้องกับเกณฑ์ของ Webb (2007) ระบุว่า ดัชนีความสมดุล (Q) ควรมีค่าตั้งแต่ $.70$ ขึ้นไป จึงถือว่ายอมรับได้

ตอนที่ 2 ผลการเทียบเคียงคะแนนแบบทดสอบวัดระดับความรู้ ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง แบ่งออกเป็น 2 ส่วน ได้แก่ ผลการเทียบเคียงคะแนนแบบทดสอบ STOU-EPT กับกรอบ CEFR ด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง ในรอบที่ 1 ถึงรอบที่ 3 และผลการประเมินกระบวนการกำหนดคะแนนจุดตัด แต่ละส่วนสรุปได้ดังนี้

2.1 ผลการเทียบเคียงคะแนนแบบทดสอบ STOU-EPT กับกรอบ CEFR ด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง ในรอบที่ 1 ถึงรอบที่ 3 แสดงผลใน 2 รูปแบบ ได้แก่ คะแนนจุดตัดแยกตามทักษะ และคะแนนจุดตัดรวม

คะแนนจุดตัดแยกตามทักษะ

ทักษะการฟัง พบว่า คะแนนจุดตัดสำหรับความสามารถระดับ B1 ในรอบที่ 3 มีค่าเท่ากับ 13.49 คะแนน หรือประมาณ 13 คะแนน ซึ่งมีค่าลดลงเมื่อเทียบกับรอบที่ 1 (15.75 คะแนน) และรอบที่ 2 (15.91 คะแนน) ในขณะที่คะแนนจุดตัดสำหรับความสามารถระดับ B2 ในรอบที่ 3 มีค่าเท่ากับ 19.24 หรือประมาณ 19 คะแนน ซึ่งมีค่าไม่แตกต่างกันเมื่อเทียบกับรอบที่ 1 (19.15 คะแนน) และรอบที่ 2 (19.30 คะแนน) ทั้งนี้ คะแนนจุดตัดสำหรับความสามารถระดับ B2 สูงกว่าระดับ B1 อยู่ 6 คะแนน และส่วนเบี่ยงเบนมาตรฐาน (SD) และค่าความคลาดเคลื่อนมาตรฐานของการตัดสินใจ (SEJ) ทั้งระดับ B1 และ B2 ในรอบที่ 3 มีค่าลดเมื่อเทียบกับรอบที่ 1 และรอบที่ 2 แสดงให้เห็นถึง ความสอดคล้องในการตัดสินใจของผู้เชี่ยวชาญ และเพิ่มความน่าเชื่อถือของคะแนนจุดตัดในรอบสุดท้าย

ทักษะโครงสร้างภาษา พบว่า คะแนนจุดตัดสำหรับความสามารถระดับ B1 ในรอบที่ 3 มีค่าเท่ากับ 20.48 คะแนน หรือประมาณ 20 คะแนน ซึ่งมีค่าไม่แตกต่างกันเมื่อเทียบกับรอบที่ 1 (19.86 คะแนน) แต่มีค่าลดลงเล็กน้อยเมื่อเทียบกับรอบที่ 2 (22.12 คะแนน) ในขณะที่คะแนนจุดตัดสำหรับความสามารถระดับ B2 ในรอบที่ 3 มีค่าเท่ากับ 27.23 หรือประมาณ 27 คะแนน ซึ่งมีค่าไม่แตกต่างกันเมื่อเทียบกับรอบที่ 2 (27.30 คะแนน) แต่มีค่าเพิ่มขึ้นเล็กน้อยเมื่อเทียบกับรอบที่ 1 (25.57 คะแนน) ทั้งนี้ คะแนนจุดตัดสำหรับความสามารถระดับ B2 สูงกว่าระดับ B1 อยู่ 7 คะแนน และส่วนเบี่ยงเบนมาตรฐาน (SD) และค่าความคลาดเคลื่อนมาตรฐานของการตัดสิน (SEJ) ทั้งระดับ B1 และ B2 ในรอบที่ 3 มีค่าลดเมื่อเทียบกับรอบที่ 1 และรอบที่ 2 แสดงให้เห็นถึง ความสอดคล้องในการตัดสินของผู้เชี่ยวชาญ และเพิ่มความน่าเชื่อถือของคะแนนจุดตัดในรอบสุดท้าย

ทักษะการอ่าน พบว่า คะแนนจุดตัดสำหรับความสามารถระดับ B1 ในรอบที่ 3 มีค่าเท่ากับ 23.26 คะแนน หรือประมาณ 23 คะแนน ซึ่งมีค่าลดลงเมื่อเทียบกับรอบที่ 1 (24.53 คะแนน) และรอบที่ 2 (26.27 คะแนน) ในขณะที่คะแนนจุดตัดสำหรับความสามารถระดับ B2 ในรอบที่ 3 มีค่าเท่ากับ 31.43 หรือประมาณ 31 คะแนน ซึ่งมีค่าเพิ่มขึ้นเล็กน้อยเมื่อเทียบกับรอบที่ 1 (30.28 คะแนน) แต่มีค่าน้อยลงเมื่อเทียบกับรอบที่ 2 (32.02 คะแนน) ทั้งนี้ คะแนนจุดตัดสำหรับความสามารถระดับ B2 สูงกว่าระดับ B1 อยู่ 8 คะแนน และส่วนเบี่ยงเบนมาตรฐาน (SD) และค่าความคลาดเคลื่อนมาตรฐานของการตัดสิน (SEJ) ทั้งระดับ B1 และ B2 ในรอบที่ 3 มีค่าลดเมื่อเทียบกับรอบที่ 1 และรอบที่ 2 แสดงให้เห็นถึง ความสอดคล้องในการตัดสินของผู้เชี่ยวชาญ และเพิ่มความน่าเชื่อถือของคะแนนจุดตัดในรอบสุดท้าย

คะแนนจุดตัดรวม

คะแนนจุดตัดรวมของแบบทดสอบ STOU-EPT พบว่า คะแนนจุดตัดสำหรับความสามารถระดับ B1 มีค่าเท่ากับ 56 คะแนน และคะแนนจุดตัดสำหรับความสามารถระดับ B2 มีค่าเท่ากับ 77 คะแนน

ผลการนำคะแนนจุดตัดมาแบ่งช่วงในแต่ละระดับความสามารถ แสดงผลใน 2 รูปแบบ ได้แก่ แบบที่ 1 คะแนนจุดตัดแยกตามทักษะ พบว่า ทักษะการฟัง ระดับ A2 มีคะแนนอยู่ในช่วง 1-12 คะแนน ระดับ B1 อยู่ในช่วง 13-18 คะแนน และระดับ B2 อยู่ในช่วง 19-25 คะแนน ทักษะโครงสร้างภาษา ระดับ A2 มีคะแนนอยู่ในช่วง 1-19 คะแนน ระดับ B1 อยู่ในช่วง 20-26 คะแนน และระดับ B2 อยู่ในช่วง 27-35 คะแนน และทักษะการอ่าน ระดับ A2 มีคะแนนอยู่ในช่วง 1-22 คะแนน ระดับ B1 อยู่ในช่วง 23-30 คะแนน และระดับ B2 อยู่ในช่วง 31-40 คะแนน และแบบที่ 2 คะแนนจุดตัดรวม พบว่า ระดับ A2 มีคะแนนอยู่ในช่วง 1-55 คะแนน ระดับ B1 อยู่ในช่วง 56-76 คะแนน และระดับ B2 อยู่ในช่วง 77-100 คะแนน

ผลการวิเคราะห์ความแม่นยำของการจำแนก (CA) ของแบบทดสอบ STOU-EPT พบว่า คะแนนจุดตัดที่กำหนดสามารถจำแนกระดับความสามารถของผู้สอบได้อย่างถูกต้องในระดับสูง โดยมีความแม่นยำในทักษะการฟัง ร้อยละ 97.20 โครงสร้างภาษา ร้อยละ 95.60 และการอ่าน ร้อยละ 93.70 ผลการวิจัยนี้เป็นหลักฐานด้านความตรงที่สนับสนุนความสอดคล้องของการตัดสินของผู้เชี่ยวชาญ และยืนยันความน่าเชื่อถือของการใช้คะแนนจุดตัดดังกล่าวในการจำแนกระดับความสามารถของผู้สอบ

2.2 ผลการประเมินกระบวนการกำหนดคะแนนจุดตัด

ผู้เชี่ยวชาญที่เข้าร่วมกำหนดคะแนนจุดตัด จำนวน 10 คน พบว่า ส่วนใหญ่มีคุณวุฒิระดับปริญญาเอก จำนวน 9 คน (ร้อยละ 90) โดยสำเร็จการศึกษาในสาขาวิชา Ph.D. (Applied Linguistics) และ Ph.D. (English) มากที่สุด มีประสบการณ์ด้านการสอนภาษาอังกฤษระดับอุดมศึกษา 21-25 ปี มากที่สุด จำนวน 5 คน (ร้อยละ 50) รองลงมา 26 ปีขึ้นไป จำนวน 3 คน (ร้อยละ 30) และมีประสบการณ์ด้านการจัดสอบหรือการสอนภาษาอังกฤษตามกรอบ CEFR มากกว่า 10 ปีขึ้นไป จำนวน 6 คน (ร้อยละ 60) รองลงมา 5-10 ปี จำนวน 2 คน (ร้อยละ 20) และน้อยกว่า 5 ปี จำนวน 2 คน (ร้อยละ 20)

ผลการประเมินกระบวนการกำหนดคะแนนจุดตัดทั้ง 12 ประเด็น พบว่า มีประเด็นที่ผู้เชี่ยวชาญมีความคิดเห็นเกี่ยวกับกระบวนการกำหนดคะแนนจุดตัดด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุงในระดับมากที่สุด (ร้อยละ 60 ขึ้นไป) 5 ประเด็น ได้แก่ (1) แบบประมาณโอกาสการทำข้อสอบถูกในรอบที่ 1 2 และ 3 ง่ายต่อการกรอกข้อมูล ร้อยละ 60 ($M = 3.60, SD = .52$) (2) ฉันทมีโอกาสแสดงความคิดเห็นเกี่ยวกับผลการกำหนดคะแนนจุดตัดแต่ละรอบ ร้อยละ 80 ($M = 3.80, SD = .42$) (3) เวลาที่ให้นั้นแสดงความคิดเห็นเกี่ยวกับผลการกำหนดคะแนนจุดตัดในแต่ละรอบมีเพียงพอ ร้อยละ 70 ($M = 3.70, SD = .48$) (4) การประชุมหลังการกำหนดคะแนนจุดตัดในรอบที่ 1 มีประโยชน์ต่อการกำหนดคะแนนจุดตัด ร้อยละ 80 ($M = 3.80, SD = .42$) และ (5) การประชุมหลังการกำหนดคะแนนจุดตัดในรอบที่ 2 มีประโยชน์ต่อการกำหนดคะแนนจุดตัด ร้อยละ 70 ($M = 3.70, SD = .48$) และประเด็นที่ผู้เชี่ยวชาญมีความคิดเห็นเกี่ยวกับกระบวนการกำหนดคะแนนจุดตัดด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุงในระดับมาก (ร้อยละ 60 ขึ้นไป) 3 ประเด็น ได้แก่ (1) ผู้ดำเนินการอธิบายวิธีการกำหนดคะแนนจุดตัดได้อย่างชัดเจน ร้อยละ 60 ($M = 3.20, SD = .63$) (2) การปฐมนิเทศและการฝึกซ้อมช่วยให้ฉันทเข้าใจวิธีการกำหนดคะแนนจุดตัดมากขึ้น ร้อยละ 60 ($M = 3.50, SD = .71$) และ (3) คู่มือสำหรับการกำหนดคะแนนจุดตัดด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง มีความชัดเจนและเข้าใจง่าย ร้อยละ 60 ($M = 3.40, SD = .84$)

5.2 อภิปรายผลการวิจัย

การวิจัยครั้งนี้ มีประเด็นที่ควรให้ความสำคัญ 3 ประเด็น ได้แก่ ประเด็นที่ 1 การศึกษาความสอดคล้องระหว่างแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีของเว็บ ประเด็นที่ 2 การเทียบเคียงคะแนนแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีการของแองกอฟที่ได้รับการปรับปรุง และประเด็นที่ 3 การกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดด้วยวิธีการของแองกอฟที่ได้รับการปรับปรุง รายละเอียดแต่ละประเด็นมีดังนี้

ประเด็นที่ 1 การศึกษาความสอดคล้องระหว่างแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีของเว็บ มีการอภิปรายใน 2 หัวข้อ ดังนี้

1.1 ข้อมูลพื้นฐานแบบทดสอบ STOU-EPT พบว่า แบบทดสอบ STOU-EPT มีข้อสอบทั้งหมด 100 ข้อ แบ่งออกเป็น 3 ทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน จุดมุ่งหมายของแบบทดสอบนี้ คือ เพื่อประเมินทักษะภาษาอังกฤษของผู้สอบตามกรอบ CEFR ซึ่งเน้นการพัฒนาทักษะทั้งความเข้าใจและการใช้ภาษาในระดับต่าง ๆ เพื่อให้ผู้สอบสามารถใช้ภาษาอังกฤษได้ใกล้เคียงกับการใช้งานจริงในชีวิตประจำวัน (Council of Europe, 2020)

ทักษะการฟัง ข้อสอบ STOU-EPT ในทักษะการฟังได้ออกแบบให้ครอบคลุมสมรรถนะที่ระบุไว้ในกรอบ CEFR ระดับ A2 B1 และ B2 (British Council & EAQUALS, 2015) โดยใช้รูปแบบข้อสอบหลากหลาย ได้แก่ แบบเลือกตอบ แบบจับคู่ แบบเติมคำ และแบบถูก-ผิด ซึ่งครอบคลุมเนื้อหาในบริบทต่าง ๆ ที่ใกล้เคียงกับสถานการณ์จริง ทั้งนี้ รูปแบบข้อสอบที่หลากหลายมีเป้าหมายเพื่อวัดความสามารถด้านการฟังของผู้สอบอย่างรอบด้าน การประเมินทักษะการฟังในลักษณะนี้ช่วยสะท้อนระดับความสามารถในการรับรู้และตอบสนองต่อข้อมูลเสียงในสถานการณ์ที่หลากหลาย ซึ่งเป็นองค์ประกอบสำคัญของการใช้ภาษาในสถานการณ์จริง นอกจากนี้ ผลการประเมินสามารถใช้เป็นข้อมูลประกอบในการวางแผนปรับปรุงและพัฒนาทักษะการฟังของผู้สอบให้สอดคล้องกับระดับความสามารถที่ต้องการ (Council of Europe, 2020)

ทักษะโครงสร้างภาษา ข้อสอบ STOU-EPT ในส่วนของทักษะโครงสร้างภาษาเป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 35 ข้อ ซึ่งออกแบบมาเพื่อประเมินความสามารถของผู้สอบในการเข้าใจและการใช้โครงสร้างทางภาษาอังกฤษได้อย่างถูกต้องและเหมาะสม โดยเฉพาะในระดับ B1 ถึง B2 ตามกรอบ CEFR ซึ่งเป็นระดับที่ให้ความสำคัญกับการใช้โครงสร้างภาษาอย่างเหมาะสมในบริบทของการพูดและการเขียน (Little, 2006) ข้อสอบโครงสร้างภาษานี้ ครอบคลุมทั้งโครงสร้างประโยค

หลักไวยากรณ์ และการใช้คำศัพท์ เพื่อวัดความสามารถของผู้สอบในการสร้างประโยคภาษาอังกฤษที่ถูกต้องตามหลักภาษา ซึ่งเป็นองค์ประกอบสำคัญของการเรียนรู้ภาษาอังกฤษตามกรอบ CEFR

ทักษะการอ่าน ข้อสอบ STOU-EPT ในส่วนของทักษะการอ่านเป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ ข้อสอบส่วนนี้ให้ความสำคัญกับการตีความและทำความเข้าใจกับเนื้อหาที่มีความซับซ้อนตามระดับที่กำหนดในกรอบ CEFR โดยเฉพาะในระดับ B1 และ B2 ซึ่งผู้สอบควรมีความสามารถในการทำความเข้าใจเนื้อหาเชิงลึกจากบทความหรือบทสนทนาที่ซับซ้อน (Council of Europe, 2018) ข้อสอบทักษะการอ่าน ครอบคลุมเนื้อหาหลากหลายประเภท ช่วยให้ผู้สอบได้แสดงความสามารถในการวิเคราะห์ สังเคราะห์ และสรุปสาระสำคัญของเนื้อหาที่มีความยาวหรือมีโครงสร้างที่ซับซ้อน ซึ่งสอดคล้องกับเป้าหมายของกรอบ CEFR ที่มุ่งเน้นให้ผู้สอบสามารถใช้ทักษะการอ่านอย่างมีประสิทธิภาพในบริบททางวิชาการและวิชาชีพ

ผลการวิจัยนี้ แสดงให้เห็นว่า แบบทดสอบ STOU-EPT เป็นเครื่องมือที่สามารถประเมินความสามารถทางภาษาอังกฤษของผู้สอบได้สอดคล้องกับกรอบ CEFR และช่วยให้สามารถกำหนดระดับความสามารถของผู้สอบโดยเทียบเคียงกับมาตรฐานสากล การแบ่งโครงสร้างแบบทดสอบออกเป็นทักษะย่อยทำให้ผู้สอบทราบจุดแข็งและจุดที่ควรพัฒนาในแต่ละทักษะได้ ซึ่งเอื้อต่อการพัฒนาภาษาอังกฤษอย่างเป็นระบบ นอกจากนี้ ผลการทดสอบ STOU-EPT ยังมีความน่าเชื่อถือและสามารถนำไปใช้เป็นข้อมูลอ้างอิงสำหรับการพัฒนาทักษะทางภาษา ตลอดจนนำไปใช้ในบริบทที่หลากหลาย

1.2 การศึกษาความสอดคล้องระหว่างแบบทดสอบ STOU-EPT กับกรอบ CEFR ด้วยวิธีของเวบบ์ พิจารณาจาก 3 เกณฑ์ ได้แก่ ความสอดคล้องของเนื้อหา ความสอดคล้องด้านความลึกของความรู้ และความสอดคล้องด้านความสมดุล

ความสอดคล้องของเนื้อหา พบว่า ข้อสอบ STOU-EPT สอดคล้องกับมาตรฐานกรอบ CEFR ทั้ง 3 ทักษะ ได้แก่ การฟัง โครงสร้างภาษา และการอ่าน ซึ่งผ่านเกณฑ์ความสอดคล้องของเนื้อหาตามวิธีของ Webb (1997) กำหนดว่า ข้อสอบควรมีอย่างน้อย 6 ข้อ สำหรับวัดความรู้ในแต่ละมาตรฐาน ในส่วนของทักษะการฟัง พบว่า ข้อสอบ STOU-EPT สอดคล้องกับมาตรฐานกรอบ CEFR ในระดับ A2 (8 ข้อ) ระดับ B1 (8 ข้อ) และระดับ B2 (9 ข้อ) และความเที่ยงระหว่างผู้ประเมิน พบว่า ผู้ประเมินประเมินสอดคล้องกันในระดับสูงมาก ($p = .902$) ทักษะโครงสร้างภาษา พบว่า ข้อสอบ STOU-EPT สอดคล้องกับมาตรฐานกรอบ CEFR ในระดับ A2 (12 ข้อ) ระดับ B1 (12 ข้อ) และระดับ B2 (10 ข้อ) อย่างไรก็ตาม ความเที่ยงระหว่างผู้ประเมิน พบว่า ผู้ประเมินประเมินสอดคล้องกันในระดับปานกลาง ($p = .588$) ทั้งนี้อาจเนื่องมาจากคำบรรยายคุณลักษณะ (descriptors) ของกรอบ CEFR มีลักษณะกว้างและไม่ได้ออกแบบมาเพื่อใช้เป็นแนวทางในการประเมินทักษะโครงสร้างภาษาโดยตรง (Al Lawati, 2023) ซึ่งอาจส่งผลให้ผู้ประเมินความสอดคล้องตีความแตกต่างกัน โดยเฉพาะในส่วน

ของโครงสร้างภาษาที่มีความซับซ้อน ซึ่งอาจนำไปสู่ความคลาดเคลื่อนหรือความไม่สอดคล้องในการให้คะแนนได้ ส่วนทักษะการอ่าน พบว่า ข้อสอบ STOU-EPT สอดคล้องกับมาตรฐานกรอบ CEFR ในระดับ A2 (6 ข้อ) ระดับ B1 (19 ข้อ) และระดับ B2 (15 ข้อ) และความเที่ยงระหว่างผู้ประเมิน พบว่า ผู้ประเมินประเมินสอดคล้องกันในระดับปานกลาง ($p = .432$) ทั้งนี้ อาจเนื่องมาจากผู้ประเมินมีความเข้าใจเกี่ยวกับมาตรฐานหรือเกณฑ์การประเมินที่กำหนดไว้แตกต่างกัน ส่งผลให้เกิดความคลาดเคลื่อนในการให้คะแนน (Cephe and Toprak, 2014; Lipnevich, Khorramdel, & Smith, 2023) ซึ่งนำไปสู่ความแปรปรวนในการให้คะแนนโดยรวม ดังนั้น การวิจัยในอนาคตควรมุ่งเน้นไปที่การพัฒนาการฝึกอบรมผู้ประเมินและการกำหนดเกณฑ์การประเมินให้ชัดเจนยิ่งขึ้น

ความสอดคล้องด้านความลึกของความรู้ พบว่า ข้อสอบ STOU-EPT ทักษะการอ่าน สอดคล้องกับความลึกของความรู้ในทุกระดับความสามารถ (A2 B1 และ B2) ส่วนทักษะการฟัง และทักษะโครงสร้างภาษา สอดคล้องกับความลึกของความรู้ในระดับความสามารถ B1 และ B2 ส่วน A2 ไม่ผ่านเกณฑ์ ตามวิธีของ Webb (1997) กำหนดว่า มาตรฐานและเครื่องมือประเมินที่ถือว่ามีความลึกของความรู้ เมื่อข้อสอบอย่างน้อย 50% ของข้อสอบทั้งหมดในมาตรฐานนั้น ๆ มีความลึกเท่ากับหรือมากกว่าความลึกที่มีในตัวชี้วัด สำหรับทักษะการฟัง ข้อสอบระดับ B1 และ B2 (% hit = 100) สอดคล้องกับความลึกของความรู้ แต่ข้อสอบระดับ A2 (% hit = 0) ไม่ผ่านเกณฑ์ ซึ่งบ่งชี้ว่าข้อสอบในระดับ A2 อาจมีความซับซ้อนทางสมองไม่เพียงพอ ทักษะโครงสร้างภาษา ข้อสอบระดับ B1 (% hit = 100) และ B2 (% hit = 92.31) สอดคล้องกับความลึกของความรู้ แต่ข้อสอบระดับ A2 (% hit = 0) ไม่ผ่านเกณฑ์ ซึ่งบ่งชี้ว่าข้อสอบในระดับ A2 อาจมีความซับซ้อนทางสมองไม่เพียงพอเช่นเดียวกับทักษะการฟัง ส่วนทักษะการอ่าน ข้อสอบระดับ A2 และ B1 (% hit = 100) และ B2 (% above = 100) สอดคล้องกับความลึกของความรู้ ซึ่งแสดงให้เห็นถึงความเหมาะสมของระดับความซับซ้อนทางสมองของข้อสอบในทักษะนี้ นอกจากนี้ ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมินทั้งทักษะการฟัง และทักษะโครงสร้างภาษา พบว่า ผู้ประเมินประเมินสอดคล้องกันในระดับปานกลาง ($p = .525$ และ $.442$ ตามลำดับ) ส่วนทักษะการอ่าน ผู้ประเมินประเมินสอดคล้องกันในระดับต่ำ ($p = .249$) ซึ่งอาจเป็นผลมาจากการตีความระดับความซับซ้อนของความคิดที่ผู้สอบจะต้องทำข้อสอบให้ถูก ซึ่งเวบบ์ (Webb, 1997) แบ่งความลึกของความรู้ออกเป็น 4 ระดับ ได้แก่ ความจำ การมีทักษะและมโนทัศน์ การคิดเชิงกลยุทธ์ และการคิดหลายขั้นตอน ความแตกต่างในการตีความนี้ อาจส่งผลให้เกิดความคลาดเคลื่อนในการให้คะแนน (Lipnevich, Khorramdel, & Smith, 2023) โดยสรุปข้อสอบ STOU-EPT สอดคล้องกับความลึกของความรู้ ในระดับ B1 และ B2 ได้เป็นอย่างดี อย่างไรก็ตาม ข้อสอบระดับ A2 ในทักษะการฟังและทักษะโครงสร้างภาษา จำเป็นต้องได้รับการปรับปรุงเพื่อให้สอดคล้องกับระดับความซับซ้อนทางสมองที่เหมาะสมยิ่งขึ้น การพัฒนาแนวทางการฝึกอบรมผู้ประเมิน และการปรับปรุง

ตัวชี้วัดสำหรับประเมินให้ชัดเจน อาจช่วยเพิ่มความแม่นยำ และความน่าเชื่อถือของการประเมินในอนาคต โดยเฉพาะในทักษะที่มีค่าความเที่ยงของผู้ประเมินในระดับต่ำ

ความสอดคล้องด้านความสมดุล พบว่า ข้อสอบ STOU-EPT ทั้ง 3 ทักษะ มีการกระจายข้อสอบในทุกระดับความสามารถอย่างสมดุล โดยค่าดัชนีความสมดุล (Q) ของทักษะการฟัง เท่ากับ .853 ทักษะโครงสร้างภาษา เท่ากับ .952 และทักษะการอ่าน เท่ากับ .817 ซึ่งเป็นไปตามเกณฑ์ของ Webb (2007) กำหนดว่า ค่าดัชนีความสมดุลควรมีค่าตั้งแต่ .70 ขึ้นไป จึงถือว่าอยู่ในระดับที่ยอมรับได้ ผลการศึกษานี้สะท้อนให้เห็นว่า ข้อสอบ STOU-EPT ได้ออกแบบให้มีการกระจายข้อสอบในแต่ละระดับความสามารถอย่างสมดุล ส่งผลให้สามารถวัดความสามารถทางภาษาของผู้สอบได้อย่างแม่นยำ ลดความลำเอียงของข้อสอบ และเป็นหลักฐานสนับสนุนความตรงของการวัดตามหลักการของ Webb (2007)

การสังเคราะห์มิติความสอดคล้องทั้ง 3 ด้าน ทำให้เห็นภาพเชิงลึกเกี่ยวกับความตรงเชิงตีความของข้อสอบ STOU-EPT ระดับ B1 และ B2 โดยพบว่ามี ความสอดคล้องกันระหว่างเกณฑ์ด้านเนื้อหา ระดับความลึกของความรู้ และความสมดุลของข้อสอบ ซึ่งสนับสนุนความเหมาะสมในการใช้เพื่อการตัดสินใจที่มีความสำคัญสูง (high-stakes) ในบริบทของมหาวิทยาลัย เช่น การสำเร็จการศึกษา การรับรองความสามารถทางภาษาอังกฤษ ซึ่งสอดคล้องกับทิศทางนโยบายระดับชาติ (คณะกรรมการมาตรฐานการอุดมศึกษา, 2559) ความครอบคลุมเนื้อหาอย่างรอบด้าน การสอดคล้องกับระดับความลึกของความรู้ และค่าดัชนี Q ที่อยู่ในระดับสูง สะท้อนให้เห็นว่าข้อสอบเป็นตัวแทนของเนื้อหาและทักษะที่ต้องการวัดตามกรอบการประเมินที่กำหนดไว้

ความไม่สอดคล้องของข้อสอบ STOU-EPT ในระดับ A2 โดยเฉพาะในทักษะการฟังและโครงสร้างภาษา ส่งผลต่อความน่าเชื่อถือในการตีความคะแนนผู้สอบที่มีความสามารถขั้นต่ำ แม้ว่าทั้งสองทักษะนี้ออกแบบข้อสอบให้มีการกระจายในแต่ละระดับความสามารถอย่างสมดุล ($Q = .853$ สำหรับการฟัง และ $Q = .952$ สำหรับโครงสร้างภาษา) แต่กลับไม่ผ่านเกณฑ์ด้านระดับความลึกของความรู้ และมีค่าความสอดคล้องระหว่างผู้ประเมินอยู่ในระดับต่ำ ข้อบกพร่องทั้งสองประการนี้ชี้ให้เห็นว่าข้อสอบระดับ A2 อาจไม่สามารถเป็นตัวแทนของเนื้อหาและทักษะที่ต้องการวัดได้อย่างเพียงพอ ซึ่งส่งผลกระทบต่อความเป็นธรรมและความแม่นยำของการจำแนกระดับความสามารถตามกรอบ CEFR (Kane, 2013; AERA et al., 2014) หากนำข้อสอบดังกล่าวไปใช้ในการกำหนดคะแนนจุดตัด อาจเพิ่มความเสี่ยงต่อการจำแนกผู้สอบผิด โดยเฉพาะในช่วงคะแนนใกล้จุดตัด ซึ่งลักษณะข้อสอบที่ไม่ชัดเจนหรือไม่ตรงกับระดับความสามารถที่กำหนด อาจทำให้ผู้สอบบางคนถูกจัดระดับสูงหรือต่ำกว่าความเป็นจริง และส่งผลกระทบต่อความแม่นยำและความเป็นธรรมในการใช้ผลสอบเพื่อการสำเร็จการศึกษา หรือการรับรองความสามารถทางภาษาอังกฤษได้ (Green, 2020)

อย่างไรก็ตาม เพื่อยกระดับคุณภาพการประเมินความสอดคล้องและความตรงเชิงตีความของแบบทดสอบ STOU-EPT จึงเสนอแนวทางพัฒนา 3 ประการ ได้แก่ (1) ปรับปรุงข้อสอบระดับ A2 ให้สอดคล้องกับนิยามความสามารถขั้นต่ำตาม CEFR และออกแบบให้กระตุ้นกระบวนการคิดในระดับที่เหมาะสม (2) เพิ่มความสอดคล้องของการประเมินความสอดคล้องข้อสอบกับตัวชี้วัด CEFR ผ่านการฝึกอบรมผู้ประเมินด้วยข้อสอบตัวอย่าง โดยเน้นการทำความเข้าใจและตีความระดับความลึกของความรู้ให้ตรงกัน และ (3) จัดทำแนวทางการพัฒนาข้อสอบที่กำหนดขั้นตอนและแนวทางปฏิบัติที่ชัดเจนสำหรับการนำตัวชี้วัด CEFR ไปใช้ในการออกแบบข้อสอบ เพื่อใช้เป็นมาตรฐานและสร้างความสอดคล้องในการประเมิน โดยเฉพาะในระดับความสามารถที่ต่ำ การดำเนินการดังกล่าวอาจช่วยเพิ่มความน่าเชื่อถือของแบบทดสอบ STOU-EPT ทั้งในด้านความเป็นธรรมและความแม่นยำ และเสริมศักยภาพให้สามารถใช้เป็นแบบทดสอบอ้างอิง CEFR ได้อย่างมั่นใจในบริบทอุดมศึกษาไทย

ประเด็นที่ 2 การเทียบเคียงคะแนนแบบทดสอบวัดระดับความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราชกับกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรปด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง มีการอภิปรายใน 2 หัวข้อ ดังนี้

2.1 การเทียบเคียงคะแนนแบบทดสอบ STOU-EPT กับกรอบ CEFR ด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง ในรอบที่ 1 ถึงรอบที่ 3 แสดงผลใน 2 รูปแบบ ได้แก่ คะแนนจุดตัดแยกตามทักษะ และคะแนนจุดตัดรวม

คะแนนจุดตัดแยกตามทักษะ

ทักษะการฟัง พบว่า คะแนนจุดตัดสำหรับความสามารถระดับ B1 ในรอบที่ 3 มีค่าเท่ากับ 13.49 คะแนน (ประมาณ 13 คะแนน) ซึ่งมีค่าลดลงเมื่อเทียบกับรอบที่ 1 (15.75 คะแนน) และรอบที่ 2 (15.91 คะแนน) ผลการวิจัยนี้สะท้อนให้เห็นว่าการตัดสินใจและการปรับคะแนนของผู้เชี่ยวชาญมีความสอดคล้องและเหมาะสมมากขึ้นในแต่ละรอบการประเมิน เช่นเดียวกับคะแนนจุดตัดความสามารถระดับ B2 ในรอบที่ 3 มีค่าเท่ากับ 19.24 (ประมาณ 19 คะแนน) ซึ่งใกล้เคียงกับรอบที่ 1 (19.15 คะแนน) และรอบที่ 2 (19.30 คะแนน) โดยคะแนนจุดตัดระดับ B2 สูงกว่าระดับ B1 อยู่ 6 คะแนน แสดงถึงการแบ่งช่วงคะแนนที่ชัดเจนระหว่างระดับความสามารถดังกล่าว นอกจากนี้ ส่วนเบี่ยงเบนมาตรฐาน (SD) และค่าความคลาดเคลื่อนมาตรฐานของการตัดสินใจ (SEJ) ทั้งระดับ B1 และ B2 ในรอบที่ 3 มีค่าลดลงเมื่อเทียบกับรอบที่ 1 และรอบที่ 2 สะท้อนถึงความสอดคล้องในการตัดสินใจของผู้เชี่ยวชาญ และเพิ่มความน่าเชื่อถือของคะแนนจุดตัดในรอบสุดท้าย สอดคล้องกับ Kampa, Wagner, and Köller (2019) พบว่า การให้ข้อมูลเกี่ยวกับความสามารถของผู้สอบร่วมกับความคิดเห็นจากผู้เชี่ยวชาญช่วยเพิ่มความสอดคล้องของคะแนนจุดตัดและการแยกแยะระดับความสามารถได้ดีขึ้น และ Cizek and Bunch (2007) กล่าวว่า การกำหนดคะแนนจุดตัดมักดำเนินการในหลายรอบ (multiple rounds) ทั้งนี้ เพื่อให้ผู้เชี่ยวชาญมีโอกาสปรับปรุงการตัดสินใจ

เมื่อมีการให้ข้อมูลเพิ่มเติมในแต่ละรอบ เช่น ผลการวิเคราะห์ข้อสอบ ผลการประเมินรอบแรก การดำเนินการดังกล่าวช่วยเพิ่มความสอดคล้องและลดความคลาดเคลื่อนของคะแนนจุดตัด ซึ่งส่งผลต่อความน่าเชื่อถือของคะแนนจุดตัดในรอบสุดท้าย

ทักษะโครงสร้างภาษา พบว่า คะแนนจุดตัดสำหรับความสามารถระดับ B1 ในรอบที่ 3 มีค่าเท่ากับ 20.48 คะแนน (ประมาณ 20 คะแนน) ซึ่งมีค่าไม่แตกต่างกันอย่างมีนัยสำคัญเมื่อเทียบกับรอบที่ 1 (19.86 คะแนน) แต่มีค่าลดลงเล็กน้อยเมื่อเทียบกับรอบที่ 2 (22.12 คะแนน) ในขณะที่คะแนนจุดตัดสำหรับความสามารถระดับ B2 ในรอบที่ 3 มีค่าเท่ากับ 27.23 (ประมาณ 27 คะแนน) ซึ่งมีค่าใกล้เคียงกับรอบที่ 2 (27.30 คะแนน) แต่เพิ่มขึ้นเล็กน้อยเมื่อเทียบกับรอบที่ 1 (25.57 คะแนน) ทั้งนี้ คะแนนจุดตัดสำหรับความสามารถระดับ B2 สูงกว่าระดับ B1 อยู่ 7 คะแนน แสดงถึงการแบ่งช่วงคะแนนที่เหมาะสมระหว่างระดับความสามารถ สอดคล้องกับ AERA (2014) กล่าวว่าการกำหนดคะแนนจุดตัดที่เป็นไปได้และเหมาะสม คะแนนจุดตัดต้องอธิบายความแตกต่างระหว่างกลุ่มผู้สอบได้อย่างแท้จริง นอกจากนี้ การลดลงของส่วนเบี่ยงเบนมาตรฐาน (SD) และค่าความคลาดเคลื่อนมาตรฐานของการตัดสิน (SEJ) ทั้งระดับ B1 และ B2 ในรอบที่ 3 เมื่อเทียบกับรอบที่ 1 และรอบที่ 2 แสดงให้เห็นถึง ความสอดคล้องในการตัดสินมากขึ้นของผู้เชี่ยวชาญ ทั้งนี้อาจเนื่องมาจากมีการให้ข้อมูลเพิ่มเติม ได้แก่ คะแนนจุดตัดในรอบที่ 1 และข้อมูลเกี่ยวกับคุณภาพข้อสอบ รายข้อ ในรอบที่ 2 รวมทั้งมีการแลกเปลี่ยนความคิดเห็นระหว่างผู้เชี่ยวชาญในแต่ละรอบ การสนับสนุนเหล่านี้ อาจช่วยให้ผู้เชี่ยวชาญเข้าใจการกำหนดคะแนนจุดตัดที่ชัดเจนและสามารถตัดสินใจได้อย่างแม่นยำมากขึ้น สอดคล้องกับ Lee et al. (2024) พบว่า การอภิปรายระหว่างรอบที่ 1 และรอบที่ 2 มีบทบาทสำคัญในการช่วยให้ผู้เชี่ยวชาญเข้าใจเกี่ยวกับผู้เข้าสอบกลุ่ม “borderline” ซึ่งช่วยให้ผู้เชี่ยวชาญสามารถปรับปรุงคะแนนจุดตัดให้สะท้อนกลุ่มเป้าหมายได้ดียิ่งขึ้น

ทักษะการอ่าน พบว่า คะแนนจุดตัดสำหรับความสามารถระดับ B1 ในรอบที่ 3 มีค่าเท่ากับ 23.26 คะแนน (ประมาณ 23 คะแนน) ซึ่งมีค่าลดลงเมื่อเทียบกับรอบที่ 1 (24.53 คะแนน) และรอบที่ 2 (26.27 คะแนน) ในขณะที่คะแนนจุดตัดสำหรับความสามารถระดับ B2 ในรอบที่ 3 มีค่าเท่ากับ 31.43 (ประมาณ 31 คะแนน) ซึ่งมีค่าเพิ่มขึ้นเล็กน้อยเมื่อเทียบกับรอบที่ 1 (30.28 คะแนน) แต่มีค่าน้อยลงเมื่อเทียบกับรอบที่ 2 (32.02 คะแนน) ทั้งนี้ คะแนนจุดตัดสำหรับความสามารถระดับ B2 สูงกว่าระดับ B1 อยู่ 8 คะแนน สะท้อนถึงการแบ่งช่วงคะแนนที่ชัดเจนระหว่างระดับความสามารถ และส่วนเบี่ยงเบนมาตรฐาน (SD) และค่าความคลาดเคลื่อนมาตรฐานของการตัดสิน (SEJ) ทั้งระดับ B1 และ B2 ในรอบที่ 3 มีค่าลดเมื่อเทียบกับรอบที่ 1 และรอบที่ 2 แสดงให้เห็นถึง ความสอดคล้องในการตัดสินของผู้เชี่ยวชาญ และเพิ่มความน่าเชื่อถือของคะแนนจุดตัดในรอบสุดท้าย สอดคล้องกับ Yim and Shin (2020) พบว่า ข้อสอบจำลองครั้งที่ 2 มีค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ต่ำกว่าข้อสอบครั้งที่ 1 แสดงถึงความสอดคล้องในการตัดสินใจของผู้เชี่ยวชาญ นอกจากนี้ การที่ผู้เชี่ยวชาญ

ร่วมกันกำหนดนิยามของความสามารถขั้นต่ำ (minimally competent person; MCP) ผ่านการอภิปรายและแลกเปลี่ยนความคิดเห็นช่วยให้การตัดสินคะแนนจุดตัดมีความแม่นยำและสะท้อนความสามารถที่แท้จริงของผู้เข้าสอบได้ดียิ่งขึ้น ผลลัพธ์ดังกล่าวสะท้อนให้เห็นว่า การสนับสนุนข้อมูลหรือการให้ข้อมูลเพิ่มเติมที่เหมาะสม สามารถเพิ่มความน่าเชื่อถือและความแม่นยำในผลการตัดสินของผู้เชี่ยวชาญได้

คะแนนจุดตัดรวม

คะแนนจุดตัดรวมของแบบทดสอบ STOU-EPT พบว่า คะแนนจุดตัดสำหรับความสามารถระดับ B1 มีค่าเท่ากับ 56 คะแนน และคะแนนจุดตัดสำหรับความสามารถระดับ B2 มีค่าเท่ากับ 77 คะแนน ซึ่งคะแนนจุดตัดแบบรวมช่วยสะท้อนความสามารถโดยรวมของผู้สอบได้อย่างเหมาะสม โดยคะแนนจุดตัดระดับ B1 เท่ากับ 56 คะแนน แสดงถึงผู้สอบที่มีความสามารถพื้นฐานในการสื่อสารและเข้าใจเนื้อหาภาษาอังกฤษในระดับทั่วไป เช่น การใช้ภาษาในชีวิตประจำวันหรือสถานการณ์เบื้องต้น (Council of Europe, 2020) ในขณะที่คะแนนจุดตัดระดับ B2 เท่ากับ 77 คะแนน สะท้อนถึงความสามารถที่สูงขึ้น ซึ่งผู้สอบในระดับนี้สามารถเข้าใจเนื้อหาที่ซับซ้อนและสื่อสารได้อย่างมีประสิทธิภาพในบริบทที่หลากหลาย (Green, 2012) นอกจากนี้ การกำหนดคะแนนจุดตัดแบบรวมช่วยส่งเสริมความยืดหยุ่นในกระบวนการสอบ กล่าวคือ ผู้สอบที่มีจุดแข็งในบางทักษะ เช่น การอ่านหรือโครงสร้างภาษา สามารถชดเชยจุดอ่อนในทักษะอื่น เช่น การฟัง ได้อย่างยุติธรรม (Papageorgiou et al., 2015) ซึ่งเป็นการลดความกังวลเกี่ยวกับข้อจำกัดเฉพาะด้าน และสนับสนุนการสอบที่เหมาะสมสำหรับผู้สอบหลากหลายกลุ่ม

การนำคะแนนจุดตัดมาแบ่งช่วงในแต่ละระดับความสามารถ แสดงผลใน 2 รูปแบบ ได้แก่ แบบที่ 1 คะแนนจุดตัดแยกตามทักษะ พบว่า ทักษะการฟัง ระดับ A2 (1-12 คะแนน) ระดับ B1 (13-18 คะแนน) และระดับ B2 (19-25 คะแนน) ทักษะโครงสร้างภาษา ระดับ A2 (1-19 คะแนน) ระดับ B1 (20-26 คะแนน) และระดับ B2 (27-35 คะแนน) และทักษะการอ่าน ระดับ A2 (1-22 คะแนน) ระดับ B1 (23-30 คะแนน) และระดับ B2 (31-40 คะแนน) ซึ่งคะแนนจุดตัดแยกตามทักษะช่วยให้การประเมินความสามารถเฉพาะทักษะได้อย่างละเอียด และในแต่ละช่วงคะแนนสะท้อนถึงความก้าวหน้าในแต่ละระดับความสามารถได้อย่างชัดเจน สอดคล้องกับ Ruegg et al. (2024) ซึ่งชี้ให้เห็นว่า ทักษะการอ่าน และการเขียน มีความสำคัญต่อความสำเร็จทางการศึกษาในระดับมหาวิทยาลัยมากกว่าทักษะอื่น โดยเฉพาะในงานวิชาการที่ต้องใช้การวิเคราะห์และสรุปข้อมูลที่ซับซ้อน เช่นเดียวกับผลการแบ่งคะแนนจุดตัดที่พบว่าทักษะการอ่านในระดับ B2 (31-40 คะแนน) มีความซับซ้อนมากขึ้นเมื่อเทียบกับระดับ A2 (1-22 คะแนน) นอกจากนี้การกำหนดคะแนนจุดตัดแยกตามทักษะการฟัง โครงสร้างภาษา และการอ่าน มีข้อดีที่ช่วยให้มหาวิทยาลัยออกแบบหลักสูตรหรือกิจกรรมการเรียนการสอนที่มุ่งเน้นทักษะเฉพาะช่วยให้ผู้เรียนสามารถพัฒนาทักษะที่สำคัญได้

อย่างเหมาะสมและสอดคล้องกับระดับความสามารถของผู้เรียน เช่น Green (2007) ซึ่งชี้ให้เห็นว่าการเรียนการสอนที่เจาะจงต่อทักษะ เช่น การอ่านและการเขียน ช่วยเตรียมความพร้อมสำหรับการเรียนในมหาวิทยาลัยได้ดียิ่งขึ้น โดยเฉพาะในระดับที่สูงขึ้น เช่น ระดับ B1 และ B2 ที่ต้องการความสามารถในการตีความและวิเคราะห์ข้อมูลที่ซับซ้อน เช่นเดียวกับ Neumann et al. (2019) ระบุว่า การออกแบบกิจกรรมที่เสริมสร้างการคิดวิเคราะห์ควบคู่กับการอ่านและการเขียน ช่วยให้ผู้เรียนต่างชาติที่ต้องใช้ภาษาอังกฤษเป็นภาษาที่สองสามารถเพิ่มศักยภาพในการเรียนได้อย่างมีประสิทธิภาพ ส่วนผู้เรียนที่อยู่ในระดับพื้นฐาน เช่น ระดับ A2 การพัฒนาทักษะโครงสร้างภาษาเป็นจุดเริ่มต้นที่สำคัญ Oliver et al. (2012) กล่าวว่า การเน้นโครงสร้างภาษาที่ถูกต้องในระดับต้นจะช่วยสร้างพื้นฐานที่มั่นคงสำหรับผู้เรียนในการพัฒนาทักษะอื่น ๆ ในระดับที่สูงขึ้น ดังนั้น การกำหนดคะแนนจุดตัดแยกตามทักษะ ไม่เพียงแต่ช่วยให้การวางแผนการเรียนการสอนมีความชัดเจน แต่ยังสามารถสนับสนุนให้ผู้เรียนได้พัฒนาทักษะในแต่ละด้านที่เหมาะสมกับระดับความสามารถของตนเอง รวมทั้งการออกแบบกิจกรรมการเรียนที่สอดคล้องกับความต้องการเฉพาะด้านของผู้เรียนในแต่ละทักษะอาจช่วยเพิ่มประสิทธิภาพในการเรียนรู้และส่งเสริมความสำเร็จทางการศึกษาในระยะยาว สำหรับแบบที่ 2 คะแนนจุดตัดรวม พบว่า ระดับ A2 (1-55 คะแนน) ระดับ B1 (56-76 คะแนน) และระดับ B2 (77-100 คะแนน) การประเมินลักษณะนี้เปิดโอกาสให้ผู้สอบที่มีจุดแข็งบางทักษะสามารถชดเชยจุดอ่อนในทักษะอื่นๆ ความยืดหยุ่นนี้ช่วยส่งเสริมความยุติธรรมในกระบวนการสอบ โดยเฉพาะในบริบทที่เน้นการคัดเลือกผู้เรียนเข้าศึกษาต่อ สอดคล้องกับ Ruegg et al. (2024) พบว่า คะแนนรวมของ IELTS มีความสามารถในการทำนายความสำเร็จทางการศึกษาของนักศึกษาได้ในระดับปานกลาง ($R^2 = 0.158$) แสดงให้เห็นว่า ผลการศึกษาดังกล่าวสนับสนุนว่า คะแนนรวมสามารถใช้เป็นตัวชี้วัดที่ดีในการพิจารณาความพร้อมด้านภาษาอังกฤษของนักศึกษาในบริบทของมหาวิทยาลัย

2.2 การประเมินกระบวนการกำหนดคะแนนจุดตัด

ผู้เชี่ยวชาญที่เข้าร่วมกำหนดคะแนนจุดตัด จำนวน 10 คน พบว่า ส่วนใหญ่มีคุณวุฒิระดับปริญญาเอก จำนวน 9 คน (90%) โดยสำเร็จการศึกษาในสาขาวิชา Ph.D. (Applied Linguistics) และ Ph.D. (English) มากที่สุด มีประสบการณ์ด้านการสอนภาษาอังกฤษในระดับอุดมศึกษา 21-25 ปี มากที่สุด จำนวน 5 คน (50%) รองลงมา 26 ปีขึ้นไป จำนวน 3 คน (30%) และมีประสบการณ์ด้านการจัดสอบหรือการสอนภาษาอังกฤษตามกรอบ CEFR มากกว่า 10 ปีขึ้นไป จำนวน 6 คน (60%) รองลงมา 5-10 ปี จำนวน 2 คน (20%) และน้อยกว่า 5 ปี จำนวน 2 คน (20%) ซึ่งการที่ผู้เชี่ยวชาญมีคุณวุฒิและประสบการณ์ตรงทั้งในด้านการสอนและการจัดสอบตามกรอบ CEFR ช่วยให้กระบวนการกำหนดคะแนนจุดตัดแบบทดสอบ STOU-EPT มีความน่าเชื่อถือและสะท้อนมาตรฐานสากลได้แม่นยำ สอดคล้องกับ Yim and Shin (2020) ซึ่งชี้ให้เห็นว่า การใช้ผู้เชี่ยวชาญที่มีประสบการณ์อย่างน้อย 5 ปีในด้านการพัฒนาข้อสอบ ประสบการณ์การสอน และประสบการณ์

ภาคปฏิบัติ มีบทบาทสำคัญในการกำหนดคะแนนจุดตัดด้วยวิธีการของแองกอฟสำหรับการสอบ
พยายาลแห่งชาติในเกาหลีใต้ช่วยให้คะแนนจุดตัดที่กำหนดมีความถูกต้องและเหมาะสมมากยิ่งขึ้น

ผลการประเมินกระบวนการกำหนดคะแนนจุดตัดทั้ง 12 ประเด็น พบว่า มีประเด็นที่
ผู้เชี่ยวชาญมีความคิดเห็นเกี่ยวกับกระบวนการกำหนดคะแนนจุดตัดด้วยวิธีการของแองกอฟที่ได้รับการ
การปรับปรุงในระดับมากที่สุด (60% ขึ้นไป) 5 ประเด็น ได้แก่ (1) แบบประมาณโอกาสการทำข้อสอบ
ถูกในรอบที่ 1 2 และ 3 ง่ายต่อการกรอกข้อมูล (60%, $M = 3.60$, $SD = .52$) (2) ฉันทมีโอกาสแสดง
ความคิดเห็นเกี่ยวกับผลการกำหนดคะแนนจุดตัดแต่ละรอบ (80%, $M = 3.80$, $SD = .42$) (3) เวลา
ที่ให้ฉันทแสดงความคิดเห็นเกี่ยวกับผลการกำหนดคะแนนจุดตัดในแต่ละรอบมีเพียงพอ (70%,
 $M = 3.70$, $SD = .48$) (4) การประชุมหลังการกำหนดคะแนนจุดตัดในรอบที่ 1 มีประโยชน์ต่อการ
กำหนดคะแนนจุดตัด (80%, $M = 3.80$, $SD = .42$) และ (5) การประชุมหลังการกำหนดคะแนน
จุดตัดในรอบที่ 2 มีประโยชน์ต่อการกำหนดคะแนนจุดตัด (70%, $M = 3.70$, $SD = .48$) และประเด็น
ที่ผู้เชี่ยวชาญมีความคิดเห็นเกี่ยวกับกระบวนการกำหนดคะแนนจุดตัดด้วยวิธีการของแองกอฟที่ได้รับการ
การปรับปรุงในระดับมาก (60% ขึ้นไป) 3 ประเด็น ได้แก่ (1) ผู้ดำเนินการอธิบายวิธีการกำหนด
คะแนนจุดตัดได้อย่างชัดเจน (60%, $M = 3.20$, $SD = .63$) (2) การปฐมนิเทศและการฝึกซ้อมช่วยให้
ฉันทเข้าใจวิธีการกำหนดคะแนนจุดตัดมากขึ้น (60%, $M = 3.50$, $SD = .71$) และ (3) คู่มือสำหรับการ
กำหนดคะแนนจุดตัดด้วยวิธีการของแองกอฟที่ได้รับการปรับปรุง มีความชัดเจนและเข้าใจง่าย (60%,
 $M = 3.40$, $SD = .84$) สอดคล้องกับ Papageorgiou et al. (2019) ได้สำรวจความคิดเห็นของ
ผู้เชี่ยวชาญที่เข้าร่วมกระบวนการกำหนดมาตรฐาน พบว่า ผู้เชี่ยวชาญส่วนใหญ่ “เห็นด้วย” และ
“เห็นด้วยอย่างยิ่ง” ว่ากระบวนการกำหนดมาตรฐานมีความชัดเจน มีการให้ข้อมูลสนับสนุนที่เพียงพอ
ที่ช่วยให้ผู้เชี่ยวชาญตัดสินใจได้อย่างเหมาะสมและมั่นใจ และมีความคิดเห็นเชิงบวกเกี่ยวกับความ
เหมาะสมของข้อมูล และวิธีการที่ใช้ในกระบวนการกำหนดมาตรฐาน เช่น การฝึกอบรมก่อนการ
ประชุม (pre-meeting training) เพื่อเพิ่มความเข้าใจและความมั่นใจในกระบวนการกำหนด
มาตรฐาน การดำเนินการกำหนดคะแนนสามรอบ (three-round process) เป็นการเปิดโอกาสให้
ผู้เชี่ยวชาญปรับปรุงการตัดสินใจในแต่ละรอบ โดยอาศัยข้อมูลย้อนกลับเพื่อเพิ่มความแม่นยำและความ
น่าเชื่อถือในการกำหนดคะแนนจุดตัด

ประเด็นที่ 3 การกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดด้วยวิธีการของแองกอฟที่ได้รับการ
ปรับปรุง

การกำหนดมาตรฐานเพื่อกำหนดคะแนนจุดตัดในครั้งนี้ ผู้วิจัยใช้วิธีการของแองกอฟที่ได้รับการ
การปรับปรุง ซึ่งถือว่าเป็นวิธีการกำหนดมาตรฐานเพื่อกำหนดคะแนนจุดตัดที่ได้รับความนิยมในการ
นำไปใช้กำหนดคะแนนจุดตัดทางภาษาอย่างกว้างขวาง เช่น งานวิจัยของ Baron and Papageorgiou
(2016), Papageorgiou et al. (2019), Tannenbaum and Baron (2011) รวมทั้งวิธีการดังกล่าว

เป็นวิธีที่ง่าย และเหมาะสมกับลักษณะข้อสอบ STOU-EPT เนื่องจากมีการให้คะแนนเป็น 2 ค่า ได้แก่ ตอบถูกให้ 1 ตอบผิดให้ 0 การกำหนดคะแนนจุดตัดในการวิจัยนี้ดำเนินการโดยผู้เชี่ยวชาญ จำนวน 10 ท่าน ที่มีประสบการณ์ด้านการสอนภาษาอังกฤษ ตั้งแต่ 5 ปี ขึ้นไป และมีประสบการณ์เกี่ยวกับการทดสอบภาษาอังกฤษตามกรอบ CEFR ซึ่งการกำหนดจุดตัดโดยผู้เชี่ยวชาญหลายคน อาจดูเหมือนว่าทำให้เกิดความยากลำบากในการที่จะให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดให้สอดคล้องกัน เนื่องจากผู้เชี่ยวชาญต้องจินตนาการว่า “ผู้สอบที่เป็นบุคคลทั่วไปที่มีความสามารถสูงสุดในระดับ B1 และ B2 จะทำข้อสอบข้อที่ i ถูกกี่เปอร์เซ็นต์” ผู้วิจัยตระหนักถึงปัญหาดังกล่าว จึงได้มีการประชุมนิเทศผู้เชี่ยวชาญก่อนกำหนดคะแนนจุดตัด โดยในการประชุมนิเทศผู้วิจัยได้เตรียมคู่มือและแบบฟอร์มที่เกี่ยวข้อง เพื่อแนะนำวิธีการใช้คู่มือการกำหนดคะแนนจุดตัด และอธิบายวิธีการบันทึกข้อมูลลงในแบบฟอร์มที่จัดเตรียมไว้ รวมถึงมีการอธิบายลักษณะของผู้สอบที่มีความสามารถอยู่ระหว่างชายขอบ โดยเน้นอธิบายลักษณะของผู้สอบที่อยู่ในระดับก้ำกึ่งระหว่างคะแนนจุดตัด 2 ระดับ ได้แก่ ระดับ B1 และระดับ B2 จำแนกตามทักษะการฟัง โครงสร้างภาษา และการอ่าน เพื่อช่วยให้ผู้เชี่ยวชาญสามารถพิจารณาและตัดสินใจในการกำหนดคะแนนจุดตัดได้อย่างแม่นยำ รวมทั้งผู้วิจัยได้เปิดโอกาสให้ผู้เชี่ยวชาญซักถาม แสดงความคิดเห็น และปรับความเข้าใจร่วมกันให้ตรงกันก่อนเริ่มกระบวนการกำหนดคะแนนจุดตัด หลังจากผู้เชี่ยวชาญทำความเข้าใจแนวทางดังกล่าวร่วมกันจนมีความชัดเจนตรงกันแล้ว ผู้วิจัยให้ผู้เชี่ยวชาญฝึกซ้อมการกำหนดคะแนนจุดตัด ก่อนเริ่มกำหนดคะแนนจุดตัดในรอบที่ 1 การกำหนดคะแนนจุดตัดในครั้งนี้ดำเนินการ 3 รอบ และหลังจากกำหนดจุดตัดแต่ละรอบ ผู้วิจัยได้นำเสนอผลและให้ข้อมูลย้อนกลับในแต่ละรอบ สำหรับในรอบที่ 2 ผู้วิจัยได้ให้ข้อมูลประกอบการตัดสินใจเพิ่มเติม ได้แก่ 1) ผลการตอบของตนเองและผลการตอบในภาพรวมของกลุ่มผู้เชี่ยวชาญ ในรอบที่ 1 ระดับ B1 และระดับ B2 จำแนกตามทักษะการฟัง โครงสร้างภาษา และการอ่าน และ 2) ค่าความยากง่ายของข้อสอบ STOU-EPT ที่วิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบด้วย Rasch testlet model รวมถึงมีการอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญ การดำเนินการเช่นนี้ทำให้ส่วนเบี่ยงเบนมาตรฐาน ค่าความคลาดเคลื่อนมาตรฐานของการตัดสินใจ มีค่าลดเมื่อเทียบกับรอบที่ 1 และรอบที่ 2 แสดงให้เห็นถึง ความสอดคล้องในการตัดสินใจของผู้เชี่ยวชาญ สอดคล้องกับ Yim and Shin (2020) สะท้อนให้เห็นว่า การกำหนดคะแนนจุดตัดที่เหมาะสมมีความสำคัญอย่างยิ่ง เพราะส่งผลโดยตรงต่อการตัดสินใจรับเข้าศึกษาและเส้นทางการเรียนของผู้สอบ Cizek and Bunch (2007) กล่าวว่า กระบวนการกำหนดมาตรฐานที่มีประสิทธิภาพเป็นสิ่งสำคัญในการรับรองว่าการประเมินผลสะท้อนความสามารถของผู้สอบได้อย่างแม่นยำ และช่วยให้ผู้สอบสามารถเข้าถึงโอกาสทางการศึกษาได้อย่างเท่าเทียม และ Wang et al. (2022) เน้นว่า คะแนนจุดตัดที่เหมาะสมสามารถช่วยให้ผู้สอบเข้าถึงโปรแกรมการศึกษาได้อย่างเท่าเทียม ในขณะที่เดียวกันยังคงรักษามาตรฐานวิชาการที่เข้มงวด นอกจากนี้ การกำหนดคะแนนจุดตัดด้วยวิธีที่โปร่งใสและเป็นระบบช่วยเพิ่มความน่าเชื่อถือ

ให้กับการประเมินทางภาษาอังกฤษ ดังนั้น การมีผู้เชี่ยวชาญเข้าร่วมในกระบวนการนี้ช่วยให้มั่นใจว่าคะแนนจุดตัดที่ได้มาจากการตัดสินใจที่รอบคอบ ไม่ใช่การตัดสินใจแบบสุ่ม ความโปร่งใสดังกล่าวยังช่วยสร้างความมั่นใจให้กับทุกฝ่ายที่เกี่ยวข้อง เช่น ผู้สอบ ผู้เรียน ผู้สอน และผู้บริหาร ว่าผลการประเมินมีความยุติธรรมและแม่นยำ

นอกจากนั้น การกำหนดมาตรฐานเพื่อหาคะแนนจุดตัดด้วยวิธีเองกอฟที่ได้รับการปรับปรุงไม่เพียงช่วยให้การแปลผลคะแนนของข้อสอบ STOU-EPT มีความแม่นยำและสะท้อนความสามารถของผู้สอบตาม CEFR ได้ดียิ่งขึ้น ยังช่วยลดข้อจำกัดในงานวิจัยของ ศจี จิระโร และ สุภมาศ อังศุโชติ (2563) ซึ่งศึกษาความตรงตามเกณฑ์สัมพันธ์ระหว่างแบบทดสอบ STOU-EPT กับแบบทดสอบมาตรฐานอื่น ๆ พบว่า คะแนนจากแบบทดสอบ STOU-EPT ในระดับ A2 มักแสดงผลสูงเกินจริงเมื่อเทียบกับแบบทดสอบมาตรฐานอื่น ๆ ทั้งนี้ อาจเกิดจากความคลาดเคลื่อนของการแปลความหมายของคะแนนแบบทดสอบ STOU-EPT โดยเฉพาะคะแนนระดับ B1 ที่ค่อนข้างกว้าง การประยุกต์ใช้วิธีเองกอฟที่ได้รับการปรับปรุงในการกำหนดคะแนนจุดตัด โดยอาศัยความเห็นจากผู้เชี่ยวชาญ จำนวน 10 คน ร่วมกับข้อมูลจากการวิเคราะห์ด้วย Rasch Testlet Model ส่งผลให้กระบวนการกำหนดคะแนนจุดตัดมีความเป็นระบบมากยิ่งขึ้น ทั้งยังมีหลักฐานเชิงประจักษ์รองรับอย่างชัดเจน วิธีการนี้ช่วยลดความคลาดเคลื่อนในการแปลผลคะแนน และเพิ่มความสอดคล้องของผลคะแนนแบบทดสอบ STOU-EPT กับระดับความสามารถตามกรอบ CEFR ได้อย่างมีนัยสำคัญ แนวทางนี้สะท้อนให้เห็นถึงความก้าวหน้าในการพัฒนาระบบการวัดผลที่น่าเชื่อถือและยุติธรรม โดยเน้นกระบวนการกำหนดมาตรฐานที่มีความรัดกุม เพื่อให้มั่นใจว่าคะแนนจุดตัดที่ได้มีความเหมาะสม และสามารถใช้อำนาจระดับความสามารถของผู้สอบได้อย่างถูกต้องและเป็นธรรม

ดังนั้น การกำหนดคะแนนจุดตัดสำหรับแบบทดสอบ STOU-EPT ด้วยวิธีการของเองกอฟที่ได้รับการปรับปรุง เป็นกระบวนการสำคัญที่ช่วยให้การประเมินความสามารถทางภาษาอังกฤษมีประสิทธิภาพ ด้วยการใช้ความคิดเห็นจากผู้เชี่ยวชาญและกระบวนการที่ชัดเจนและเป็นระบบ มหาวิทยาลัยสามารถกำหนดเกณฑ์มาตรฐานที่น่าเชื่อถือสำหรับการวัดความสามารถทางภาษาอังกฤษได้อย่างแม่นยำ เกณฑ์เหล่านี้ช่วยสนับสนุนความสำเร็จของผู้เรียน พร้อมทั้งรักษามาตรฐานทางวิชาการที่สูง ในขณะเดียวกันยังช่วยให้มหาวิทยาลัยสามารถแข่งขันในยุคที่มีการเปลี่ยนแปลงและความท้าทายเพิ่มขึ้นได้ การให้ความสำคัญกับกระบวนการกำหนดคะแนนจุดตัดที่ยุติธรรมและโปร่งใสจึงเป็นสิ่งจำเป็น เพื่อสร้างโอกาสที่เท่าเทียมและสนับสนุนให้ผู้เรียนทุกคนประสบความสำเร็จ

5.3 ข้อเสนอแนะ

ข้อเสนอแนะจากการวิจัย แบ่งออกเป็น 2 ส่วน ได้แก่ ข้อเสนอแนะในการนำผลการวิจัยไปใช้ และข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้ มี 4 ข้อ ดังนี้

1.1 ผลการศึกษา พบว่า ข้อสอบระดับ A2 ในทักษะการฟังและโครงสร้างภาษา ไม่ผ่านเกณฑ์ความลึกของความรู้ของ Webb (1997) ซึ่งบ่งชี้ว่าอาจมีระดับความซับซ้อนไม่เพียงพอและไม่สะท้อนทักษะของผู้สอบได้อย่างถูกต้อง ดังนั้น ควรปรับปรุงข้อสอบให้เหมาะสมกับระดับ A2 สำหรับทักษะการฟังควรเน้นความเข้าใจและใช้ภาษาในบริบทจริง เช่น การฟังประกาศสั้น ๆ หรือการติดตามเรื่องราวต่าง ๆ ในข่าว ซึ่งต้องอาศัยทักษะที่สูงกว่าความจำ แต่ยังไม่ถึงระดับการคิดวิเคราะห์เชิงลึก ขณะที่ทักษะโครงสร้างภาษาควรมุ่งเน้นการใช้ไวยากรณ์พื้นฐานเพื่อการสื่อสารในชีวิตประจำวัน เช่น การเล่าเรื่อง การตั้งคำถาม การให้คำแนะนำ ซึ่งต้องการความเข้าใจมากกว่าความจำเพียงอย่างเดียว แต่ยังไม่ถึงระดับการวิเคราะห์ ข้อสอบควรได้รับการออกแบบใหม่มุ่งเน้นการประยุกต์ใช้ไวยากรณ์ในบริบทจริง แทนการท่องจำกฎไวยากรณ์ เพื่อให้สามารถประเมินทักษะทางภาษาของผู้สอบระดับ A2 ได้อย่างแม่นยำและครอบคลุม การปรับปรุงข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบอาจช่วยให้แบบทดสอบ STOU-EPT สามารถประเมินทักษะทางภาษาได้สอดคล้องกับกรอบ CEFR

1.2 ผลการศึกษา พบว่า ข้อสอบ STOU-EPT สอดคล้องกับเนื้อหาหรือมาตรฐานกรอบ CEFR ทั้งสามทักษะ (การฟัง โครงสร้างภาษา และการอ่าน) และมีค่าดัชนีความสมดุล (Q) อยู่ในระดับที่ยอมรับได้ อย่างไรก็ตาม เพื่อให้การพัฒนาแบบทดสอบ STOU-EPT เป็นไปอย่างมีมาตรฐาน ควรมีการจัดทำผังการออกข้อสอบ (Test Blueprint) ซึ่งเป็นเครื่องมือสำคัญในการควบคุมความสมดุลของข้อสอบ ลดความคลาดเคลื่อนของระดับความสามารถ และกำหนดจำนวนข้อสอบที่เหมาะสมในแต่ละทักษะ นอกจากนี้ ผังการออกข้อสอบยังช่วยกำหนดระดับความยาก-ง่ายของข้อสอบให้สอดคล้องกับกรอบ CEFR ทำให้สามารถแยกแยะความสามารถของผู้สอบได้อย่างแม่นยำ ลดอคติในการออกข้อสอบ และส่งเสริมความเป็นธรรมในการสอบ การดำเนินการนี้จะช่วยให้แบบทดสอบ STOU-EPT มีความน่าเชื่อถือ เป็นระบบ และสามารถพัฒนาให้สอดคล้องกับมาตรฐานการสอบระดับสากลได้อย่างต่อเนื่อง

1.3 ผลการวิจัยชี้ให้เห็นถึงความสำคัญของการกำหนดเกณฑ์และแนวทางที่ชัดเจนสำหรับการออกแบบข้อสอบ STOU-EPT โดยเฉพาะในด้านความลึกของความรู้และระดับ CEFR เพื่อให้ข้อสอบสามารถประเมินทักษะของผู้สอบได้อย่างครอบคลุมและแม่นยำ ดังนั้น ควรมีการฝึกอบรมผู้พัฒนาข้อสอบให้มีความรู้ความเข้าใจอย่างลึกซึ้งเกี่ยวกับกรอบ CEFR และหลักการวัดผลที่

ถูกต้อง ทั้งนี้ เพื่อให้ผู้พัฒนาข้อสอบสามารถออกแบบข้อสอบได้อย่างมีประสิทธิภาพ และสอดคล้องกับมาตรฐานที่กำหนด

1.4 คณะแนจุดตัดที่กำหนดขึ้น เป็นคณะแนจุดตัดสำหรับแบบทดสอบ STOU-EPT ซึ่งพัฒนาโดยคณาจารย์แขนงวิชาภาษาอังกฤษ สาขาวิชาศิลปศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช ภายใต้โครงการปรับปรุงข้อสอบวัดความรู้ภาษาอังกฤษด้วยระบบคอมพิวเตอร์ STOU-EPT (E-Testing) ในปี พ.ศ. 2566 ซึ่งคณะแนจุดตัดนี้สามารถนำไปใช้กับแบบทดสอบ STOU-EPT ที่มีโครงสร้างข้อสอบและผังการออกข้อสอบเดียวกับแบบทดสอบที่พัฒนาขึ้นใหม่นี้ อย่างไรก็ตาม หากมีการปรับเปลี่ยนโครงสร้างข้อสอบหรือผังการสร้างข้อสอบในอนาคต ควรทำการกำหนดคณะแนจุดตัดใหม่ เพื่อให้การแปลผลสะท้อนความสามารถของผู้สอบได้อย่างถูกต้องและเหมาะสม

2. ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป มี 3 ข้อ ดังนี้

2.1 แบบทดสอบ STOU-EPT โดยเฉพาะในส่วนของทักษะการฟัง มีข้อสอบหลายรูปแบบ ได้แก่ แบบเลือกตอบ แบบจับคู่ แบบเติมคำ และแบบถูก-ผิด ดังนั้น ควรดำเนินการวิจัยเพื่อวิเคราะห์ความสอดคล้องของข้อสอบในแต่ละรูปแบบกับกรอบ CEFR โดยมุ่งเน้นการประเมินว่าข้อสอบแต่ละรูปแบบสามารถสะท้อนและวัดผลตามระดับความสามารถทางภาษา (A2 B1 และ B2) ได้อย่างแม่นยำและครอบคลุมเพียงใด รวมถึงศึกษาว่ารูปแบบคำถามแต่ละประเภทมีความเหมาะสมกับผู้สอบในระดับต่าง ๆ หรือไม่ เพื่อพัฒนาข้อสอบให้สามารถประเมินความสามารถทางภาษาของผู้สอบได้อย่างมีประสิทธิภาพและสอดคล้องกับเป้าหมายของการวัดผลทางภาษา

2.2 ควรมีการศึกษาเชิงคุณภาพกับผู้เชี่ยวชาญหลังจากกำหนดมาตรฐานเพื่อหาคณะแนจุดตัดด้วยวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง เพื่อตรวจสอบความตรงของกระบวนการกำหนดคณะแนจุดตัด เช่น การวิเคราะห์ความคิดเห็นของผู้เชี่ยวชาญเกี่ยวกับความยากของข้อสอบในแต่ละระดับ CEFR ซึ่งจะช่วยเพิ่มความลึกในการวิเคราะห์กระบวนการที่ใช้ และช่วยระบุจุดแข็งหรือข้อจำกัดวิธีการของแองกอฟท์ที่ได้รับการปรับปรุง อันจะนำผลที่ได้ไปพัฒนากระบวนการกำหนดคณะแนจุดตัดให้มีความแม่นยำและเหมาะสมมากยิ่งขึ้น

2.3 ควรนำคณะแนจุดตัดที่ได้จากการวิจัยนี้ไปทดสอบคณะแนจุดตัดกับกลุ่มตัวอย่างผู้สอบจริงที่มีความหลากหลายในระดับความสามารถทางภาษา เช่น A2 B1 และ B2 เพื่อวิเคราะห์ว่าคณะแนจุดตัดที่กำหนดขึ้นสามารถสะท้อนความสามารถทางภาษาในแต่ละระดับได้อย่างถูกต้องหรือไม่ ซึ่งจะช่วยยืนยันความถูกต้องของคณะแนจุดตัดและตรวจสอบว่าข้อสอบสามารถวัดได้ตรงตามระดับที่กำหนดไว้