

การตรวจสอบคุณสมบัติทางจิตมิติของข้อสอบวัดความรู้ภาษาอังกฤษของ  
มหาวิทยาลัยสุโขทัยธรรมมาธิราช

Evaluation of the Psychometric Properties of Sukhothai Thammathirat Open  
University's English Proficiency Test

ศุกลรัตน์ อิงชาติเจริญ<sup>1\*</sup>

Sukolrat Ingchatcharoen<sup>1\*</sup>

(Received: October 17, 2023; Revised: November 14, 2023; Accepted: November 16, 2023)

บทคัดย่อ

การวิจัยนี้มีจุดมุ่งหมายเพื่อ 1) วิเคราะห์คุณภาพของข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ (STOU-EPT) ใน 3 พารามิเตอร์ ได้แก่ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าโอกาสการเดา (c) ตามทฤษฎีการตอบสนองข้อสอบ (IRT) และ 2) ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ กลุ่มตัวอย่าง ได้แก่ ผู้ที่เข้าสอบ STOU-EPT ใช้วิธีการเลือกแบบเจาะจงตามเกณฑ์ คือ มีข้อมูลทุติยภูมิผลการสอบ STOU-EPT ร่วมกับมีข้อมูลเพศ ภูมิลำเนา และประสบการณ์ในการสอบ จำนวน 1,272 คน เครื่องมือวิจัยเป็นแบบทดสอบ STOU-EPT จำนวน 100 ข้อ ประกอบด้วย ทักษะการฟัง 25 ข้อ ทักษะโครงสร้างไวยากรณ์ 35 ข้อ และทักษะการอ่าน 40 ข้อ วิเคราะห์ข้อมูลด้วย RStudio ผลการวิจัยพบว่า

1. คุณภาพข้อสอบ STOU-EPT ได้แก่ 1) ด้านทักษะการฟัง ค่า a อยู่ในช่วง -9.62 ถึง 4.50 ซึ่งผ่านเกณฑ์  $a > 0$  จำนวน 22 ข้อ (88%) ค่า b อยู่ในช่วง -1.80 ถึง 8.19 ซึ่งผ่านเกณฑ์  $-2 \leq b \leq +2$  จำนวน 22 ข้อ (88%) และค่า c อยู่ในช่วง 0 ถึง 0.61 2) ด้านทักษะโครงสร้างไวยากรณ์ ค่า a อยู่ในช่วง -3.40 ถึง 7.65 ซึ่งผ่านเกณฑ์  $a > 0$  จำนวน 31 ข้อ (88.57%) ค่า b อยู่ในช่วง -4.21 ถึง 3.06 ซึ่งผ่านเกณฑ์  $-2 \leq b \leq +2$  จำนวน 22 ข้อ (62.86%) และค่า c อยู่ในช่วง 0 ถึง 0.44 และ 3) ด้านทักษะการอ่าน ค่า a อยู่ในช่วง -8.75 ถึง 7.87 ซึ่งผ่านเกณฑ์  $a > 0$  จำนวน 36 ข้อ (90%) ค่า b อยู่ในช่วง -2.02 ถึง 23.75 ซึ่งผ่านเกณฑ์  $-2 \leq b \leq +2$  จำนวน 35 ข้อ (87.50%) และค่า c อยู่ในช่วง -0 ถึง 0.63

<sup>1</sup> ผู้ช่วยศาสตราจารย์ สาขาวิชาสถิติ วิจัย การวัด และการประเมินผลการศึกษา สำนักทะเบียนและวัดผล  
มหาวิทยาลัยสุโขทัยธรรมมาธิราช

<sup>1</sup> Assistant Professor, Education Statistics, Research, Measurement and Evaluation, Office of  
Registration, Records and Evaluation, Sukhothai Thammathirat Open University  
งานวิจัยนี้ได้รับทุนวิจัยจากมหาวิทยาลัยสุโขทัยธรรมมาธิราช ปีงบประมาณ 2565

This research was funded by Sukhothai Thammathirat Open University, Fiscal Year 2022.

\* Corresponding Author E-mail: sukolrating@gmail.com

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ STOU-EPT ได้แก่ วิธี LR, SIBTEST และ MH โดยเปรียบเทียบจำนวนข้อที่พบ DIF ตรงกันตั้งแต่ 2 วิธีขึ้นไป พบว่า 1) ด้านทักษะการฟัง มี DIF ระหว่างเพศ (หญิง = focal) ภูมิสำเนา (กทม.และปริณทล = focal) และประสบการณ์ในการสอบ (> 1 ครั้ง = focal) จำนวนอย่างละ 3 ข้อเท่ากัน (12%) 2) ด้านทักษะโครงสร้างไวยากรณ์ มี DIF ระหว่างเพศ ภูมิสำเนา และประสบการณ์ในการสอบ จำนวน 3, 5 และ 1 ข้อ ตามลำดับ (8.57%, 14.29%, และ 2.86%) และ 3) ด้านทักษะการอ่าน มี DIF ระหว่างเพศ ภูมิสำเนา และประสบการณ์ในการสอบ จำนวน 1, 3 และ 0 ข้อ ตามลำดับ (2.50%, 7.50%, และ 0%)

**คำสำคัญ** คุณสมบัติทางจิตมิติ การประมาณค่าพารามิเตอร์ของข้อสอบ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ข้อสอบวัดความรู้ภาษาอังกฤษ

### Abstract:

The objectives of this research were to 1) analyze the quality of Sukhothai Thammathirat Open University's English proficiency test (STOU-EPT) using a computer system in three parameters: the discrimination parameter (a), the item difficulty parameter (b), and the guessing parameter (c) following Item Response Theory (IRT); and 2) test differential item functioning (DIF) in STOU-EPT. The sample was 1,272 persons who took STOU-EPT, using purposive sampling in compliance with the criteria of having secondary data, as well as data on gender, dwelling place, and experience in taking STOU-EPT. The research instrument was the 100-question STOU-EPT which consists of 25 listening skills questions, 35 grammar skills questions, and 40 reading skills questions. The data was examined using RStudio. The study's findings showed that:

1. The quality of STOU-EPT included: 1) for listening skills, the a-parameter ranged from -9.62 to 4.5, with 22 items (88%) passing the criterion  $a > 0$ , the b-parameter ranged from -1.80 to 8.19, with 22 items (88%) passing the criterion  $-2 \leq b \leq +2$ , and the c-parameter ranged from 0 to 0.61; 2) regarding grammar skills, the a-parameter ranged from -3.40 to 7.65, with 31 items (88.57%) passing the criterion  $a > 0$ , the b-parameter ranged from -4.21 to 3.06, with 22 items (62.86%) passing the criterion  $-2 \leq b \leq +2$ , and the c-parameter ranged from 0 to 0.44; and 3) reading skills, the a-parameter ranged from -8.75 to 7.87, with 36 items (90%) passing the criterion  $a > 0$ , the b-parameter ranged from -2.02 to 23.75, with 35 items (87.50%) passing the criterion  $-2 \leq b \leq +2$ , and the c-parameter ranged from -0 to 0.63.

2. Testing the DIF in STOU-EPT using LR, SIBTEST, and MH, by comparing the number of items that found the DIF matched from two methods and over, revealed that 1) listening

skills had the DIF between gender (female = focal), dwelling place (Bangkok and surrounding areas = focal), and experience in taking the exam ( $> 1$  time = focal), with the same number of 3 items each (1.2%); 2) grammar skills had the DIF between gender, dwelling place, and experience in taking the exam, with 3, 5, and 1 items, respectively (8.57%, 14.29%, and 2.86%), and 3) reading skills had the DIF between gender, dwelling place, and experience in taking the exam, with 1, 3, and 0 items, respectively (2.50%, 7.50%, and 0%).

**Keywords:** psychometric properties, item parameter estimation, differential item functioning, English proficiency test

## บทนำ

การสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ STOU-EPT (E-Testing) ได้ถูกพัฒนาขึ้นโดยมหาวิทยาลัยสุโขทัยธรรมมาธิราชโดยได้กำหนดแนวทางยกระดับมาตรฐานภาษาอังกฤษของนักศึกษาระดับปริญญาตรีเพื่อให้สอดคล้องกับนโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษาของสำนักงานคณะกรรมการการอุดมศึกษา พ.ศ. 2559 ซึ่งคณาจารย์สาขาวิชาศิลปศาสตร์เป็นผู้ดำเนินการออกข้อสอบและทางสำนักทะเบียนและวัดผลเป็นผู้ดำเนินการพัฒนาระบบการจัดสอบจัดสอบด้วยระบบคอมพิวเตอร์ในครั้งแรกเมื่อ พ.ศ. 2560 โดยแบบทดสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ได้จัดสร้างขึ้นตามมาตรฐานการประเมินความสามารถทางภาษาที่สหภาพยุโรปจัดทำขึ้นเพื่อเป็นแนวทางในการจัดการเรียนรู้ การสอน และการประเมินภาษาที่สองหรือภาษาต่างประเทศซึ่งเรียกว่า กรอบอ้างอิงทางภาษาของสหภาพยุโรป (The Common European Framework of Reference for Languages - CEFR) (Council of Europe, 2001) ลักษณะของแบบวัดเป็นรูปแบบมัลติมีเดียและสามารถประมวลผลคะแนนได้ด้วยระบบคอมพิวเตอร์ ทำให้ผู้เข้าสอบสามารถทราบผลได้ทันทีหลังสอบเสร็จ โดยโครงสร้างของข้อสอบมีจำนวน 100 ข้อ (100 คะแนน) แบ่งเป็น 3 ด้าน ทักษะการฟัง (listening) จำนวน 25 ข้อ ทักษะโครงสร้างไวยากรณ์ (structure) จำนวน 35 ข้อ และทักษะการอ่าน (reading) จำนวน 40 ข้อ

การทดสอบเพื่อวัดทักษะทางภาษาอังกฤษเป็นการวัดเพื่อให้ได้สารสนเทศในการนำไปใช้ประโยชน์อันได้แก่ การตรวจสอบความพร้อมก่อนเริ่มเรียน การประเมินความก้าวหน้า ตลอดจนการประเมินเพื่อตัดสินผล โดยผู้สอบสามารถนำผลคะแนนไปใช้เพื่อพัฒนาตนเอง เพื่อสมัครเข้าศึกษาต่อ เพื่อสำเร็จการศึกษา หรือเพื่อสมัครงานตามข้อกำหนดของแต่ละสถานศึกษาหรือหน่วยงาน ซึ่งการทดสอบนี้จัดว่ามีผลกระทบสูงต่อผู้สอบและผู้ที่เกี่ยวข้อง (high-stakes test) ดังนั้นกระบวนการสร้างแบบทดสอบและเกณฑ์การตัดสินผลคะแนนจึงมีความสำคัญ โดยแบบทดสอบจะต้องมีความเที่ยง ความตรง และไม่มีความลำเอียง (bias) ไปยังผู้สอบกลุ่มใดกลุ่มหนึ่ง ซึ่งที่ผ่านมาแบบทดสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ได้ผ่านการตรวจสอบคุณภาพโดยมีงานวิจัยรองรับ คือ 1) การตรวจสอบคุณภาพข้อสอบในด้าน

กระบวนการสร้างแบบทดสอบตามเกณฑ์ของ CEFR 2) การพิจารณาโครงสร้างข้อสอบโดยผู้เชี่ยวชาญจากสถาบันภาษาจุฬาลงกรณ์มหาวิทยาลัย จำนวน 4 ท่าน โดยใช้โปรแกรม Lexitutor และโปรแกรม AWL list 3) การวิเคราะห์คุณภาพข้อสอบค่าความยากง่ายและความเที่ยงตามแนวทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) (Jiraro & Aungsuchoti, 2021) แต่ทั้งนี้ยังไม่มีการตรวจสอบคุณภาพข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ที่สามารถให้ทั้งสารสนเทศของข้อสอบเป็นรายข้อ (item parameter) โดยโมเดลแบบสามพารามิเตอร์ (Three-Parameter Logistic Model) จะสามารถประมาณค่าพารามิเตอร์ของข้อสอบได้ทั้งค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และค่าโอกาสการเดาของข้อสอบ (c) ซึ่งค่าพารามิเตอร์ของข้อสอบนี้จะไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบที่มีความสามารถต่างกันที่เป็นจุดอ่อนของการวิเคราะห์ข้อสอบแบบทฤษฎีการทดสอบแบบดั้งเดิม นอกจากนี้ยังสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) หรือความลำเอียงของข้อสอบ (item bias) ซึ่งเป็นการตรวจสอบด้านความยุติธรรมของข้อสอบที่จะเกิดขึ้นเมื่อกลุ่มต่าง ๆ ที่มีความสามารถเท่าเทียมกันแต่มีโอกาของการตอบข้อสอบได้ถูกต้องแตกต่างกัน หรือมีความได้เปรียบหรือเสียเปรียบทั้งที่มีความสามารถเท่า ๆ กัน โดยข้อสอบนั้นอาจมีความลำเอียงเข้าข้างกลุ่มใดกลุ่มหนึ่ง (Kanjawasee, 2020) ดังตัวอย่างในงานวิจัยที่ศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความรู้ภาษาอังกฤษที่พบว่ามีความลำเอียงในด้านต่าง ๆ เช่น ความลำเอียงด้านเพศ (Geramipour & Shahmirzadi, 2019; Liu, 2017; Council of Europe, 2001; Ryan & Bachman, 1992) ความลำเอียงด้านภูมิลำเนา (Liu, 2017; Council of Europe, 2001; Ryan & Bachman, 1992) และความลำเอียงด้านประสบการณ์ในการสอบ (Liao & Yao, 2021; Liu, 2017; Council of Europe, 2001) เป็นต้น

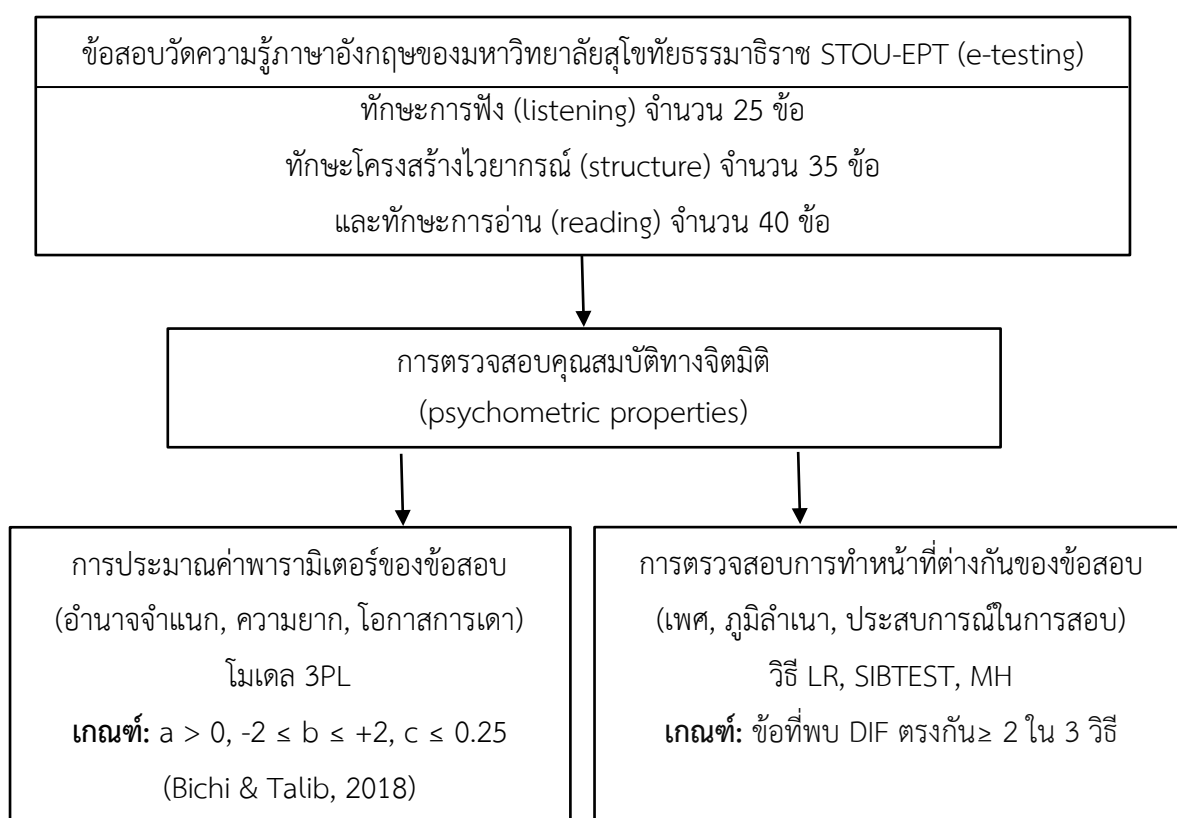
ดังนั้นผู้วิจัยจึงต้องการวิเคราะห์คุณภาพของข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราชด้วยระบบคอมพิวเตอร์ โดยใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อวิเคราะห์คุณภาพของข้อสอบใน 3 พารามิเตอร์ ได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และค่าโอกาสการเดาของข้อสอบ (c) และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธี logistic regression (LR), Simultaneous Item Bias Test (SIBTEST), และ Mantel-Haenszel statistic (MH) เนื่องจากเป็นวิธีที่นิยมใช้มากและการเลือกใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่แตกต่างกันก็จะให้ผลการวิเคราะห์ที่ต่างกัน (Berrío et al., 2020) จึงควรใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มากกว่า 1 วิธี แล้วพิจารณาตัดสินความสอดคล้องกันของผลการวิเคราะห์ เพื่อให้ได้สารสนเทศในการนำไปใช้ปรับปรุงและพัฒนาข้อสอบทำให้ข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราชด้วยระบบคอมพิวเตอร์มีคุณภาพและมีมาตรฐานเป็นที่ยอมรับ และผลจากการวิจัยนี้สามารถใช้เป็นหลักฐานเชิงประจักษ์สำหรับคุณภาพของข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราช

### ความมุ่งหมายของการวิจัย

1. เพื่อวิเคราะห์คุณภาพของข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ ใน 3 พารามิเตอร์ ได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และค่าโอกาสการเดาของข้อสอบ (c) ตามหลักการของทฤษฎีการตอบสนองข้อสอบ (IRT)
2. เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์โดยใช้วิธี logistic regression (LR), Simultaneous Item Bias Test (SIBTEST), และ Mantel-Haenszel statistic (MH)

### กรอบแนวคิดในการวิจัย

กรอบแนวคิดในการวิจัยเพื่อตรวจสอบคุณสมบัติทางจิตมิติของข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชมีดังนี้



### วิธีดำเนินการวิจัย

#### 1. ประชากรและกลุ่มตัวอย่าง

ประชากร ได้แก่ ผู้ที่เข้าสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ (STOU-EPT)

กลุ่มตัวอย่าง ได้แก่ ผู้ที่เข้าสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ (STOU-EPT) คัดเลือกตัวอย่างด้วยวิธีการเลือกแบบเจาะจง (purposive sampling) ตามเกณฑ์ คือ มีข้อมูลทุติยภูมิ (secondary data) ที่เป็นผลการสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ในช่วงระหว่างวันที่ 13 พฤศจิกายน 2554 ถึง 24 เมษายน 2555 ร่วมกับมีข้อมูลเพศ ภูมิลำเนา และประสบการณ์ในการสอบ รวมจำนวน 1,272 คน

## 2. เครื่องมือที่ใช้ในการวิจัย

แบบทดสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ จัดสร้างขึ้นเพื่อมุ่งวัดความพร้อม ผลสัมฤทธิ์ หรือพัฒนาการตามมาตรฐานการประเมินความสามารถทางภาษา ที่สหภาพยุโรปจัดทำขึ้น ซึ่งเรียกว่ากรอบอ้างอิงทางภาษาของสหภาพยุโรป (The Common European Framework of Reference for Languages - CEFR) ลักษณะของแบบวัดเป็นรูปแบบมัลติมีเดียและสามารถประมวลผลคะแนนได้ด้วยระบบคอมพิวเตอร์ทำให้ทราบผลได้ทันทีหลังสอบเสร็จ โดยโครงสร้างของข้อสอบ มีจำนวน 100 ข้อ (100 คะแนน) แบ่งเป็น 3 ด้าน ทักษะการฟัง (listening) จำนวน 25 ข้อ ซึ่งมีรูปแบบข้อสอบ ได้แก่ แบบ 4 ตัวเลือก จำนวน 8 ข้อ แบบจับคู่ จำนวน 12 ข้อ และแบบถูกผิด จำนวน 5 ข้อ ส่วนทักษะโครงสร้างไวยากรณ์ (structure) จำนวน 35 ข้อ และทักษะการอ่าน (reading) จำนวน 40 ข้อ จะมีรูปแบบข้อสอบเป็นแบบตัวเลือก 4 ตัวเลือก มีการตรวจสอบคุณภาพข้อสอบโดยผู้เชี่ยวชาญจากสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย จำนวน 4 ท่าน โดยใช้โปรแกรม Lexitutor ในการตรวจสอบความเหมาะสมของระดับความสามารถทางภาษา (A2-B2) และโปรแกรม AWL list เพื่อตรวจสอบความเหมาะสมของการใช้คำศัพท์ทางวิชาการของข้อสอบ และมีผลการวิเคราะห์คุณภาพข้อสอบค่าความยากง่ายอยู่ในเกณฑ์ใช้ได้ ค่า  $p$  อยู่ระหว่าง 0.20 – 0.80 และมีค่าสัมประสิทธิ์ความเที่ยง KR-20 อยู่ระหว่าง 0.35 ถึง 0.89 โดยทักษะการอ่าน มีความเที่ยงสูงสุดอยู่ระหว่าง 0.85 ถึง 0.893 ส่วนทักษะการฟังมีความเที่ยงต่ำสุดเท่ากับ 0.35 (Jiraro & Aungsuchoti, 2021)

## 3. การเก็บรวบรวมข้อมูล

3.1 ผู้วิจัยเสนอโครงการวิจัยเพื่อขอรับการพิจารณาจริยธรรมการวิจัยจากมหาวิทยาลัยราชภัฏ นครราชสีมา (เลขที่ใบรับรอง HE-143-2565)

3.2 ผู้วิจัยทำหนังสือขอความอนุเคราะห์ในการเก็บรวบรวมข้อมูลงานวิจัยส่งไปยังอธิการบดี มหาวิทยาลัยสุโขทัยธรรมมาธิราช

3.3 ผู้วิจัยรวบรวมข้อมูลผลการสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ (STOU-EPT) โดยคัดเลือกตัวอย่างด้วยวิธีการเลือกแบบเจาะจง (purposive sampling) ตามเกณฑ์ คือ ข้อมูลทุติยภูมิ (secondary data) ที่เป็นผลการสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ ซึ่งใช้แบบทดสอบฉบับปรับปรุง ฉบับที่ 1 เก็บข้อมูลในช่วงเวลาเดือนพฤศจิกายน 2564 ถึงเมษายน 2565 โดยมีข้อมูลเพศ ภูมิลำเนา และประสบการณ์ในการสอบ รวมด้วย

#### 4. การวิเคราะห์ข้อมูล

4.1 วิเคราะห์ข้อมูลเบื้องต้นตามตัวแปรที่ศึกษา โดยใช้สถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าต่ำสุด ค่าสูงสุด ความถี่ ร้อยละ ด้วยโปรแกรม RStudio

4.2 ตรวจสอบข้อตกลงเบื้องต้นของ IRT ได้แก่ ความเป็นเอกมิติ (unidimensionality) และความเป็นอิสระจากกัน (local independence) ด้วยโปรแกรม RStudio Package 'irtGUI'

4.3 วิเคราะห์คุณภาพข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ใน 3 พารามิเตอร์ ได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และค่าโอกาสการเดาของข้อสอบ (c) ตามหลักการของทฤษฎีการตอบสนองข้อสอบ (IRT) ด้วยโปรแกรม RStudio Package 'irtGUI'

4.4 ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ตามตัวแปรที่ศึกษาจำนวน 3 ตัวแปร ได้แก่ เพศ ภูมิภาค และประสบการณ์ในการสอบ ด้วยโปรแกรม RStudio Package 'difR'

#### ผลการวิจัย

1. ผลการวิเคราะห์ข้อมูลเบื้องต้นตามตัวแปรที่ศึกษา พบว่า กลุ่มตัวอย่างส่วนใหญ่เป็นเพศหญิงจำนวน 822 คน คิดเป็นร้อยละ 64.62 มีภูมิลำเนาอยู่จังหวัดกรุงเทพฯ และปริมณฑล จำนวน 1,272 คน คิดเป็นร้อยละ 50.79 และมีประสบการณ์ในการสอบเป็นครั้งแรก จำนวน 1,120 คน คิดเป็นร้อยละ 88.05

2. ผลการตรวจสอบข้อตกลงเบื้องต้นของ IRT ด้วยโปรแกรม R โดยจำแนกข้อมูลเป็น 3 ชุด ได้แก่ ทักษะการฟัง (listening) จำนวน 25 ข้อ ทักษะโครงสร้างไวยากรณ์ (structure) จำนวน 35 ข้อ และทักษะการอ่าน (reading) จำนวน 40 ข้อ ผลการตรวจสอบ พบว่า ทั้ง 3 ชุดข้อมูลต่างก็มีความเป็นเอกมิติ กล่าวคือแต่ละชุดข้อมูลมีจำนวนเพียง 1 องค์กรประกอบที่มีค่าไอเกน (eigen value) มากกว่า 1 ส่วนการตรวจสอบความเป็นอิสระจากกัน (local independence) พบว่า ทั้ง 3 ชุดข้อมูลมีค่า residual correlation ไม่เกิน 0.20 แสดงว่า ไม่ละเมิดข้อตกลงด้านความเป็นอิสระจากกัน และผลการตรวจสอบความสอดคล้องของโมเดลแบบสามพารามิเตอร์ (3PL) กับข้อมูล (model fit) จากค่า log likelihood ได้ค่าเท่ากับ -17802.39, -24904.56, -30215.29 ตามลำดับ

3. ผลการวิเคราะห์คุณภาพข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) ในโมเดลแบบสามพารามิเตอร์ (3PL) ได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) โดยใช้เกณฑ์  $a > 0$  ค่าความยากของข้อสอบ (b) โดยใช้เกณฑ์  $-2 \leq b \leq +2$  และค่าโอกาสการเดาของข้อสอบ (c) ในกรณีรูปแบบข้อสอบเป็นแบบตัวเลือก 4 ตัวเลือกจะใช้เกณฑ์ค่า c ไม่ควรเกิน 0.25 (Bichi & Talib, 2018)

3.1 ทักษะการฟัง (listening) จำนวน 25 ข้อ พบว่า ค่าอำนาจจำแนกของข้อสอบ (a) อยู่ในช่วง -9.62 ถึง 4.50 มีข้อที่ผ่านเกณฑ์  $a > 0$  จำนวน 22 ข้อ ค่าความยากของข้อสอบ (b) อยู่ในช่วง

-1.80 ถึง 8.19 มีข้อที่ผ่านเกณฑ์  $-2 \leq b \leq +2$  จำนวน 22 ข้อ และค่าโอกาสการเดาของข้อสอบ (c) อยู่ในช่วง 0 ถึง 0.61 (ข้อสอบเป็นแบบตัวเลือก 3 ตัวเลือก แบบจับคู่ และแบบถูกผิด)

3.2 ทักษะโครงสร้างไวยากรณ์ (structure) จำนวน 35 ข้อ พบว่า ค่าอำนาจจำแนกของข้อสอบ (a) อยู่ในช่วง -3.40 ถึง 7.65 มีข้อที่ผ่านเกณฑ์  $a > 0$  จำนวน 31 ข้อ ค่าความยากของข้อสอบ (b) อยู่ในช่วง -4.21 ถึง 3.06 มีข้อที่ผ่านเกณฑ์  $-2 \leq b \leq +2$  จำนวน 22 ข้อ และค่าโอกาสการเดาของข้อสอบ (c) อยู่ในช่วง 0 ถึง 0.44 มีข้อที่ผ่านเกณฑ์  $c \leq 0.25$  จำนวน 24 ข้อ

3.3 ทักษะการอ่าน (reading) จำนวน 40 ข้อ พบว่า ค่าอำนาจจำแนกของข้อสอบ (a) อยู่ในช่วง -8.75 ถึง 7.87 มีข้อที่ผ่านเกณฑ์  $a > 0$  จำนวน 36 ข้อ ค่าความยากของข้อสอบ (b) อยู่ในช่วง -2.02 ถึง 23.75 มีข้อที่ผ่านเกณฑ์  $-2 \leq b \leq +2$  จำนวน 35 ข้อ และค่าโอกาสการเดาของข้อสอบ (c) อยู่ในช่วง 0 ถึง 0.63 มีข้อที่ผ่านเกณฑ์  $c \leq 0.25$  จำนวน 25 ข้อ รายละเอียดดังตาราง 1

**ตาราง 1** ผลการวิเคราะห์คุณภาพข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) ในโมเดลแบบสามพารามิเตอร์ (3PL)

| ค่าพารามิเตอร์ 3PL |                      | ข้อสอบวัดความรู้ภาษาอังกฤษ STOU-EPT              |                                                                                    |                                                                         |
|--------------------|----------------------|--------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
|                    |                      | ทักษะการฟัง                                      | ทักษะโครงสร้าง                                                                     | ทักษะการอ่าน                                                            |
|                    |                      | 25 ข้อ<br>(ข้อที่ 1-25)                          | ไวยากรณ์ 35 ข้อ<br>(ข้อที่ 26-60)                                                  | 40 ข้อ<br>(ข้อที่ 61-100)                                               |
| ช่วงค่า a          |                      | -9.62 ถึง 4.50                                   | -3.40 ถึง 7.65                                                                     | -8.75 ถึง 7.87                                                          |
| a > 2.00           | Quite satisfactory   | 4, 5, 7, 8, 10,<br>14, 16, 18, 19,<br>20, 22, 23 | 2, 5, 6, 7, 8, 12, 14,<br>17, 18, 20, 21, 22,<br>23, 24, 25, 26, 27,<br>28, 32, 34 | 1, 2, 3, 4, 5, 14, 15,<br>16, 17, 18, 21, 23,<br>24, 27, 29, 34, 37, 39 |
| 1.35 ≤ a ≤ 2.00    | Good item            | 2, 3, 6, 9, 12, 15                               | 3, 9, 10, 15, 31                                                                   | 6, 10, 11, 25, 26, 32,<br>33, 36                                        |
| 0.65 ≤ a ≤ 1.34    | Moderate             | 11                                               | 4, 11, 13, 16, 19                                                                  | 8, 9, 13, 19, 22, 30                                                    |
| 0 ≤ a ≤ 0.64       | Marginal, revision   | 1, 13, 17                                        | 30                                                                                 | 12, 20, 28, 40                                                          |
| a < 0              | Poor item, revision  | 21, 24, 25                                       | 1, 29, 33, 35                                                                      | 7, 31, 35, 38                                                           |
| ช่วงค่า b          |                      | -1.80 ถึง 8.19                                   | -4.21 ถึง 3.06                                                                     | -2.02 ถึง 23.75                                                         |
| -3.00 < b < -2.00  | Very easy            | -                                                | 29, 33, 35                                                                         | 38                                                                      |
| -2.00 ≤ b < -1.00  | Easy                 | 21, 24, 25                                       | 1                                                                                  | 7, 31, 35                                                               |
| -1.00 ≤ b ≤ 1.00   | Moderately difficult | 1, 2, 3, 5, 8, 9,<br>10, 15, 18, 20              | 3, 5, 12, 14, 15, 19,<br>20, 30                                                    | 1, 2, 4, 5, 6, 10, 13,<br>14, 18, 19, 20, 22,<br>25, 30, 36             |
| 1.00 < b ≤ 2.00    | Difficult            | 4, 6, 7, 11, 12,<br>13, 14, 16, 19               | 4, 6, 7, 8, 9, 10, 16,<br>21, 22, 24, 28, 31, 32                                   | 3, 8, 9, 11, 15, 16,<br>17, 21, 23, 24, 26,<br>27, 29, 32, 33, 34, 39   |



ตาราง 1 (ต่อ)

| ค่าพารามิเตอร์ 3PL |                | ข้อสอบวัดความรู้ภาษาอังกฤษ STOU-EPT    |                                                     |                                                               |
|--------------------|----------------|----------------------------------------|-----------------------------------------------------|---------------------------------------------------------------|
|                    |                | ทักษะการฟัง<br>25 ข้อ<br>(ข้อที่ 1-25) | ทักษะโครงสร้าง<br>ไวยากรณ์ 35 ข้อ<br>(ข้อที่ 26-60) | ทักษะการอ่าน<br>40 ข้อ<br>(ข้อที่ 61-100)                     |
| b > 2.00           | Very difficult | 17, 22, 23                             | 2, 11, 13, 17, 18, 23,<br>25, 26, 27, 34            | 12, 28, 37, 40                                                |
| ช่วงค่า c          |                | 0 ถึง 0.61                             | 0 ถึง 0.44                                          | 0 ถึง 0.63                                                    |
| c > 0.25           |                | NA (ข้อสอบไม่ใช่<br>4 ตัวเลือก)        | 3, 4, 6, 8, 9, 14, 17,<br>22, 31, 32, 35            | 1, 3, 7, 10, 11, 13,<br>14, 15, 18, 25, 27,<br>29, 31, 35, 39 |

4. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ ตามตัวแปรที่ศึกษาจำนวน 3 ตัวแปร ได้แก่ เพศ ภูมิภาค และประสบการณ์ในการสอบ ด้วยการวิเคราะห์ใน 3 วิธี ได้แก่ LR (Logistic regression) SIBTEST (Simultaneous Item Bias Test) และ MH (Mantel-Haenszel) โดยแบ่งเป็นกลุ่มอ้างอิง (reference group) ซึ่งเป็นตัวแทนกลุ่มหลักในประชากร ที่คาดว่ากลุ่มนี้จะได้ประโยชน์จากการตอบข้อสอบหรือมีโอกาสในการตอบข้อสอบถูกต้องมากกว่าผู้เข้าสอบกลุ่มเปรียบเทียบ และกลุ่มเปรียบเทียบ (focal group) ซึ่งเป็นกลุ่มที่คาดว่าจะเสียประโยชน์จากการตอบข้อสอบ โดยเป็นกลุ่มที่สนใจจะศึกษาการทำหน้าที่ต่างกันของข้อสอบ ซึ่งมีโอกาสตอบข้อสอบถูกต้องได้น้อยกว่าผู้สอบกลุ่มอ้างอิง

4.1 ทักษะการฟัง (listening) จำนวน 25 ข้อ พบว่า มีจำนวนข้อที่พบ DIF ระหว่างเพศตรงกันตั้งแต่ 2 ใน 3 วิธีขึ้นไป (หญิง = focal) จำนวน 3 ข้อ ได้แก่ ข้อที่ 12, 15, 23 คิดเป็นร้อยละ 12 ของข้อสอบทั้งหมด โดยข้อที่ 12 และ 15 เข้าข้างเพศหญิงมากกว่าเพศชาย ส่วนข้อที่ 23 เข้าข้างเพศชายมากกว่าเพศหญิง ส่วนจำนวนข้อที่พบ DIF ระหว่างภูมิภาคตรงกันตั้งแต่ 2 ใน 3 วิธีขึ้นไป (กทม.และปริมณฑล = focal) มีจำนวน 3 ข้อ ได้แก่ ข้อที่ 7, 9, 21 คิดเป็นร้อยละ 12 ของข้อสอบทั้งหมด โดยข้อที่ 7 และ 9 เข้าข้างกลุ่มที่มีภูมิลำเนา กทม.และปริมณฑลมากกว่าภูมิภาคอื่น ส่วนข้อที่ 21 เข้าข้างกลุ่มที่มีภูมิลำเนาอื่นมากกว่าภูมิภาค กทม.และปริมณฑล และมีจำนวนข้อที่พบ DIF ระหว่างประสบการณ์ในการสอบ ตรงกันตั้งแต่ 2 ใน 3 วิธีขึ้นไป (> 1 ครั้ง = focal) จำนวน 3 ข้อ ได้แก่ ข้อที่ 9, 21, 24 คิดเป็นร้อยละ 12 ของข้อสอบทั้งหมด โดยข้อที่ 9 เข้าข้างกลุ่มผู้ที่มีประสบการณ์ในการสอบมากกว่าผู้ที่ไม่เคยมีประสบการณ์ในการสอบ ส่วนข้อที่ 21 และ 24 เข้าข้างผู้ที่ไม่เคยมีประสบการณ์ในการสอบมากกว่าผู้ที่มีประสบการณ์ในการสอบ

4.2 ทักษะโครงสร้างไวยากรณ์ (structure) จำนวน 35 ข้อ พบว่า มีจำนวนข้อที่พบ DIF ระหว่างเพศ ตรงกันตั้งแต่ 2 ใน 3 วิธีขึ้นไป (หญิง = focal) มีจำนวน 3 ข้อ ได้แก่ ข้อที่ 36, 42, 49 คิดเป็นร้อยละ 8.57 ของข้อสอบทั้งหมด โดยข้อที่ 36 และ 49 เข้าข้างเพศชายมากกว่าเพศหญิง ส่วนข้อที่ 42 เข้าข้างเพศหญิงมากกว่าเพศชาย ส่วนจำนวนข้อที่พบ DIF ระหว่างภูมิภาคตรงกันตั้งแต่ 2 ใน 3 วิธีขึ้นไป (กทม.และปริมณฑล = focal) มีจำนวน 5 ข้อ ได้แก่ ข้อที่ 33, 34, 39, 40, 53 คิดเป็นร้อยละ 14.29 ของข้อสอบทั้งหมด

โดยข้อที่ 33, 39, และ 53 เข้าข้างกลุ่มที่มีภูมิลำเนาหม.และปริมณฑลมากกว่าภูมิลำเนาอื่น ส่วนข้อที่ 34 และ 40 เข้าข้างกลุ่มที่มีภูมิลำเนาอื่นมากกว่าภูมิลำเนาหม.และปริมณฑล และมีจำนวนข้อที่พบ DIF ระหว่างประสบการณ์ในการสอบตรงกันตั้งแต่ 2 ใน 3 วิธีขึ้นไป ( $> 1$  ครั้ง = focal) จำนวน 1 ข้อ ได้แก่ ข้อที่ 27 คิดเป็นร้อยละ 2.86 ของข้อสอบทั้งหมด ซึ่งเข้าข้างผู้ที่มีประสบการณ์ในการสอบมากกว่าผู้ที่ไม่เคยมีประสบการณ์ในการสอบ

4.3 ทักษะการอ่าน (reading) จำนวน 40 ข้อ พบว่า มีจำนวนข้อที่พบ DIF ระหว่างเพศตรงกันตั้งแต่ 2 ใน 3 วิธีขึ้นไป (หญิง = focal) จำนวน 1 ข้อ ได้แก่ ข้อที่ 83 คิดเป็นร้อยละ 2.50 ของข้อสอบทั้งหมด โดยเข้าข้างเพศหญิงมากกว่าเพศชาย ส่วนจำนวนข้อที่พบ DIF ระหว่างภูมิลำเนาตรงกันตั้งแต่ 2 ใน 3 วิธีขึ้นไป (กทม.และปริมณฑล = focal) มีจำนวน 3 ข้อ ได้แก่ ข้อที่ 63, 73, 74 คิดเป็นร้อยละ 7.50 ของข้อสอบทั้งหมด โดยทั้ง 3 ข้อเข้าข้างกลุ่มที่มีภูมิลำเนาอื่นมากกว่าภูมิลำเนาหม.และปริมณฑล และมีจำนวนข้อที่พบ DIF ระหว่างประสบการณ์ในการสอบตรงกันตั้งแต่ 2 ใน 3 วิธีขึ้นไป ( $> 1$  ครั้ง = focal) จำนวน 0 ข้อ

**ตาราง 2** ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัย  
ธรรมาธิราชด้วยระบบคอมพิวเตอร์

| การทำหน้าที่ต่างกันของข้อสอบ                     |                               | ข้อสอบวัดความรู้ภาษาอังกฤษ STOU-EPT |                                   |                           |
|--------------------------------------------------|-------------------------------|-------------------------------------|-----------------------------------|---------------------------|
|                                                  |                               | ทักษะการฟัง                         | ทักษะโครงสร้าง                    | ทักษะการอ่าน              |
|                                                  |                               | 25 ข้อ<br>(ข้อที่ 1-25)             | ไวยากรณ์ 35 ข้อ<br>(ข้อที่ 26-60) | 40 ข้อ<br>(ข้อที่ 61-100) |
| DIF เพศ<br>(หญิง = focal)                        | LR                            | 12, 15, 23                          | 36, 42                            | 62, 63, 81, 83,<br>94, 96 |
|                                                  | SIBTEST                       | 12, 15, 23                          | 36, 41, 42, 49                    | 83                        |
|                                                  | MH                            | 4, 15, 23                           | 30, 36, 37, 42,<br>49, 53         | 83                        |
|                                                  | ข้อที่ DIF ตรงกัน 2 ใน 3 วิธี | 12, 23                              | 49                                | -                         |
|                                                  | ข้อที่ DIF ตรงกันทั้ง 3 วิธี  | 15                                  | 36, 42                            | 83                        |
|                                                  |                               |                                     |                                   |                           |
| DIF ภูมิลำเนา<br>(กทม.และปริมณฑล = focal)        | LR                            | 7, 9, 19, 21                        | 26, 40, 50, 52, 53                | 63, 73, 74, 84,<br>87, 96 |
|                                                  | SIBTEST                       | 9, 13, 14, 21                       | 33, 34, 39                        | 63, 73, 74, 76            |
|                                                  | MH                            | 7, 21                               | 33, 34, 39                        | 63, 73, 74, 85            |
|                                                  | ข้อที่ DIF ตรงกัน 2 ใน 3 วิธี | 7, 9                                | 33, 34, 39, 40, 53                | -                         |
|                                                  | ข้อที่ DIF ตรงกันทั้ง 3 วิธี  | 21                                  | -                                 | 63, 73, 74                |
|                                                  |                               |                                     |                                   |                           |
| DIF ประสบการณ์ในการสอบ<br>( $> 1$ ครั้ง = focal) | LR                            | 9, 21, 24                           | 26, 27                            | 62, 91                    |
|                                                  | SIBTEST                       | 9, 21, 24                           | 31, 51, 52, 59                    | 69, 78                    |
|                                                  | MH                            | 9, 21                               | 27                                | 68, 81                    |
|                                                  | ข้อที่ DIF ตรงกัน 2 ใน 3 วิธี | 24                                  | 27                                | -                         |
|                                                  | ข้อที่ DIF ตรงกันทั้ง 3 วิธี  | 9, 21                               | -                                 | -                         |
|                                                  |                               |                                     |                                   |                           |

## อภิปรายผล

1. ผลการวิเคราะห์คุณภาพข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) ในโมเดลแบบสามพารามิเตอร์ (3PL) พบว่า 1) ด้านทักษะการฟัง มีค่าอำนาจจำแนก (a) อยู่ในช่วง -9.62 ถึง 4.5 มีข้อที่ผ่านเกณฑ์  $a > 0$  จำนวน 22 ข้อ คิดเป็นร้อยละ 88 ค่าความยาก (b) อยู่ในช่วง -1.80 ถึง 8.19 ซึ่งผ่านเกณฑ์  $-2 \leq b \leq +2$  จำนวน 22 ข้อ คิดเป็นร้อยละ 88 และค่าโอกาสการเดา (c) อยู่ในช่วง 0 ถึง 0.61 (ข้อสอบเป็นแบบตัวเลือก 3 ตัวเลือก แบบจับคู่ และแบบถูกผิด) 2) ด้านทักษะโครงสร้างไวยากรณ์ จำนวน 35 ข้อ มีค่าอำนาจจำแนก (a) อยู่ในช่วง -3.40 ถึง 7.65 ซึ่งผ่านเกณฑ์  $a > 0$  จำนวน 31 ข้อ คิดเป็นร้อยละ 88.57 ค่าความยาก (b) อยู่ในช่วง -4.21 ถึง 3.06 ซึ่งผ่านเกณฑ์  $-2 \leq b \leq +2$  จำนวน 22 ข้อ คิดเป็นร้อยละ 62.86 และค่าโอกาสการเดา อยู่ในช่วง 0 ถึง 0.44 ซึ่งผ่านเกณฑ์  $c \leq 0.25$  จำนวน 24 ข้อ คิดเป็นร้อยละ 68.57 3) ด้านทักษะการอ่าน จำนวน 40 ข้อ มีค่าอำนาจจำแนก (a) อยู่ในช่วง -8.75 ถึง 7.87 ซึ่งผ่านเกณฑ์  $a > 0$  จำนวน 36 ข้อ คิดเป็นร้อยละ 90 ค่าความยากของข้อสอบ (b) อยู่ในช่วง -2.02 ถึง 23.75 ซึ่งผ่านเกณฑ์  $-2 \leq b \leq +2$  จำนวน 35 ข้อ คิดเป็นร้อยละ 87.50 และค่าโอกาสการเดาของข้อสอบ (c) อยู่ในช่วง -0 ถึง 0.63 ซึ่งผ่านเกณฑ์  $c \leq 0.25$  จำนวน 25 ข้อ คิดเป็นร้อยละ 62.50 ซึ่งสอดคล้องใกล้เคียงกับงานวิจัยของ Jiraro & Aungsuchoti (2021) ที่วิเคราะห์คุณภาพข้อสอบค่าความยากง่ายและความเที่ยงตามแนวทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ของแบบทดสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ซึ่งเป็นแบบทดสอบชุดเดียวกับงานวิจัยนี้ จากกลุ่มตัวอย่างเป็นบุคคลทั่วไปจำนวน 800 คน ผลการวิเคราะห์พบว่า ด้านทักษะการฟังมีข้อสอบที่ค่อนข้างยากจำนวน 12 ข้อ ยากง่ายพอเหมาะจำนวน 8 ข้อ ค่อนข้างง่ายจำนวน 5 ข้อ และมีข้อสอบที่อยู่ในเกณฑ์ใช้ได้ 25 ข้อ (100%) ด้านทักษะโครงสร้างไวยากรณ์มีข้อสอบที่ยากมากจำนวน 3 ข้อ ค่อนข้างยากจำนวน 11 ข้อ ยากง่ายพอเหมาะจำนวน 16 ข้อ ค่อนข้างง่ายจำนวน 5 ข้อ และมีข้อสอบที่อยู่ในเกณฑ์ใช้ได้ 32 ข้อ (91.43%) ด้านทักษะการอ่านข้อสอบที่ยากมากจำนวน 2 ข้อ ค่อนข้างยากจำนวน 13 ข้อ ยากง่ายพอเหมาะจำนวน 20 ข้อ ค่อนข้างง่ายจำนวน 5 ข้อ และมีข้อสอบที่อยู่ในเกณฑ์ใช้ได้ 38 ข้อ (95%) ซึ่งแสดงว่าผลการตรวจสอบคุณภาพของแบบทดสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบได้ผลการวิเคราะห์คุณภาพด้านความยากง่ายของข้อสอบที่สอดคล้องกันใกล้เคียงกัน แต่การวิเคราะห์คุณภาพข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบจะสามารถให้สารสนเทศได้มากกว่า กล่าวคือ สามารถวิเคราะห์พารามิเตอร์ค่าโอกาสการเดาได้ ซึ่งไม่สามารถวิเคราะห์ได้ในทฤษฎีการทดสอบแบบดั้งเดิม ดังนั้นค่าพารามิเตอร์ของข้อสอบทั้งค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าโอกาสการเดา (c) จึงสามารถนำไปใช้ในการปรับปรุงคุณภาพข้อสอบได้อย่างครอบคลุมตามเกณฑ์มาตรฐานสำหรับการทดสอบทางการศึกษาและจิตวิทยา ฉบับล่าสุดในปี 2014 ที่ได้กล่าวว่าไว้ว่า ต้องมีการประเมินคุณภาพของแบบทดสอบทั้งในด้านความตรง ความเที่ยง และความยุติธรรม (AERA, APA, & NCME, 2014)

2. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยระบบคอมพิวเตอร์ทั้ง 3 วิธี ได้แก่ LR SIBTEST และ MH โดยการ

เปรียบเทียบจำนวนข้อที่พบ DIF ตรงกันตั้งแต่ 2 วิธีขึ้นไป พบว่า 1) ด้านทักษะการฟัง (listening) มีข้อสอบที่ทำหน้าที่ต่างกันระหว่างเพศ (หญิง = focal) ภูมิภาค (กทม.และปริมณฑล = focal) และประสบการณ์ในการสอบ (> 1 ครั้ง = focal) จำนวนอย่างละ 3 ข้อ คิดเป็นร้อยละ 12 2) ด้านทักษะโครงสร้างไวยากรณ์ (structure) มีข้อสอบที่ทำหน้าที่ต่างกันระหว่างเพศ ภูมิภาค และประสบการณ์ในการสอบ จำนวน 3, 5 และ 1 ข้อ คิดเป็นร้อยละ 8.57, 14.29 และ 2.86 ตามลำดับ 3) ด้านทักษะการอ่าน (reading) มีข้อสอบที่ทำหน้าที่ต่างกันระหว่างเพศ ภูมิภาค และประสบการณ์ในการสอบ จำนวน 1, 3 และ 0 ข้อ คิดเป็นร้อยละ 2.50, 7.50 และ 0 ตามลำดับ สอดคล้องกับงานวิจัยของ Geramipour & Shahmirzadi (2019) ที่ศึกษาตัวแปรเพศในการส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาอังกฤษระดับชาติ สำหรับชั้น K-11 ซึ่งเป็นคำถามปลายเปิดจำนวน 30 ข้อ ที่วัดทักษะการอ่าน ทักษะการเขียน คำศัพท์ และไวยากรณ์ กลุ่มตัวอย่างกำลังศึกษาระดับอนุปริญญาในประเทศอิหร่าน วิเคราะห์ข้อมูลด้วย iterative hybrid ordinal logistic regression โดยใช้โปรแกรม R ผลการวิจัยพบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบมี 3 ข้อ โดยเพศหญิงจะได้เปรียบในการทำข้อสอบมากกว่าเพศชาย Ryan & Bachman (1992) ศึกษาการทำหน้าที่ต่างกันของข้อสอบ TOEFL ที่วัดทักษะด้านการฟัง ไวยากรณ์ คำศัพท์และการอ่าน จำนวน 117 ข้อ กลุ่มตัวอย่างเป็นอาสาสมัครที่มีภูมิภาคภาษาแม่ 2 กลุ่ม คือ อินโดยูโรเปียน (Indoeuropean) และ ไม่ใช่อินโดยูโรเปียน (non-Indoeuropean) พบว่า ตัวแปรเพศและภูมิภาคภาษาแม่ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาอังกฤษ และ Liao & Yao (2021) ศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบทักษะการฟังภาษาอังกฤษในกลุ่มตัวอย่างนักเรียนที่มีประสบการณ์ต่างกัน คือ เรียนอยู่เกรด 5 และ 6 วิเคราะห์ข้อมูลโดยใช้โปรแกรม R packages “difR” และ “difNLR” 5 วิธี ได้แก่ Lord’s chi-square, Raju’s area tests, MH, LR, และ nonlinear regression ผลการศึกษาพบว่า ทั้ง 5 วิธีพบข้อสอบที่มีการทำหน้าที่แตกต่างกันระหว่างระดับชั้นปีตรงกันจำนวน 16 ข้อ

## ข้อเสนอแนะ

### 1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1.1 ผู้ที่เกี่ยวข้องในการพัฒนาข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราช ด้วยระบบคอมพิวเตอร์ ควรนำผลการวิจัยนี้ไปใช้เป็นข้อมูลในการปรับปรุงคุณภาพข้อสอบให้ดียิ่งขึ้น โดยอาจพิจารณาปรับเลือกใช้ข้อสอบที่มีคุณภาพดี และปรับปรุงข้อสอบที่ยังไม่ได้คุณภาพดีตามเกณฑ์ รวมถึงทบทวนการจัดความเป็นไปได้ของสาเหตุที่ทำให้เกิดการทำหน้าที่ต่างกันของข้อสอบ เช่น คำศัพท์ที่เกี่ยวข้องกับเพศชายหรือหญิง ภาษาที่ใช้ในการเขียนข้อสอบ ลักษณะเนื้อหาที่นำมาสร้างข้อสอบ และลักษณะเฉพาะของกลุ่มผู้สอบ (Posing & Erawan, 2023)

1.2 ผู้ที่เกี่ยวข้องในการพัฒนาข้อสอบวัดความรู้ภาษาอังกฤษของมหาวิทยาลัยสุโขทัยธรรมาธิราช ด้วยระบบคอมพิวเตอร์ ควรนำข้อสอบที่ตรวจพบว่ามีการทำหน้าที่ต่างกันของข้อสอบ (DIF) นี้ไปตรวจสอบวินิจฉัยอีกครั้งโดยให้ผู้เชี่ยวชาญพิจารณาตัดสินจากเนื้อหาเพื่อเป็นการยืนยัน เนื่องจากสถิติที่ใช้ทดสอบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ต่างวิธีกันก็ให้ผลที่ต่างกัน ทั้งนี้ควรระมัดระวังการเปิดเผยข้อมูล

เกี่ยวกับผลการตรวจสอบทางสถิติให้ผู้เชี่ยวชาญทราบ เพื่อให้เกิดผลการตรวจสอบในสภาพความเป็นจริงได้อย่างถูกต้องที่สุด (Haruethaipun et al., 2015)

## 2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ในการทำวิจัยเพื่อวิเคราะห์คุณภาพของข้อสอบใน 3 พารามิเตอร์ ได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และค่าโอกาสการเดาของข้อสอบ (c) ตามหลักการของทฤษฎีการตอบสนองข้อสอบ (IRT) ควรใช้โปรแกรม RStudio Package ‘irtGUI’ เนื่องจากสามารถวิเคราะห์ทั้งคุณภาพของข้อสอบตามหลักการของทฤษฎีการตอบสนองข้อสอบ (IRT) และสามารถตรวจสอบข้อตกลงเบื้องต้นของ IRT ได้แก่ ความเป็นเอกมิติ (unidimensionality) และความเป็นอิสระจากกัน (local independence) โดยทำได้ง่ายและสะดวกรวดเร็ว (user-friendly graphic user interface) มากกว่าการใช้แพ็คเกจอื่น เช่น ltm, eRm, MCMC ซึ่งมีปัญหาเรื่องการเขียนคำสั่งแล้ววิเคราะห์ไม่ออก

2.2 ในการทำวิจัยเพื่อวิเคราะห์ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ผู้วิจัยควรใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมากกว่าหนึ่งวิธี เนื่องจากสถิติที่ใช้ทดสอบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีหลายวิธี โดยวิธีที่ต่างกันก็ให้ผลการวิเคราะห์ที่แตกต่างกัน เช่น ข้อที่เกิด DIF ไม่ตรงกัน และจำนวนข้อที่พบ DIF ไม่เท่ากัน ทั้งนี้อาจเปรียบเทียบผลการวิเคราะห์ระหว่างจำนวนข้อที่พบ DIF ของแต่ละวิธีโดยใช้ฟังก์ชัน dichodif ในโปรแกรม RStudio Package ‘difR’

## References

- AERA, APA, & NCME. (2014). *Standards for Educational and Psychological Testing: National Council on Measurement in Education*. American Educational Research Association.
- Berrío, Á. I., Gómez-Benito, J., & Arias-Patiño, E. M. (2020). Developments and trends in research on methods of detecting differential item functioning. *Educational Research Review*, 31, 100340. <https://doi.org/10.1016/j.edurev.2020.100340>.
- Bichi, A. A., & Talib, R. (2018). Item Response Theory: An Introduction to Latent Trait Models to Test and Item Development. *International Journal of Evaluation and Research in Education (IJERE)*, 7(2), 142-151. <https://doi.org/10.11591/ijere.v7i2.12900>.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press.
- Geramipour, M., & Shahmirzadi, N. (2019). A Gender-Related Differential Item Functioning Study of an English Test. *The Journal of Asia TEFL*, 16(2), 674-682.
- Liao, L., & Yao, D. (2021). Grade-Related Differential Item Functioning in General English Proficiency Test-Kids Listening. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.767244>

Liu, L. M. (2017). *Differential Item Functioning in Large-scale Mathematics Assessments: Comparing the Capabilities of the Rasch Trees Model to Traditional Approaches* [Doctoral dissertation]. University of Toledo.

Ryan, K. E., & Bachman, L. F. (1992). Differential item functioning on two tests of EFL proficiency. *Language Testing*, 9(1), 12-29.  
<https://doi.org/10.1177/026553229200900103>

### Translate Thai References

Haruethaipun, C., Kanjanawasee, S., & Pasiphol, S. (2015). A Development of the Differential Item Functioning Detection Methods by Expert Judgment. *Journal of Education Studies*, 43(1), 1-18. (in Thai)

Jiraro, S., & Aungsuchoti. (2021). Validating Sukhothai Thammathirat Open University's Computer-Based English Proficiency Test (STOU-EPT) Under the Common European Framework of Reference for Languages (CEFR). *Journal of Social Sciences in Measurement Evaluation Statistics and Research*, 2(1) 27-35. (in Thai)

Kanjanawasee, S. (2020). *Classical Test Theory* (3th ed.). Chulalongkorn University Press. (in Thai)

Posing, S. & Erawan, W. (2023). Detecting Differential Item Functioning of Grade 9 Science Literacy Test. *Journal of Educational Measurement Mahasarakham University*, 29(1), 231-244. (in Thai)