

บทความวิจัย (Research Article)

การพัฒนาแบบทดสอบมาตรฐานด้านการฟังภาษาอังกฤษ สำหรับนักศึกษา
มหาวิทยาลัยราชภัฏพิบูลสงคราม
A Development of the Standard English Listening Test of
Pibulsongkram Rajabhat University

นิภาวรรณ นาวาตัน
Nipawan Navawatana

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อพัฒนาแบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม และ 2) เพื่อหาคุณภาพของแบบทดสอบมาตรฐานและค่าสารสนเทศการฟังภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม กลุ่มตัวอย่าง ได้แก่ นักศึกษาชั้นปีที่ 3 ปีการศึกษา 2564 จำนวน 1,498 คน จาก 8 คณะ เครื่องมือที่ใช้ในการวิจัย ได้แก่ แบบทดสอบแบบปรนัย จำนวน 2 ฉบับ ๆ ละ 100 ข้อ รวม 200 ข้อ และมีการวิเคราะห์คุณภาพของแบบทดสอบโดยใช้การวิเคราะห์ตามแนวทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) สถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ ค่าร้อยละ (Percentage) ค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) การหาค่าความยาก (b) อำนาจจำแนก (a) ค่าโอกาสการเดา (c) และความเชื่อมั่นของแบบทดสอบ (Reliability) ผลการวิจัยพบว่า

1. แบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม จำนวน 2 ฉบับ มีการแบ่งเนื้อหาออกเป็น 4 ตอน จำนวน 200 ข้อ โดยคัดเลือกคำศัพท์ ร้อยละ 95 ของคำศัพท์ที่มีความถี่สูงสุดทั้งหมด และมีค่าดัชนีความสอดคล้อง (IOC) เท่ากับ 0.99 และค่าสัมประสิทธิ์ความเชื่อมั่นแอลฟาครอนบาค (Coefficient Alpha) ทั้งฉบับ เท่ากับ 0.904

2. ผลการวิเคราะห์คุณภาพของแบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม ตามแนวทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) พบว่า แบบทดสอบฉบับที่ 1 มีความยาก (b) ตั้งแต่ 0.70 ถึง 2.47 ค่าอำนาจจำแนก (a) ตั้งแต่ -2.23 ถึง 1.62 และค่าโอกาสการเดา (c) ตั้งแต่ 0.00 ถึง 0.26 ส่วนแบบทดสอบฉบับที่ 2 มีค่าความยาก (b) ตั้งแต่ 0.31 ถึง 2.48 ค่าอำนาจจำแนก (a) ตั้งแต่ -1.47 ถึง 2.62 และค่าโอกาสการเดา (c) ตั้งแต่ 0.00 ถึง 0.29 แบบทดสอบฉบับที่ 1 ผ่านเกณฑ์ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) มีจำนวน 56 ข้อ ส่วนแบบทดสอบฉบับที่ 2 ผ่านเกณฑ์ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) มีจำนวน 76 ข้อ รวมจำนวนข้อสอบจากทั้ง 2 ฉบับที่ได้คุณภาพคือ 132 ข้อ

คำสำคัญ: แบบทดสอบมาตรฐาน ความสามารถในการฟัง ทฤษฎีการตอบสนองข้อสอบ

คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏพิบูลสงคราม จังหวัดพิษณุโลก 65000

Faculty of Humanities and Social Sciences, Pibulsongkram Rajabhat University, Phitsanulok 65000

Corresponding author; email: nipawanbell@hotmail.com

(Received: 4 September 2023; Revised: 24 October 2023; Accepted: 4 January 2024)

DOI: <https://doi.org/10.14456/psruhss.2024.34>

Abstract

The purposes of this research were as follows: 1) to develop a standard English listening test for students at Pibulsongkram Rajabhat University and 2) to assess the quality of the standard English listening test for students at Pibulsongkram Rajabhat University. The sample group consisted of 1,498 third-year students in the academic year of 2564 from 8 faculties. The research instrument comprised two sets of standard English listening test items, totaling 100 items each and 200 items overall. Additionally, the standard English listening test was analyzed using the Item Response Theory method. Statistical analysis included percentage, mean, standard deviation, difficulty index, discrimination power, guessing, and test reliability. The research findings indicated that:

1. The two sets of the standard English listening test for Pibulsongkram Rajabhat University are divided into four sections with a total of 100 items each. High-frequency vocabulary, useful for learners and recommended for prior learning, was included in the test by selecting 95% of the most frequently used words. The Index of Consonance (IOC) is 0.99, and the Cronbach's Alpha coefficient for both versions is 0.904.

2. The quality assessment of the standard English listening test for Pibulsongkram Rajabhat University, based on the Item Response Theory (IRT) method, revealed that Set 1 had item difficulty scores ranging from 0.70 to 2.47, a discrimination power (b) ranging from -2.23 to 1.62, and distractor functionality ranging from 0.00 to 0.26. Meanwhile, Set 2 had item difficulty scores ranging from 0.31 to 2.48, discrimination power (a) ranging from -1.47 to 2.62, and guessing ranging from 0.00 to 0.29. A total of 56 items in Set 1 and 76 items in Set 2 passed the IRT criteria. The overall number of qualified test items from both sets is 132.

Keywords: Standardized test, English listening ability, Item response theory (IRT)

บทนำ

ในปัจจุบัน สถาบันอุดมศึกษาหลายแห่งทั้งของรัฐและเอกชนต่างนิยมใช้การทดสอบมาตรฐานเพื่อวัดความสามารถทางภาษาอังกฤษมากยิ่งขึ้น เนื่องจากภาพรวมความสามารถทางภาษาอังกฤษของคนไทยจากการสำรวจของ EF (Education First) ปี 2020 พบว่า ทักษะภาษาอังกฤษของคนไทยอยู่ในลำดับที่ 89 จาก 100 ประเทศของโลกและลำดับที่ 20 ของเอเชียจากจำนวน 24 ประเทศ ค่าคะแนนเฉลี่ย 419 คะแนน จัดอยู่ในกลุ่มความสามารถต่ำมาก (Education First, 2020) ดังนั้น สถาบันอุดมศึกษาส่วนใหญ่จึงต้องการทดสอบความสามารถทางภาษาอังกฤษที่ได้มาตรฐานเพื่อประเมินความรู้ของผู้เรียน เพราะตระหนักถึงความจำเป็นในการใช้ภาษาอังกฤษทั้งในด้านการเรียน การทำงาน และการศึกษาต่อในระดับที่สูงขึ้น จากเหตุผลดังกล่าว สำนักงานการศึกษาขั้นพื้นฐานได้กำหนดการจัดการเรียนการสอนภาษาอังกฤษในฐานะภาษาต่างประเทศไว้ในหลักสูตรแกนกลางขั้นพื้นฐาน พุทธศักราช 2551 ได้กำหนดนโยบายการปฏิรูปแนวทางการเรียนการสอนภาษาอังกฤษในยุทธศาสตร์การปฏิรูปการศึกษาขั้นพื้นฐาน (พ.ศ 2558-2563) และแนวทางการสอนภาษาอังกฤษแนวใหม่ โดยกำหนดคุณภาพผู้เรียนตามกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรป (The Common European Framework of Reference for Language: CEFR) ให้ปรับการเรียนการสอน จากการเน้นไวยากรณ์มาเป็นการสอนภาษาเพื่อการสื่อสาร โดยเน้นทักษะการฟัง การพูด การอ่าน และการเขียน ตามลำดับ

ดังนั้นเพื่อค้นหากลยุทธ์และนวัตกรรมในการแก้ปัญหา โดย 1 ใน 5 ยุทธศาสตร์การปฏิรูปการศึกษา ภาษาอังกฤษของประเทศไทย คือ การออกแบบแบบทดสอบเพื่อวัดความสามารถภาษาอังกฤษ (English Proficiency Tests) เมื่อศึกษาต่อในระดับต่าง ๆ เพื่อให้กฎเกณฑ์ดังกล่าวเป็นตัวกระตุ้นความสามารถด้านการใช้ภาษาต่างประเทศของคนไทย (สำนักงานคณะกรรมการส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม (สกสว.), 6 ตุลาคม 2562) และจากการที่มหาวิทยาลัยราชภัฏทั่วประเทศได้ร่วมกันทำงานเพื่อพัฒนานักศึกษาและท้องถิ่นมายาวนาน ทำให้ในปี พ.ศ. 2560 ได้ร่วมกันจัดทำยุทธศาสตร์มหาวิทยาลัยราชภัฏ ระยะ 20 ปี (2560-2579) โดยหนึ่งในโครงการร่วมคือโครงการพัฒนาความรู้ทักษะด้านภาษาอังกฤษในศตวรรษที่ 21 ซึ่งผลการดำเนินการมาจนถึงปัจจุบัน (2566) ระยะเวลา 6 ปี การพัฒนาความรู้และทักษะภาษาอังกฤษของนักศึกษามีความก้าวหน้าและเป็นรูปธรรมมากขึ้น โดยในปี 2563 แต่ละมหาวิทยาลัยได้พัฒนาข้อสอบสำหรับวัดความสามารถทางภาษาอังกฤษของนักศึกษาในแต่ละมหาวิทยาลัยรวมถึงมหาวิทยาลัยราชภัฏพิบูลสงคราม แต่เนื่องจากกรอบเวลาการทำงานที่ถูกจำกัด การพัฒนาข้อสอบที่ผ่านมายังไม่ได้เข้าสู่กระบวนการวิจัยเพื่อทดสอบคุณภาพ ในปีงบประมาณ 2564 นี้ คณะมนุษยศาสตร์และสังคมศาสตร์จึงได้จัดสรรงบประมาณในการพัฒนาชุดข้อสอบให้ได้มาตรฐานยิ่งขึ้นโดยแบ่งออกเป็น 2 ชุดข้อสอบ คือ ชุดข้อสอบการฟังและชุดข้อสอบการอ่าน โดยในงานวิจัยมีวัตถุประสงค์เพื่อพัฒนาชุดข้อสอบการฟังโดยเพิ่มกระบวนการศึกษาเพื่อให้ได้มาซึ่งคำศัพท์และกลุ่มคำศัพท์ที่จำเป็นสำหรับการใช้จริงจากแหล่งข้อมูลที่หลากหลายเพื่อให้มั่นใจว่าเนื้อหาข้อสอบจะเป็นเนื้อหาที่เป็นมาตรฐานในการใช้งานจริงของภาษาอังกฤษในชีวิตประจำวันและในการทำงานซึ่งกระบวนการคัดคำศัพท์แบ่งกลุ่มคำศัพท์ตามเกณฑ์ CEFR และนำมาเป็นเนื้อหาในข้อสอบได้ผ่านกระบวนการวิจัยและเผยแพร่ในวารสารอย่างเป็นทางการ ซึ่งกระบวนการคัดคำศัพท์ก่อนนำมาพัฒนาเป็นข้อสอบในงานวิจัยนี้เผยแพร่อยู่ในงานของณัฐกานต์ เสงี่ยม (2566)

แบบทดสอบมาตรฐานทางการศึกษา (Standardized Test) เป็นแบบทดสอบที่มีระเบียบที่ถูกกำหนดเกี่ยวกับการดำเนินงาน การบริหารจัดการ เอกสาร และการให้คะแนนที่แน่นอน เพื่อให้การจัดการประเมินเหมือนกันในการทดสอบที่มีสถานที่สอบและเวลาอันแตกต่างกัน (Nitko, 2004) แบบทดสอบมาตรฐานทางการศึกษาเป็นรูปแบบของการวัดที่ต้องอ้างอิงถึงกลุ่มประชากรเฉพาะเจาะจง ความเป็นมาตรฐานประกอบไปด้วย การบริหารการทดสอบ โดยนำแบบทดสอบให้กับกลุ่มผู้สอบและคำนวณค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน และคะแนนเปอร์เซ็นต์ไทล์ออกมาเพื่อนำไปเปรียบเทียบกับคะแนนของผู้สอบรายบุคคลกับความสามารถที่เป็นคะแนนปกติของกลุ่ม (A Glossary of Measurement Term Eric Digest, 1989) ดังนั้น การสร้างแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยให้มีมาตรฐานระดับสากลสามารถทำได้ด้วยการปฏิบัติตามทฤษฎีการออกแบบและสร้างข้อสอบ 3 ขั้นตอน ตาม Bachman & Palmer (1996) ดังนี้ 1) ขั้นตอนออกแบบ (Design Stage) ขั้นตอนนี้ผู้พัฒนาแบบทดสอบได้อธิบายส่วนประกอบต่าง ๆ ในการออกแบบทดสอบอย่างละเอียดชัดเจน 2) ขั้นตอนปฏิบัติการ (Operationalization) ขั้นตอนนี้เกี่ยวกับการพัฒนาโครงสร้างลักษณะของแบบทดสอบ พร้อมทั้งมีการบรรยายถึงหลักการที่ใช้ในแบบทดสอบ เช่น การเขียนคำสั่งและการระบุวิธีตรวจให้คะแนนแบบทดสอบ และ 3) ขั้นตอนการบริหารการสอบ (Test Administration) การบริหารการสอบเกี่ยวข้องกับการจัดสอบกับกลุ่มเป้าหมาย การเก็บรวบรวมข้อมูลและการวิเคราะห์ข้อมูล โดยสอดคล้องกับ Genesee et al. (1996) เสนอขั้นตอนการพัฒนาแบบทดสอบมาตรฐาน (Standardized Test Development) ดังนี้ ขั้นตอนที่ 1 การกำหนดจุดประสงค์การสอบ การพัฒนาแบบทดสอบมาตรฐานเหมือนกับการพัฒนาแบบทดสอบประเภทอื่น ๆ ขั้นตอนที่ 2 การจัดทำโครงสร้างแบบทดสอบและการสร้างชุดคำถาม เนื่องจากแบบทดสอบมาตรฐานได้อิงวัตถุประสงค์ ขั้นตอนที่ 3 การทดลองใช้แบบทดสอบกับกลุ่มทดลอง หลังจากสร้างข้อคำถามเสร็จแล้ว นักพัฒนาแบบทดสอบนำชุดข้อคำถามที่ผลิตได้ไปทดลองสอบกับกลุ่มทดลอง ภายใต้สภาพแวดล้อมที่ใกล้เคียงที่สุดกับสถานการณ์ที่แบบทดสอบจะถูกใช้จริง ขั้นตอนที่ 4 การวิเคราะห์

แบบทดสอบเบื้องต้นและการคัดเลือกแบบทดสอบเพื่อสร้างแบบทดสอบฉบับจริง ขั้นตอนที่ 5 การกำหนดเกณฑ์การประเมินผลการสอบ และการกิจสุดท้ายในการพัฒนาแบบทดสอบมาตรฐานแบบอิงกลุ่ม คือ การกำหนดเกณฑ์ระดับคะแนน

ทฤษฎีที่ใช้ในการวิเคราะห์ข้อสอบได้มีการพัฒนาอย่างต่อเนื่องจนถึงปัจจุบัน ทฤษฎีการตอบสนองข้อสอบ (Item response theory: IRT) เป็นการนำหลักการความน่าจะเป็นมาวิเคราะห์ข้อสอบ สามารถแยกผลการวิเคราะห์ข้อสอบแต่ละข้อเป็นอิสระจากข้อสอบข้ออื่นในการสอบเดียวกันได้ รวมถึงค่าความยากและค่าอำนาจจำแนกของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง ทำให้ทฤษฎีการตอบสนองข้อสอบ มีข้อได้เปรียบกว่าทฤษฎีแบบดั้งเดิม ซึ่งสอดคล้องกับ Rivioelli & Hambleton (1977) ที่กล่าวว่า ด้วยคุณสมบัติด้านความไม่เปลี่ยนแปลงตามกลุ่มตัวอย่าง (Invariance) ของค่าพารามิเตอร์ของข้อสอบจากการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ จึงทำให้ค่าฟังก์ชันสารสนเทศเหมาะสมที่จะใช้เป็นตัวบ่งบอกคุณภาพของข้อสอบและแบบทดสอบแทนการหาค่าความเที่ยงและความคลาดเคลื่อนมาตรฐานของการวัดตามทฤษฎีแบบดั้งเดิม สารสนเทศของข้อสอบ (Item Information) เป็นดัชนีสมมติที่สร้างจากดัชนีคุณลักษณะของข้อสอบหลายลักษณะ ประกอบด้วยค่าพารามิเตอร์ระดับขั้นความยากของการตอบ ค่าพารามิเตอร์ความชัน และค่าความแปรปรวนของคะแนนรายข้อ (ศิริชัย กาญจนวาสี, 2563) สารสนเทศของแบบสอบ (Test Information) เป็นความถูกต้องแม่นยำในการประมาณความสามารถของผู้ตอบ ซึ่งเป็นค่าฟังก์ชันสารสนเทศของแบบสอบ อันเกิดจากผลรวมเชิงพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกันทั้งฉบับ ณ ตำแหน่ง θ เดียวกันสารสนเทศของข้อสอบแต่ละข้อ จึงมีส่วนอย่างเป็นอิสระจากกันต่อค่าสารสนเทศของแบบสอบ ลักษณะเช่นนี้ไม่ได้เกิดกับทฤษฎีการทดสอบแบบดั้งเดิม เนื่องจากค่าความยากและค่าอำนาจจำแนกของข้อสอบแต่ละข้อต่างก็ส่งผลต่อค่าความเที่ยงของแบบสอบทั้งฉบับ แต่ไม่สามารถคำนวณค่าของแต่ละข้อได้อย่างเป็นอิสระจากกัน ดังนั้นคะแนนที่ได้จึงขึ้นอยู่กับลักษณะเฉพาะของกลุ่มข้อสอบ และแบบสอบเฉพาะฉบับที่เลือกมาใช้เนื่องจากค่าสารสนเทศมีความสัมพันธ์ผกผันกับความคลาดเคลื่อนมาตรฐานของการประมาณค่า ดังนั้น ถ้าค่าสารสนเทศของแบบสอบมีค่าสูงในช่วง θ ใดก็必将มีความถูกต้องแม่นยำสูง ในการประมาณค่าความสามารถของผู้ตอบในช่วง θ นั้น ๆ โดยมีความคลาดเคลื่อนมาตรฐานของการประมาณค่าต่ำ

งานวิจัยนี้มีความแตกต่างจากงานวิจัยพัฒนาข้อสอบก่อนหน้านี้ กล่าวคือ การใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) ในการวิเคราะห์ เนื่องจากใช้กลุ่มตัวอย่างขนาดใหญ่ คือนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม จำนวน 1,498 คน จาก 8 คณะ และมีกระบวนการคัดเลือกคำศัพท์ และกลุ่มคำศัพท์ที่ใช้ในข้อสอบด้วยกระบวนการวิจัยแบบภาษาศาสตร์คลังข้อมูล (ณัฐกานต์ เสงี่ยม, 2566)

วัตถุประสงค์

1. เพื่อพัฒนาแบบทดสอบมาตรฐานวัดความสามารถด้านฟังภาษาอังกฤษ สำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม
2. เพื่อหาคุณภาพของแบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษ สำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม

วิธีดำเนินการวิจัย

ประชากรและกลุ่มตัวอย่าง

ประชากร ได้แก่ นักศึกษาของมหาวิทยาลัยราชภัฏพิบูลสงคราม ชั้นปีที่ 3 ที่ลงทะเบียนและกำลังศึกษาอยู่ในปีการศึกษา 2564 จำนวน 2,625 คน จาก 8 คณะ ได้แก่ 1) คณะครุศาสตร์ 2) คณะวิทยาศาสตร์และเทคโนโลยี 3) คณะมนุษยศาสตร์และสังคมศาสตร์ 4) คณะวิทยาการจัดการ 5) คณะเทคโนโลยีการเกษตรและอาหาร 6) คณะเทคโนโลยีอุตสาหกรรม 7) คณะสังคมศาสตร์และการพัฒนาท้องถิ่น และ 8) คณะพยาบาล

กลุ่มตัวอย่างในการวิจัยครั้งนี้ ผู้วิจัยใช้การสุ่มตัวอย่างแบบ 2 ขั้นตอน ได้แก่ ขั้นตอนที่ 1 การเลือกกลุ่มตัวอย่างแบบเจาะจง (Purposive Sampling) เป็นการเลือกกลุ่มตัวอย่างจากผลการทดสอบ Pre-test ของนักศึกษาที่มีผลการสอบระดับตั้งแต่ A2 ขึ้นไป ขั้นตอนที่ 2 การสุ่มตัวอย่างโดยอาสาสมัคร (Voluntary Sampling) เป็นการเลือกตัวอย่างจากการสมัครใจของนักศึกษาที่มีความสนใจเข้ารับการทดสอบวัดระดับความสามารถภาษาอังกฤษของนักศึกษาชั้นปีที่ 3 ปีการศึกษา 2564 จำนวน 1,498 คน ประกอบด้วย 8 คณะ ดังตาราง 1

ตาราง 1 ประชากรและกลุ่มตัวอย่าง

คณะ	นักศึกษา	
	ประชากร (คน)	กลุ่มตัวอย่าง (คน)
คณะครุศาสตร์	465	264
คณะวิทยาศาสตร์และเทคโนโลยี	242	144
คณะมนุษยศาสตร์และสังคมศาสตร์	296	185
คณะวิทยาการจัดการ	713	437
คณะเทคโนโลยีการเกษตรและอาหาร	66	30
คณะเทคโนโลยีอุตสาหกรรม	397	201
คณะสังคมศาสตร์และการพัฒนาท้องถิ่น	399	208
คณะพยาบาล	47	29
รวม	2,625	1,498

ตัวแปรที่ศึกษา

1. แบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม จำนวน 2 ฉบับ

2. คุณภาพของแบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ มีจำนวน 2 ฉบับ ได้แก่

1. แบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงครามที่สร้างขึ้นโดยคณะอาจารย์สาขาวิชาภาษาอังกฤษที่มีประสบการณ์ในการสอนภายใต้การชี้แนะของผู้เชี่ยวชาญภาษาอังกฤษ และการวัดผลภาษาอังกฤษ และผู้ทรงคุณวุฒิได้ตรวจสอบดัชนีความสอดคล้อง (IOC) แบบทดสอบทั้งหมดมี 2 ฉบับ ๆ ละ 100 ข้อ รวม 200 ข้อ ซึ่งมีลักษณะแบบปรนัยโดยเทียบเคียงจากข้อสอบมาตรฐาน TOEIC แบ่งเนื้อหาออกเป็น 4 ส่วน คือ Part 1 : Photographs จำนวน 12 ข้อ Part 2 : Question – Response จำนวน 50 ข้อ Part 3 : Conversations จำนวน 78 ข้อ และ Part 4 : Short Talks จำนวน 60 ข้อ รวมทั้งสิ้น 200 ข้อ โดยแบบทดสอบมีคุณภาพด้านคุณประโยชน์ตามเกณฑ์ที่กำหนด มีค่าดัชนีความสอดคล้อง (IOC) เท่ากับ 0.99

การเก็บรวบรวมข้อมูลและการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ใช้ระยะเวลาดำเนินการ ตั้งแต่เดือนกันยายน 2564 – 2565 ซึ่งผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูล โดยมี 3 ขั้นตอนในการพัฒนาแบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษตามแนวคิดของ Bachman & Palmer (2010) และ Genesee & Upshur (1996) คือ ขั้นที่ 1 การออกแบบแบบทดสอบ (Test Design Stage) ขั้นที่ 2 การปฏิบัติการพัฒนาแบบทดสอบ (Test Operationalization Stage) และขั้นที่ 3 การบริหารการสอบ (Test Administration Stage) มีรายละเอียด แต่ละขั้น ดังนี้

ขั้นตอนที่ 1 การออกแบบแบบทดสอบ (Test Design Stage)

1.1 ศึกษาข้อมูลจากเอกสาร ตำรา และงานวิจัยจากแหล่งข้อมูลต่าง ๆ เช่น กรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรป (CEFR) (Council of Europe, 2018) โครงสร้างแบบทดสอบวัดความสามารถภาษาอังกฤษมาตรฐานสากล ได้แก่ TOEFL, TOEIC, IELTS, CU-TEP, TU-GET ฯลฯ หลักการและทฤษฎีการพัฒนาแบบทดสอบวัดความสามารถภาษาอังกฤษ (Bachman & Palmer, 2010; Genesee & Upshur, 1996) งานวิจัยกระบวนการคัดคำศัพท์เพื่อนำมาพัฒนาเป็นข้อสอบด้วยกระบวนการภาษาศาสตร์คลังข้อมูล (ณัฐกานต์ เสงี่ยม, 2566) รวมถึงหลักการและทฤษฎีการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) (ศิริชัย กาญจนวสี, 2563) หลังจากนั้นประมวลความรู้เพื่อนำมากำหนดกรอบแนวคิดเพื่อใช้ในการดำเนินการศึกษาวิจัย

1.2 ประชุมสัมมนาเชิญผู้ทรงคุณวุฒิซึ่งเป็นผู้เชี่ยวชาญด้านการสร้างแบบทดสอบภาษาอังกฤษ จำนวน 2 ท่าน และด้านสถิติการวัดและประเมินผล จำนวน 2 ท่าน เพื่อกำหนดทักษะที่ต้องการวัดผู้สอบในแต่ละตอนของแบบทดสอบการฟัง และจำนวนข้อในการพัฒนาข้อสอบเพื่อเป็นกรอบในการออกข้อสอบ

ขั้นตอนที่ 2 การปฏิบัติการพัฒนาแบบทดสอบ (Test Operationalization Stage)

ขั้นปฏิบัติการพัฒนาแบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษสำหรับนักศึกษา มหาวิทยาลัยราชภัฏพิบูลสงคราม ผู้วิจัยแบ่งการดำเนินงานสร้างและพัฒนาเครื่องมือในการดังนี้

2.1 แบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม ผู้วิจัยดำเนินงานดังนี้

2.1.1 จัดทำโครงสร้างลักษณะแบบทดสอบมาตรฐาน (Test Specifications) วัดความสามารถด้านการฟังอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม ผู้วิจัยได้กำหนดโครงสร้างของแบบทดสอบตามรายการหัวข้อที่ควรปรากฏในโครงสร้างลักษณะของแบบทดสอบ และมีการนำผลการวิจัยที่ได้จากกระบวนการศึกษาของณัฐกานต์ เสงี่ยม (2566) มาเป็นข้อมูลชุดคำศัพท์ในการพัฒนาเป็นข้อสอบการฟัง โดยมีจำนวนคำและระดับความยากง่าย ตามเกณฑ์การแบ่งของ CEFR พบว่า คำศัพท์ส่วนใหญ่ที่พบในแนวข้อสอบ TOEIC ส่วนของการฟัง (Listening Part) อยู่ในระดับ B1 มากที่สุด โดยณัฐกานต์ เสงี่ยม (2566) เลือกคำศัพท์ร้อยละ 95 ตามแนวคิดของ Laufer (1989) ของคำศัพท์ที่มีความถี่สูงสุดจากคำศัพท์ 3,987 คำ ได้คำศัพท์ทั้งหมด 3,788 คำ

2.1.2 การพัฒนาแบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษ ผู้วิจัยดำเนินการตามขั้นตอน ดังนี้

1) ผู้วิจัยนำข้อมูลจากระยะที่ 1 มาพัฒนาเป็นแบบทดสอบ จำนวน 200 ข้อ โดยจัดทำตารางวิเคราะห์เนื้อหา กำหนดจำนวนพาร์ทและข้อที่เหมาะสมโดยเทียบเคียงจากข้อสอบมาตรฐาน TOEIC โดย แบ่งเนื้อหาออกเป็น 4 ตอน คือ ตอนที่ 1 : Photographs จำนวน 12 ข้อ Part 2 : Question – Response จำนวน 50 ข้อ Part 3 : Conversations จำนวน 78 ข้อ และ Part 4 : Short Talks จำนวน 60 ข้อ รวมทั้งสิ้น 200 ข้อ คัดเลือกคำศัพท์ที่ได้จากงานวิจัยของณัฐกานต์ เสงี่ยม (2566) คำศัพท์ทั่วไประดับ K 1 (General Service List) ที่อยู่ในระดับ A2 – B2 มาใช้สำหรับเป็นกรอบของการพัฒนาแบบทดสอบ

2) นำแบบทดสอบทั้ง 2 ฉบับไปให้ผู้เชี่ยวชาญตรวจสอบดัชนีความสอดคล้อง (IOC) โดยผู้ทรงคุณวุฒิทางด้านภาษาอังกฤษ ด้านการวัดและประเมินผล ที่มีประสบการณ์ด้านการสอนวิชาภาษาอังกฤษไม่น้อยกว่า 10 ปี รวมทั้งสิ้น 3 คน โดยผู้เชี่ยวชาญพิจารณาความสอดคล้องในแง่ของภาษาและประเด็นคำถาม และตัวเลือกตอบ รวมทั้งประเมินประโยชน์ จากผลการวิเคราะห์สามารถนำข้อสอบไปใช้ได้ จำนวน 198 ข้อ และตัดทิ้ง/ปรับปรุง จำนวน 2 ข้อ

3) นำแบบทดสอบมาปรับปรุงแก้ไขตามข้อเสนอแนะให้มีความชัดเจนมากยิ่งขึ้น

ขั้นตอนที่ 3 การบริหารการสอบ (Test Administration Stage)

ขั้นตอนนี้เป็นการบริหารการสอบโดยมีกิจกรรมหลักที่สำคัญคือก่อนการเก็บข้อมูล ขณะเก็บข้อมูล และการวิเคราะห์ประสิทธิภาพแบบทดสอบ มีรายละเอียดดังนี้

3.1 กิจกรรมก่อนการเก็บข้อมูล ผู้วิจัยมีการปฐมนิเทศผู้เข้าสอบ จัดคนคุมสอบ และผู้ตรวจแบบทดสอบ มีการประชุมผู้คุมสอบและกับตรวจแบบทดสอบ อบรมผู้ดำเนินการสอบพูด และผู้ตรวจแบบทดสอบการเขียน จัดเตรียมอุปกรณ์และห้องสอบ และตรวจสอบความพร้อมก่อนวันสอบ

3.2 กิจกรรมขณะเก็บข้อมูล ผู้วิจัยกำหนดวันและทำหนังสือขอความอนุเคราะห์จากทั้ง 8 คณะในมหาวิทยาลัยเพื่อขออนุญาตให้นักศึกษาเข้าร่วมการทดสอบ และดำเนินการทดสอบด้วยแบบทดสอบที่พัฒนาขึ้น

3.3 การวิเคราะห์ประสิทธิภาพของแบบทดสอบ ประกอบด้วยคะแนนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความยาก อำนาจจำแนก โอกาสการเดา และความเชื่อมั่นแบบทดสอบ

3.3.1 นำแบบทดสอบไปหาค่าความยาก (b) และหาค่าอำนาจจำแนก (a) โดยใช้โปรแกรมคอมพิวเตอร์ตามทฤษฎีการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) รายข้อ โดยมีการคัดเลือกข้อสอบที่มีค่าความยาก (b) ตั้งแต่ 0.3 ถึง 2.5 และข้อสอบที่มีค่าอำนาจจำแนก (a) ตั้งแต่ - 3.0 ถึง 3.0 นอกจากนี้ยังได้นำแบบทดสอบเพื่อหาค่าโอกาสการเดา (c) เป็นการวิเคราะห์หาความน่าจะเป็นของบุคคลหนึ่งที่ไม่มีความสามารถในการตอบข้อสอบนั้นได้ถูกต้อง เป็นค่าที่แสดงถึงโอกาสการตอบข้อสอบถูก โดยไม่มีความรู้ในเรื่องนั้น มีค่าจาก 0 ถึง 1 จะคัดเลือกเอาข้อสอบที่มีค่าไม่เกิน 0.3

3.3.2 นำข้อสอบที่ผ่านการคัดเลือกให้ได้ข้อสอบ จำนวน 200 ข้อ เพื่อใช้เป็นชุดข้อสอบภาษาอังกฤษมาตรฐานที่ได้พัฒนาแล้วไปใช้

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลมีการใช้สถิติ ดังนี้

1. สถิติพื้นฐาน ได้แก่ ค่าร้อยละ (Percentage) คะแนนเฉลี่ย (Mean) และความเบี่ยงเบนมาตรฐาน (Standard deviation)

2. สถิติที่ใช้หาคุณภาพของแบบทดสอบ ได้แก่ ความยาก (b) อำนาจจำแนก (a) โอกาสการเดา (c) และความเชื่อมั่นของแบบทดสอบ (Reliability)

ผลการวิจัย

การวิจัยเรื่อง การพัฒนาแบบทดสอบมาตรฐานการฟังภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม ผู้วิจัยนำเสนอผลการวิจัยออกเป็น 2 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์แบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษ สำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม โดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) รายละเอียดดังนี้

1. ผลการวิเคราะห์แบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงครามตามทฤษฎีการตอบสนองข้อสอบ

ตาราง 2 จำนวนแบบทดสอบฉบับที่ 1 จำแนกตามเกณฑ์การคัดเลือกตามทฤษฎีการตอบสนองข้อสอบ

ความยาก (b)	การเดา (c)	อำนาจจำแนก (a)			รวม
		$a < 0.3$	$0.3 \leq a \leq 2.50$	$a > 2.50$	
$b > 3.00$	$c \geq 0.30$	-	-	-	-
	$c < 0.30$	-	-	-	-
$-3.00 \leq b \leq 3.00$	$c \geq 0.30$	-	29	2	31
	$c < 0.30$	3	18	2	23
$b < -3.00$	$c \geq 0.30$	-	-	-	-
	$c < 0.30$	-	-	2	2
รวม		3	47	6	56

จากตาราง 2 แสดงค่าอำนาจจำแนก ค่าความยาก และค่าโอกาสการเดาของแบบทดสอบที่มีค่าสารสนเทศของแบบทดสอบและข้อสอบ จำนวน 56 ข้อ พบว่า ข้อสอบที่มีค่าอำนาจจำแนกอยู่ระหว่าง $0.3 - 2.5$ มีจำนวน 47 ข้อ ข้อสอบที่มีค่าความยากอยู่ระหว่าง -3.00 ถึง $+3.00$ จำนวน 54 ข้อ และข้อสอบที่มีค่าโอกาสการเดาอยู่ต่ำกว่า 0.3 จำนวน 25 ข้อ เมื่อพิจารณารายข้อ โดยพิจารณาค่าอำนาจจำแนกและค่าความยาก ในแต่ละข้อ พบว่ามีข้อสอบที่ผ่านเกณฑ์สามารถนำไปใช้ได้จำนวน 47 ข้อ

ตาราง 3 จำนวนแบบทดสอบฉบับที่ 2 จำแนกตามเกณฑ์การคัดเลือกตามทฤษฎีการตอบสนองข้อสอบ

ความยาก (b)	การเดา (c)	อำนาจจำแนก (a)			รวม
		$a < 0.3$	$0.3 \leq a \leq 2.50$	$a > 2.50$	
$b > 3.00$	$c \geq 0.30$	-	-	-	-
	$c < 0.30$	-	-	-	-
$-3.00 \leq b \leq 3.00$	$c \geq 0.30$	-	12	1	13
	$c < 0.30$	4	50	6	60
$b < -3.00$	$c \geq 0.30$	-	-	-	-
	$c < 0.30$	2	1	-	3
รวม		6	63	7	76

จากตาราง 3 แสดงค่าอำนาจจำแนก ค่าความยาก และค่าโอกาสการเดาของแบบทดสอบที่มีค่าสารสนเทศของแบบทดสอบและข้อสอบ จำนวน 76 ข้อ พบว่า ข้อสอบที่มีค่าอำนาจจำแนกอยู่ระหว่าง $0.3 - 2.5$ มีจำนวน 63 ข้อ ข้อสอบที่มีค่าความยากอยู่ระหว่าง -3.00 ถึง $+3.00$ จำนวน 75 ข้อ และข้อสอบที่มีค่าโอกาสการเดาอยู่ต่ำกว่า 0.3 จำนวน 87 ข้อ เมื่อพิจารณารายข้อ โดยพิจารณาค่าอำนาจจำแนกและค่าความยาก ในแต่ละข้อ พบว่ามีข้อสอบที่ผ่านเกณฑ์สามารถนำไปใช้ได้จำนวน 63 ข้อ และข้อสอบที่ไม่เป็นไปตามเกณฑ์จำนวน 13 ข้อ

ตอนที่ 2 ผลการวิเคราะห์ข้อมูลสารสนเทศของแบบทดสอบ (Item Information) มาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษสำหรับนักศึกษามหาวิทยาลัยราชภัฏภูเก็ต

ตาราง 4 ผลการวิเคราะห์ข้อมูลสารสนเทศของแบบทดสอบมาตรฐานการฟังภาษาอังกฤษตามทฤษฎีการตอบสนองข้อสอบแบบทดสอบฉบับที่ 1

ค่าความสามารถของผู้สอบ (θ)	ค่าสารสนเทศของ แบบทดสอบ (θ)	ความคลาดเคลื่อน มาตรฐาน	ระดับความสามารถการฟัง
2.8	2.469	0.636	ความสามารถสูง
2.4	4.766	0.458	
2	12.182	0.287	
1.6	26.602	0.194	ความสามารถค่อนข้างสูง
1.2	29.612	0.184	
0.8	30.169	0.093	
0.4	18.074	0.235	ความสามารถปานกลาง
0	8.233	0.349	
-0.4	5.187	0.439	
-0.8	3.804	0.513	ความสามารถค่อนข้างต่ำ
-1.2	3.171	0.562	
-1.6	3.383	0.544	
-2	3.765	0.515	ความสามารถต่ำ
-2.4	2.546	0.627	
-2.8	1.409	0.843	

จากตาราง 4 ผลการวิเคราะห์ค่าสารสนเทศของแบบทดสอบ พบว่า ข้อสอบจำนวน 56 ข้อมีค่าสารสนเทศของแบบทดสอบกับความสามารถของผู้สอบ โดยให้สารสนเทศครอบคลุมความสามารถของผู้สอบ (θ) ตั้งแต่ - 2.8 ถึง 2.8 มีค่าสารสนเทศของแบบทดสอบโดยรวมอยู่ระหว่าง 1.409 – 30.169 มีความคลาดเคลื่อนในการวัดอยู่ระหว่าง 0.093 – 0.843 และสารสนเทศของแบบทดสอบจะมีค่าสูงเมื่อทำการวัดกับผู้สอบที่มีความสามารถ (θ) ตั้งแต่ 0.0 ถึง 2.0 โดยมีค่าสารสนเทศของแบบทดสอบอยู่ระหว่าง 8.233 – 30.169 มีความคลาดเคลื่อนในการวัดอยู่ที่ 0.093 – 0.349 แสดงว่าแบบทดสอบจะมีความถูกต้องแม่นยำสูงในการประมาณค่าความสามารถของผู้สอบเมื่อนำไปใช้กับผู้สอบที่มีระดับความสามารถ (θ) ในกลุ่มผู้มีความสามารถระดับปานกลางถึงระดับค่อนข้างสูง

ตาราง 5 ผลการวิเคราะห์ความแม่นยำในการประมาณค่าระดับความสามารถของแบบทดสอบมาตรฐานการฟัง
ภาษาอังกฤษตามทฤษฎีการตอบสนองข้อสอบแบบทดสอบฉบับที่ 1

ความสามารถของผู้สอบ	จำนวนข้อสอบ
- ระดับสูง	2
- ระดับค่อนข้างสูงถึงระดับสูง	1
- ระดับปานกลางถึงระดับสูง	12
- ระดับปานกลางถึงระดับค่อนข้างสูง	1
- ระดับค่อนข้างต่ำถึงระดับสูง	12
- ระดับต่ำถึงระดับปานกลาง	16
- ระดับต่ำถึงระดับสูง	12
รวม	56

จากตาราง 5 พบว่าข้อคำถามที่แสดงถึงความแม่นยำในการประมาณค่าระดับความสามารถของผู้สอบทุกระดับระดับ จำนวน 12 ข้อ และความแม่นยำในการประมาณค่าระดับความสามารถของผู้สอบระดับต่ำถึงระดับปานกลาง มีจำนวนข้อคำถามมากที่สุด จำนวน 16 ข้อ

ตาราง 6 ผลการวิเคราะห์ข้อมูลสารสนเทศของแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบแบบทดสอบฉบับที่ 2

ค่าความสามารถของ ผู้สอบ (θ)	ค่าสารสนเทศของ แบบทดสอบ (θ)	ความคลาดเคลื่อน มาตรฐาน	ระดับความสามารถการฟัง
2.8	7.142	0.374	ความสามารถสูง
2.4	13.745	0.27	
2	22.839	0.209	
1.6	36.979	0.164	ความสามารถค่อนข้างสูง
1.2	36.67	0.165	
0.8	36.47	0.166	
0.4	22.157	0.212	ความสามารถปานกลาง
0	16.932	0.243	
-0.4	13.378	0.271	
-0.8	11.192	0.299	ความสามารถค่อนข้างต่ำ
-1.2	9.585	0.323	
-1.6	7.765	0.359	
-2	5.823	0.414	ความสามารถต่ำ
-2.4	4.228	0.486	
-2.8	3.112	0.567	

จากตาราง 6 ผลการวิเคราะห์ค่าสารสนเทศของแบบทดสอบ พบว่า ข้อสอบจำนวน 76 ข้อมีค่าสารสนเทศของแบบทดสอบกับความสามารถของผู้สอบ โดยให้สารสนเทศครอบคลุมความสามารถของผู้สอบ (θ) ตั้งแต่ - 2.8 ถึง 2.8 มีค่าสารสนเทศของแบบทดสอบโดยรวมอยู่ระหว่าง 3.112 – 36.979 มีความคลาดเคลื่อนในการวัดอยู่ระหว่าง 0.164 – 0.567 และสารสนเทศของแบบทดสอบจะมีค่าสูงเมื่อทำการวัดกับผู้สอบที่มีความสามารถ (θ) ตั้งแต่ -0.8 ถึง 2.4 โดยมีค่าสารสนเทศของแบบทดสอบอยู่ระหว่าง 11.192 – 36.470 มีความคลาดเคลื่อนในการวัด

อยู่ที่ 0.164 – 0.299 แสดงว่าแบบทดสอบจะมีความถูกต้องแม่นยำสูงในการประมาณค่าความสามารถของผู้สอบ เมื่อนำไปใช้กับผู้สอบที่มีระดับความสามารถ (θ) ในกลุ่มผู้มีความสามารถระดับปานกลางถึงระดับค่อนข้างสูง

ตาราง 7 ผลการวิเคราะห์ความแม่นยำในการประมาณค่าระดับความสามารถของแบบทดสอบมาตรฐานการฟังภาษาอังกฤษตามทฤษฎีการตอบสนองข้อสอบแบบทดสอบฉบับที่ 2

ความสามารถของผู้สอบ	จำนวนข้อสอบ
- ระดับสูง	0
- ระดับค่อนข้างสูงถึงระดับสูง	4
- ระดับปานกลางถึงระดับสูง	14
- ระดับปานกลางถึงระดับค่อนข้างสูง	0
- ระดับค่อนข้างต่ำถึงระดับสูง	13
- ระดับต่ำถึงระดับปานกลาง	0
- ระดับต่ำถึงระดับสูง	45
รวม	76

จากตาราง 7 พบว่า ข้อคำถามส่วนใหญ่มีความแม่นยำในการประมาณค่าระดับความสามารถของผู้สอบทุกระดับ จำนวน 45 ข้อ และความแม่นยำในการประมาณค่าระดับความสามารถของผู้สอบระดับค่อนข้างสูงถึงระดับสูง มีจำนวนข้อคำถามน้อยที่สุด จำนวน 4 ข้อ

สรุปและอภิปรายผล

งานวิจัยนี้ดำเนินการพัฒนาแบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษ สำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม โดยเติมเต็มสิ่งที่ขาดไปในงานวิจัยก่อนหน้านี้ที่เกี่ยวข้องกับการพัฒนาข้อสอบในประเด็นเพิ่มกระบวนการคิดคำศัพท์ด้วยกระบวนการวิจัยแบบภาษาศาสตร์คลังข้อมูลที่ปรากฏในงานของณัฐกานต์ เสงี่ยม (2566) ซึ่งได้คำศัพท์ที่มีความถี่สูงและเทียบเคียงได้กับมาตรฐาน CEFR นอกจากนี้ ยังได้ปริมาณกลุ่มตัวอย่างที่ทำการทดสอบแบบทดสอบมาตรฐานวัดความสามารถด้านการฟังภาษาอังกฤษจำนวนมากเป็นกลุ่มตัวอย่างขนาดใหญ่ซึ่งเหมาะสมกับการใช้สถิติทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)

จากการวิจัยในครั้งนี้สามารถอภิปรายผลของการวิจัยได้ ดังนี้

1. การพัฒนาชุดข้อสอบการฟังภาษาอังกฤษมาตรฐานสำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงคราม ซึ่งต่อยอดผลการวิจัยของณัฐกานต์ เสงี่ยม (2566) โดยชุดข้อสอบการฟังภาษาอังกฤษมาตรฐานฉบับนี้ได้้นำคำศัพท์ที่มีความถี่สูงซึ่งเป็นคำที่มีประโยชน์ต่อผู้เรียนและเป็นคำศัพท์ที่ควรเลือกมาเรียนก่อน โดยทำการเลือกคำศัพท์ร้อยละ 95 ของคำศัพท์ที่มีความถี่สูงสุดทั้งหมดตามแนวคิดของ Laufer (1989) ที่กล่าวว่า การรู้คำศัพท์ร้อยละ 95 ในเรื่องที่อ่าน เพียงพอสำหรับการเดาความหมายของคำศัพท์ที่ไม่รู้ได้ ผู้วิจัยจึงกำหนดสัดส่วนจำนวนตามแนวคิดของ Laufer โดยให้สัดส่วนตามความถี่ของคำศัพท์แต่ละประเภทคำศัพท์ได้ดังนี้

1.1 คำศัพท์ที่ทั่วไป (General Service List GSL) ระดับ K1 คือคำศัพท์ที่มีความถี่สูงสุด 1,000 คำแรก จำนวน 785 คำ คิดเป็นร้อยละ 50 และคำศัพท์ที่ทั่วไป (General Service List GSL) ระดับ K2 คือคำศัพท์ที่มีความถี่สูงสุด 1,001-2,000 คำ จำนวน 233 คำ คิดเป็นร้อยละ 30 คำศัพท์วิชาการ (Academic Word List AWL) จำนวน 21 คำ คิดเป็นร้อยละ 5 และคำศัพท์เฉพาะ (Outside Word List OWL-OFF LIST) จำนวน 130 คำ คิดเป็นร้อยละ 10 รวมทั้งสิ้น 1,169 คำ คิดเป็นร้อยละ 95 และคัดเลือกคำศัพท์ทั่วไป ระดับ K1 ที่มีความถี่สูงสุดจำนวน 100 คำแรก คำศัพท์วิชาการ จำนวน 10 คำแรก และคำศัพท์เฉพาะจำนวน 30 คำแรก ที่อยู่ในระดับ A2 – B2 มาใช้สำหรับเป็นกรอบของการสร้างชุดข้อสอบมาตรฐานแบบปรนัย

1.2 นำแบบทดสอบแบบปรนัย จำนวน 200 ข้อ ไปให้ผู้เชี่ยวชาญตรวจสอบดัชนีความสอดคล้อง (IOC) โดยผู้ทรงคุณวุฒิทางด้านภาษาอังกฤษ ด้านการวัดและประเมินผล ที่มีประสบการณ์ด้านการสอนวิชาภาษาอังกฤษ รวมทั้งสิ้น 3 คน โดยผู้เชี่ยวชาญพิจารณาความสอดคล้องในแง่ของภาษาและประเด็นคำถาม และตัวเลือกตอบ ซึ่งพิจารณาจาก IOC ครมมากกว่าหรือเท่ากับ .80 และจากผลการวิเคราะห์สามารถนำข้อสอบไปใช้ได้ จำนวน 198 ข้อ และปรับปรุง จำนวน 2 ข้อ ซึ่งผู้วิจัยได้แก้ไขตามคำแนะนำของผู้เชี่ยวชาญ โดย Pan Yi-Ching (2009) ได้เห็นสอดคล้องว่า แบบทดสอบวัดความสามารถภาษาอังกฤษของนักศึกษาที่ออกแบบและสร้างขึ้นควรให้เหมาะสมและสอดคล้องกับผู้สอบเฉพาะกลุ่ม ซึ่งจะนำมาซึ่งแบบทดสอบที่มีความน่าเชื่อถือและมีความเที่ยงตรง และยิ่งไปกว่านั้น ผลการสอบจะสามารถเทียบเคียงคุณภาพกับมาตรฐานการอ้างอิงในระดับสากลได้อีกด้วย

2. การหาคุณภาพของแบบทดสอบวัดความสามารถด้านการฟังภาษาอังกฤษมาตรฐาน สำหรับนักศึกษา มหาวิทยาลัยราชภัฏพิบูลสงคราม วิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) พบว่า จำนวนข้อสอบที่มีสารสนเทศของแบบทดสอบจำนวน 132 ข้อ

2.1 แบบทดสอบฉบับที่ 1 จำนวน 56 ข้อ มีค่าสารสนเทศของแบบทดสอบกับความสามารถของผู้สอบ (θ) ตั้งแต่ -2.8 ถึง 2.8 ค่าสารสนเทศของแบบทดสอบโดยรวมอยู่ระหว่าง 1.409 ถึง 30.169 ความคลาดเคลื่อนการวัดอยู่ระหว่าง 0.093 ถึง 0.843 และแบบทดสอบฉบับที่ 2 จำนวน 76 ข้อ มีค่าสารสนเทศของแบบทดสอบกับความสามารถของผู้สอบ (θ) ตั้งแต่ -2.8 ถึง 2.8 มีค่าสารสนเทศของแบบทดสอบโดยรวมอยู่ระหว่าง 3.112 ถึง 36.979 มีความคลาดเคลื่อนการวัดอยู่ระหว่าง 0.164 ถึง 0.567 เมื่อวิเคราะห์ข้อมูลตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) พบว่าแบบทดสอบฉบับที่ 1 มีข้อสอบที่มีค่าอำนาจจำแนกอยู่ระหว่าง 0.3 – 2.5 มีจำนวน 47 ข้อ ข้อสอบที่มีค่าความยากอยู่ระหว่าง -3.00 ถึง +3.00 จำนวน 54 ข้อ และข้อสอบที่มีค่าโอกาสการเดาอยู่ต่ำกว่า 0.3 จำนวน 25 ข้อ เมื่อพิจารณารายข้อ โดยพิจารณาค่าอำนาจจำแนกและค่าความยาก ในแต่ละข้อ พบว่า มีข้อสอบที่ผ่านเกณฑ์สามารถนำไปใช้ได้จำนวน 47 ข้อ และข้อสอบที่ไม่เป็นไปตามเกณฑ์จำนวน 9 ข้อ

2.2 แบบทดสอบฉบับที่ 2 ข้อสอบที่มีค่าอำนาจจำแนกอยู่ระหว่าง 0.3 – 2.5 มีจำนวน 63 ข้อ ข้อสอบที่มีค่าความยากอยู่ระหว่าง -3.00 ถึง +3.00 จำนวน 75 ข้อ และข้อสอบที่มีค่าโอกาสการเดาอยู่ต่ำกว่า 0.3 จำนวน 87 ข้อ เมื่อพิจารณารายข้อ โดยพิจารณาค่าอำนาจจำแนกและค่าความยาก ในแต่ละข้อ พบว่า มีข้อสอบที่ผ่านเกณฑ์สามารถนำไปใช้ได้จำนวน 63 ข้อ และข้อสอบที่ไม่เป็นไปตามเกณฑ์จำนวน 13 ข้อ เมื่อพิจารณาข้อสอบที่ผ่านเกณฑ์ค่าความยาก และอำนาจจำแนกของทั้ง 2 ฉบับ จำนวน 68 ข้อ แบ่งเป็นข้อสอบในแบบทดสอบฉบับที่ 1 จำนวน 18 ข้อ และข้อสอบในแบบทดสอบฉบับที่ 2 จำนวน 50 ข้อ โดยมีค่าความยาก (b) ของข้อสอบ มีค่าตั้งแต่ 0.31 ถึง 2.47 คือข้อสอบที่ยากง่ายปานกลางไปจนถึงข้อสอบที่ค่อนข้างง่าย ค่าอำนาจจำแนก (a) ตั้งแต่ -2.23 ถึง 2.62 และค่าโอกาสการเดา (c) ตั้งแต่ 0.00 ถึง 0.29

โดยแบบทดสอบทั้ง 2 ฉบับมีความน่าจะเป็นที่จะตอบถูกด้วยการเดาไม่เกินร้อยละ 30 ซึ่งเป็นค่าที่อยู่ในเกณฑ์ที่ยอมรับได้ มีความสอดคล้องกับ เบญจมาภรณ์ เสนารัตน์ และคณะ (2559) ได้ศึกษาเรื่อง การพัฒนาเครื่องมือประเมินความสามารถทางการวิจัยการศึกษาของนักศึกษาครูมหาวิทยาลัยราชภัฏในภาคตะวันออกเฉียงเหนือ พบว่า ค่าโอกาสการเดามีค่าอยู่ในช่วง 0.10 ถึง 0.30 จะเป็นข้อสอบที่มีระดับความยากครอบคลุมระดับทุกช่วงความสามารถได้อย่างเหมาะสม โดยในระบบนี้ผู้วิจัยใช้หลักเกณฑ์ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ คือ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าโอกาสการเดา (c) มาใช้ในการจัดทำแบบทดสอบ เนื่องจากจะช่วยค้นหาหาความสามารถของผู้เรียนได้อย่างมีประสิทธิภาพมากที่สุด และการคัดเลือกข้อมูลชุดคำศัพท์ในการพัฒนาเป็นข้อสอบการฟัง โดยมีจำนวนคำและระดับความยากง่าย ตามเกณฑ์การแบ่งของ CEFR และคัดเลือกคำศัพท์ที่ได้จากงานวิจัยของ ญัฐกานต์ เส่งขึ้น (2566)

เพื่อให้แบบทดสอบมีเนื้อหาเหมาะสมกับผู้เข้าสอบ โดยไม่ง่ายหรือยากจนเกินไปของระดับความสามารถของผู้เข้าสอบ (Genesee & Upshur, 1996) และเมื่อพิจารณาความเชื่อมั่นของแบบทดสอบวัดความสามารถด้านการฟังภาษาอังกฤษมาตรฐาน สำหรับนักศึกษามหาวิทยาลัยราชภัฏพิบูลสงครามมีค่าความเชื่อมั่นสูง อาจจะมาจกแบบทดสอบความสามารถด้านการฟังภาษาอังกฤษมีความสมจริง (Authenticity) และการมีปฏิสัมพันธ์ (Interactiveness) เพราะเนื้อหาของแบบทดสอบเกี่ยวข้องกับการใช้ภาษาอังกฤษในชีวิตประจำวัน (Bachman & Palmer, 2010; Brown, 2004) ซึ่งสอดคล้องกับ Chen (2009) ในการหาคุณภาพของแบบทดสอบ ที่พบว่าแบบทดสอบวัดความสามารถทักษะการฟังภาษาอังกฤษระดับมหาวิทยาลัยของประเทศจีนทั้ง 2 ชุดข้อสอบ มีความสมจริง (Authenticity) มากพอสมควร และมีหัวข้อที่มีความหลากหลายและเป็นที่คุ้นเคยของผู้เข้าทดสอบอีกด้วย

ข้อเสนอแนะ

ข้อเสนอแนะทั่วไป

1. ควรมีการนำแบบทดสอบที่พัฒนาขึ้นไปใช้สอบเพื่อคัดเลือกบุคคลเข้าทำงาน หรือศึกษาต่อในระดับที่สูงขึ้นของมหาวิทยาลัยราชภัฏพิบูลสงครามด้วย
2. ควรพัฒนาจำนวนแบบทดสอบไว้ใช้มากกว่า 1 ฉบับไว้ในคลังข้อสอบ

ข้อเสนอแนะในการวิจัยครั้งต่อไป

ควรมีการศึกษาเปรียบเทียบความสามารถในการฟังของแบบทดสอบที่พัฒนาขึ้นกับคณะต่าง ๆ ว่าจะมีค่าสถิติแสดงการทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) กันหรือไม่

เอกสารอ้างอิง

- ณัฐกานต์ เสงี่ยม. (2566). การศึกษาค้นข้อมูลคำศัพท์ในแนวข้อสอบ TOEIC เพื่อการพัฒนารายการคำศัพท์สำหรับการสอบฟัง. *สัปดาห์: วารสารมนุษยศาสตร์และสังคมศาสตร์*, 29(1), 74-90.
- เบญจมาภรณ์ เสนารัตน์, บุญชม ศรีสะอาด และจริยา รักตรารัตน์. (2559). การพัฒนาเครื่องมือประเมินความสามารถทางการวิจัยการศึกษาของนักศึกษาครูมหาวิทยาลัยราชภัฏในภาคตะวันออกเฉียงเหนือ. *วารสารการวัดการศึกษา มหาวิทยาลัยมหาสารคาม*, 22(1), 164-182.
- ศิริชัย กาญจนวาสี. (2563). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 5). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สำนักงานคณะกรรมการส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม (สกสว.). (2562). *เอกสารประกอบการเสวนาโครงการวิจัย "ปัญหาความสามารถในการใช้ภาษาอังกฤษของคนไทย: กลยุทธ์และนวัตกรรมในการแก้ปัญหา"* วันอาทิตย์ที่ 6 ตุลาคม 2562 ณ โรงแรมเอทิส ลุมพินี.
- A Glossary of Measurement Terms Eric Digest. (1989). *American Institutes for Research Washington DC*. Retrieved May 10, 2023, from <http://ericae.net/digests/ed315430> ERIC Clearinghouse on Tests Measurement and Evaluation Washington DC.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. (86-89). Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Brown, H. D. (2004). *Language Assessment: Principle and Classroom Practices*. New York: Pearson Education.

- Chen, L. (2009). *Assessing the Test Usefulness: A Comparison Between the Old and the New College English Test Band 4 (CET-4) in China*. Retrieved May 2, 2023 from <http://www.diva-portal.org/smash/get/diva2:228454/FULLTEXT01.pdf>
- Education First. (2020). *EF English Proficiency Index: A Ranking of 100 Countries and Regions by English Skills*. Retrieved April 29, 2023 from <https://www.ef.com/wwen/epi/>
- Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Laufer, B. (1989). What Percentage of Text-Lexis Is Essential for Comprehension? In: Lauren, C. and Nordman, M., Eds., *Special Language: From Human Thinking to Thinking Machines*, Multilingual Matters, Clevedon, 316-323.
- Nitko, A. J. (1996). *Educational assessment of students*. Prentice-Hall Order Processing Center, PO Box 11071, Des Moines, IA 50336-1071.
- Pan, Y.-C. (2009). Evaluating the appropriateness and consequences of test use. *Colombian Applied Linguistics Journal*, (11), 93–105. <https://doi.org/10.14483/22487085.156>
- Rovinelli, R. J., & Hambleton, R. K. (1977) On the Use of Content Specialists in the Assessment of Criterion-Referenced Test Item Validity. *Tijdschrift Voor Onderwijs Research*, 2, 49-60.
- The European Association for Quality Language Services. (2009). *Calibrating tests to the CEFR: A simplified guide for EAQUALS Members on calibrating entry or progress tests to the CEFR*. Retrieved April 20, 2023, from http://clients.squareeye.net/uploads/eaquals2011/documents/EAQUALS_Calibrating_tests_to_the_CEFR.pdf