

**การตรวจสอบคุณภาพข้อสอบวัดสมิทธิภาพในการใช้ภาษาอังกฤษของ
มหาวิทยาลัยสุโขทัยธรรมมาธิราช ด้วยระบบคอมพิวเตอร์
ตามกรอบการอ้างอิงทางภาษาของสหภาพยุโรป**
**Validating Sukhothai Thammathirat Open University's Computer-Based
English Proficiency Test (STOU-EPT) Under the Common European
Framework of Reference for Languages (CEFR)**

ศจี จิระโร^{1*} และสุภมาส อังคุโชติ²

¹ ค.ด./ผู้ช่วยศาสตราจารย์ ดร. มหาวิทยาลัยสุโขทัยธรรมมาธิราช sajee.jir@stou.ac.th

² ค.ด./รองศาสตราจารย์ ดร. supamas.ang@stou.ac.th

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อตรวจสอบคุณภาพข้อสอบ STOU-EPT (E-Testing) ตามกรอบอ้างอิงทางภาษาของสหภาพยุโรป (The Common European Framework of Reference for Languages: CEFR) ในประเด็นเกี่ยวกับ (1) กระบวนการสร้างแบบทดสอบ (2) โครงสร้างแบบทดสอบแต่ละ part ตามเกณฑ์การสร้างข้อสอบของ CEFR (3) ความยากง่ายของข้อสอบรายข้อ (4) สัมประสิทธิ์ความเที่ยงของแบบทดสอบแต่ละ Part แหล่งข้อมูลในการวิจัยประกอบด้วย แหล่งข้อมูลปฐมภูมิ ได้แก่ ผู้เชี่ยวชาญที่มีประสบการณ์เกี่ยวกับการทดสอบภาษาอังกฤษตามกรอบ CEFR จำนวน 4 คน คณะกรรมการออกข้อสอบจำนวน 1 คน และแหล่งข้อมูลทุติยภูมิ ได้แก่ ผลการตอบข้อสอบของผู้เข้าสอบวัดความรู้ภาษาอังกฤษ STOU-EPT (E-Testing) จำนวน 800 คน ผลการวิจัยจำแนกตามประเด็นการตรวจสอบคุณภาพได้ดังนี้ (1) ขั้นตอนการพัฒนาแบบทดสอบมีการดำเนินการเริ่มจาก มหาวิทยาลัยได้ส่งตัวแทนคณะกรรมการออกข้อสอบเข้ารับการฝึกอบรมการพัฒนาแบบทดสอบมาตรฐานภาษาอังกฤษ สถาบันที่มีความเชี่ยวชาญ จากประเทศสหราชอาณาจักร และได้พัฒนารอบของแบบทดสอบ (test specifications) ในแต่ละสมรรถนะตามเกณฑ์ของ CEFR โดยมีผู้เชี่ยวชาญซึ่งเป็นวิทยากรอบรม เป็นผู้ตรวจสอบและให้ข้อเสนอแนะ (2) การพิจารณาโครงสร้างข้อสอบได้ผ่านความเห็นชอบจากผู้เชี่ยวชาญ เพื่อให้สอดคล้องกับกรอบของ CEFR และการดำเนินการของมหาวิทยาลัย การตรวจสอบคุณภาพของข้อสอบก่อนนำไปใช้ พิจารณาความถูกต้อง เหมาะสม และความสอดคล้องตามกรอบของ CEFR โดยนักวิชาการที่มีชื่อเสียงด้านการวัดและประเมินผลทางภาษาของประเทศไทย การใช้โปรแกรม Lexitutor ในการตรวจสอบความเหมาะสมของระดับภาษา และ AWL list ตรวจสอบความเหมาะสมของการใช้คำศัพท์ทางวิชาการ ของข้อสอบส่วนมากมีระดับสูงกว่าเกณฑ์ CEFR (3) ความยากของแบบทดสอบ พบว่า มีความยากอยู่ในเกณฑ์ใช้ได้ดีคิดเป็นร้อยละ 88.00-100.00 (4) ผลการวิเคราะห์สัมประสิทธิ์ความเที่ยงของแบบทดสอบ โดยสูตร KR-20 Part 3 Reading มีความเที่ยงสูงสุด ค่าความเที่ยงอยู่ระหว่าง .853 ถึง .893

คำสำคัญ : การสอบวัดสมิทธิภาพในการใช้ภาษาอังกฤษด้วยระบบคอมพิวเตอร์, การตรวจสอบคุณภาพข้อสอบวัดสมิทธิภาพในการใช้ภาษาอังกฤษ

Abstract

The purposes of this research study is validating Sukhothai Thammathirat Open University's English Proficiency Test or STOU-EPT (E-testing) under the Common European Framework of Reference for Languages (CEFR) in the terms of (1) test development (2) test structure according to the CEFR's test level criteria and (3) item difficulty analysis (4) reliability coefficient of each part of the test. The sources of data consisted of the primary sources including four experts with the experiences in English tests under the CEFR, a voluntary test development committee as well as the secondary sources including the answers from the forms of the STOU-EPT (e-testing) given by 800 examinees. The findings were as follows: (1) The test development began with a team of STOU test development committee members participating in the standard English test development training in the United Kingdom, which led to the development of the test specifications using the CEFR criteria under the guidance of an expert in English language testing, also a training leader, who checked test accuracy and congruence with university operation, the test specifications were approved. (2) A world-renowned scholar in the fields of language testing and evaluation was also involved in judging the consistency of the test items with the CEFR criteria. Then the Lexitutor program was used for checking the appropriateness of the language level, and the AWL list was used for checking the appropriateness of the academic words. Most of the test had higher level of items than the CEFR's criteria.. (3) Regarding the item difficulty of the test, it was found that item difficulty had ranging from 88.00% to 100.00% (4) It was found that the KR-20 reliability coefficient of Part 3 of Reading had the highest reliability ranging from .853 to .893.

Keywords: The English proficiency test with the computer system, validating the English proficiency test

บทนำ

มหาวิทยาลัยสุโขทัยธรรมาธิราชได้พัฒนาข้อสอบวัดความรู้ภาษาอังกฤษ STOU-EPT โดยสร้างข้อสอบให้สอดคล้องตามกรอบอ้างอิงทางภาษาของสหภาพยุโรป (The Common European Framework of Reference for Languages -- CEFR; Council of Europe, 2001) ซึ่งเป็นกรอบอ้างอิงทางภาษาที่เป็นสากล และสอดคล้องกับนโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษาของสำนักงานคณะกรรมการการอุดมศึกษา ตามประกาศคณะกรรมการการอุดมศึกษา พ.ศ. 2559 ข้อสอบที่พัฒนาขึ้นมีการวัดทักษะภาษาอังกฤษ 3 ด้าน ได้แก่ ทักษะการฟัง (Listening) ทักษะโครงสร้างไวยากรณ์ (Structure) และทักษะการอ่าน (Reading) การวัดคุณลักษณะดังกล่าว มีรูปแบบข้อคำถามและรูปแบบการตอบที่หลากหลาย รวมทั้งมีการวัดทักษะการฟัง ซึ่งต้องควบคุมคุณภาพเสียงให้เป็นไปในแนวทางเดียวกันทุกสนามสอบ เพื่อให้การสอบมีความเป็นมาตรฐาน และวัดความรู้ความสามารถของผู้เข้าสอบได้สอดคล้องกับคุณลักษณะตามแต่ละทักษะมากยิ่งขึ้น สำนักทะเบียนและวัดผลจึงได้พัฒนารูปแบบการทดสอบความรู้ภาษาอังกฤษ STOU-EPT จากการทดสอบแบบกระดาษ เป็นการสอบด้วยระบบอิเล็กทรอนิกส์ และมีชื่อเรียกการทดสอบนี้ว่า Sukhothai Thammathirat Open University English Proficiency Test (E-Testing) หรือที่เรียกสั้น ๆ ว่า STOU-EPT (E-Testing) โดยเป็นการสอบผ่านคอมพิวเตอร์ตามศูนย์สอบที่มหาวิทยาลัยจัดขึ้น (อลิสรา วาณิชดี และคณะ, 2551)

ข้อสอบที่อยู่ในคลังข้อสอบของ STOU-EPT (E-Testing) มีโครงสร้างข้อสอบเป็นไปตามกรอบอ้างอิงทางภาษาของสหภาพยุโรป (CEFR) ซึ่งได้รับการพัฒนาโดยคณาจารย์ผู้เชี่ยวชาญในการสร้างข้อสอบตามกรอบอ้างอิงทางภาษาของสหภาพยุโรป (CEFR) ผู้เข้าสอบจึงเชื่อมั่นในคุณภาพของข้อสอบได้ระดับหนึ่ง ทั้งนี้ เพื่อให้ข้อสอบมีความน่าเชื่อถืออย่างสมบูรณ์ตามหลักการทางวัดและประเมินผล จึงควรมีการตรวจสอบคุณภาพของข้อสอบเพื่อยืนยันความถูกต้อง เหมาะสมของข้อสอบตามกรอบอ้างอิงทางภาษาของสหภาพยุโรป (CEFR) และเพื่อให้กระบวนการดังกล่าวดำเนินการอย่างเป็นเหตุเป็นผล มีความถูกต้อง น่าเชื่อถือ ผ่านการตรวจสอบโดยผู้ทรงคุณวุฒิ และมีกระบวนการตรวจสอบคุณภาพข้อสอบกับเกณฑ์ของ CEFR โดยผู้ทรงคุณวุฒิภายนอก การดำเนินการดังกล่าว จึงควรดำเนินการผ่านกระบวนการวิจัย โดยนำหลักทางการวัดและประเมิน ผลที่เกี่ยวข้องกับการตรวจสอบคุณภาพ เป็นแนวทางในการดำเนินการทำให้ได้ข้อสอบที่ผ่านกระบวนการตรวจสอบคุณภาพเป็นไปตามกรอบอ้างอิงทางภาษาของสหภาพยุโรป (CEFR) โดยสมบูรณ์ เป็นที่ยอมรับของนักศึกษาและผู้สนใจทั่วไป สำหรับนำไปใช้ในการพัฒนาความรู้ความสามารถทางภาษาอังกฤษได้ รวมทั้งนำผลการสอบไปใช้อ้างอิงแก่หน่วยงานหรือสถานศึกษาอื่นได้

วัตถุประสงค์ของการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อตรวจสอบคุณภาพข้อสอบ STOU-EPT (E-Testing) ตามกรอบอ้างอิงทางภาษาของสหภาพยุโรป (The Common European Framework of Reference for Languages -- CEFR) ในประเด็นเกี่ยวกับ (1) กระบวนการสร้างแบบทดสอบ (2) โครงสร้างแบบทดสอบแต่ละ part ตามเกณฑ์การสร้างข้อสอบของ CEFR (3) ความยากง่ายของข้อสอบรายข้อ และ (4) สัมประสิทธิ์ความเที่ยงของแบบทดสอบแต่ละ Part

วิธีดำเนินการวิจัย

การดำเนินการวิจัยเป็นไปตามลำดับของวัตถุประสงค์การวิจัย ดังนี้ (1) สัมภาษณ์คณะกรรมการออกข้อสอบเกี่ยวกับกระบวนการสร้างข้อสอบ เพื่อตรวจสอบกระบวนการสร้างแบบทดสอบ (2) จัดประชุมเชิงปฏิบัติการผู้เชี่ยวชาญที่มีประสบการณ์เกี่ยวกับการทดสอบภาษาอังกฤษตามกรอบอ้างอิงทางภาษาของสหภาพยุโรปเพื่อตรวจสอบโครงสร้างของข้อสอบ STOU-EPT (E-Testing) กับกรอบอ้างอิงทางภาษาของสหภาพยุโรป (The Common European Framework of Reference for Languages -- CEFR) ที่วัดทักษะภาษาอังกฤษ 3 ด้าน ได้แก่ ทักษะการฟัง (listening) ทักษะโครงสร้างไวยากรณ์ (structure) และทักษะการอ่าน (reading) โดยผู้เชี่ยวชาญจากสถาบันภาษาจุฬาลงกรณ์มหาวิทยาลัย จำนวน 4 ท่าน ประชุมพิจารณาภาพรวมโครงสร้างของแบบทดสอบ STOU-EPT (E-Testing) โดยพิจารณาร่วมกับผลการวิเคราะห์ข้อสอบด้วย โดยกำหนดเกณฑ์ในการตรวจสอบระดับของคำศัพท์ที่ใช้โดย Cambridge Dictionary เมื่อผู้เชี่ยวชาญแต่ละท่านตรวจสอบเสร็จ มีการประชุมหาข้อสรุปเพื่อให้ได้ข้อยุติอีกครั้งหนึ่ง (3) นำแบบทดสอบไปทดลองสอบกับผู้เข้าสอบ จำนวน 800 คน เพื่อวิเคราะห์คุณภาพข้อสอบรายข้อ และวิเคราะห์สัมประสิทธิ์ความเที่ยงของแบบทดสอบ

รายละเอียดขั้นตอนการวิจัย กลุ่มตัวอย่าง วิธีดำเนินการวิจัย และการวิเคราะห์ข้อมูล ตามวัตถุประสงค์ของการวิจัย กล่าวโดยสรุปดังตาราง 1

ตาราง 1 ขั้นตอนการวิจัย กลุ่มตัวอย่าง วิธีดำเนินการวิจัย และการวิเคราะห์ข้อมูล ตามวัตถุประสงค์ของการวิจัย

วัตถุประสงค์ การวิจัย	ขั้นตอนการวิจัย	กลุ่มตัวอย่าง	วิธีดำเนินการวิจัย	การวิเคราะห์ข้อมูล
1. เพื่อตรวจสอบ กระบวนการสร้าง แบบทดสอบ	ตรวจสอบ กระบวนการสร้าง ข้อสอบ STOU-EPT	ประธานกรรมการ ออกข้อสอบ	สัมภาษณ์กรรมการออกข้อสอบ STOU-EPT (E-Testing) ในประเด็น เกี่ยวกับ ขั้นตอนการพัฒนาแบบทดสอบ การกำหนดโครงสร้างและระดับความ ยากง่ายของข้อสอบ การตรวจสอบ ความเหมาะสมของข้อสอบก่อนการ นำไปใช้	วิเคราะห์เนื้อหา
2. เพื่อตรวจสอบ โครงสร้าง แบบทดสอบ แต่ละ Part	ตรวจสอบคุณภาพ ข้อสอบในประเด็น เกี่ยวกับ โครงสร้าง แบบทดสอบ ระดับ ของคำศัพท์ ความ ถูกต้อง เหมาะสม สอดคล้องกับระดับ ความสามารถ ตั้งแต่ A2-B2	ผู้เชี่ยวชาญจาก สถาบันภาษา จุฬาลงกรณ์ มหาวิทยาลัย จำนวน 4 ท่าน	นำข้อสอบที่ผ่านการหาความยากง่าย ของข้อสอบแต่ละข้อ เสนอให้ผู้เชี่ยวชาญ ด้านภาษาอังกฤษและมีความเชี่ยวชาญ ด้าน CEFR พิจารณาโครงสร้าง แบบทดสอบ ระดับของคำศัพท์ ความถูกต้อง เหมาะสมสอดคล้องกับ ระดับความสามารถ ตั้งแต่ A2-B2	วิเคราะห์เนื้อหา
3. เพื่อตรวจสอบ ความยากง่ายของ ข้อสอบ 4. เพื่อตรวจสอบ สัมประสิทธิ์ความ เที่ยงของ แบบทดสอบ	วิเคราะห์คุณภาพ ข้อสอบรายข้อ และวิเคราะห์ สัมประสิทธิ์ความ เที่ยงของ แบบทดสอบ	บุคคลทั่วไป จำนวน 800 คน	นำแบบทดสอบไปทดลองสอบกับผู้เข้า สอบ จำนวน 800 คน	วิเคราะห์ค่าความยาก ตามทฤษฎี Classical Test Theory และวิเคราะห์ค่าความ เที่ยงของแบบทดสอบ จำแนกตามแบบ โดยใช้ สูตร KR-20

ผลการวิจัย

จากการดำเนินการวิจัยตามลำดับขั้นตอนดังกล่าว มีผลการวิจัยดังนี้

1. การตรวจสอบกระบวนการสร้างข้อสอบ STOU-EPT

จากการสัมภาษณ์กรรมการออกข้อสอบ STOU-EPT (E-Testing) ในประเด็นเกี่ยวกับ ขั้นตอนการพัฒนาแบบทดสอบ การกำหนดโครงสร้างและระดับความยากง่ายของข้อสอบ การตรวจสอบความเหมาะสมของข้อสอบก่อนการนำไปใช้ กระบวนการพัฒนาแบบทดสอบ กล่าวโดยสรุปได้ว่า กระบวนการสร้างข้อสอบมีการดำเนินการเริ่มจากมหาวิทยาลัยได้ส่งตัวแทน คณะกรรมการออกข้อสอบ เข้ารับการฝึกอบรมการพัฒนาแบบทดสอบมาตรฐานภาษาอังกฤษ ณ Centre for Research in English Language Learning and Assessment University of Bedfordshire สหราชอาณาจักร เนื้อหาในการฝึกอบรมมุ่งเน้น การทำความเข้าใจเกี่ยวกับเกณฑ์มาตรฐานของ CEFR และการพัฒนาแบบทดสอบมาตรฐานตามกรอบอ้างอิงทางภาษาของ สหภาพยุโรป (The Common European Framework of Reference for Languages หรือ CEFR) ทุกทักษะ และการกำหนด

เกณฑ์เทียบเคียงสำหรับแบบทดสอบกับเกณฑ์ของ CEFR และได้พัฒนากรอบของแบบทดสอบ (Test specifications) ในแต่ละสมรรถนะตามเกณฑ์ของ CEFR โดยมีผู้เชี่ยวชาญซึ่งเป็นวิทยากรอบรม เป็นผู้ตรวจสอบและให้ข้อเสนอแนะ

การพิจารณาโครงสร้างข้อสอบได้ผ่านความเห็นชอบจากผู้เชี่ยวชาญ เพื่อให้สอดคล้องกับกรอบของ CEFR และการดำเนินการของมหาวิทยาลัย โดยข้อสอบ STOU-EPT (E-testing) เป็นการวัดสมรรถนะ 3 ด้าน ประกอบด้วย การฟัง (listening) โครงสร้างไวยากรณ์ (structure) และการอ่าน (reading) จากนั้นตัวแทนคณะกรรมการออกข้อสอบ ได้ถ่ายทอดความรู้ให้แก่คณาจารย์ของมหาวิทยาลัย และแต่งตั้งคณะกรรมการออกข้อสอบขึ้น ได้พัฒนาแบบทดสอบ จำนวน 4 ฉบับ ฉบับละ 100 ข้อ รวม 400 ข้อ วัดสมรรถนะ 3 ด้าน ประกอบด้วย การฟัง (listening) โครงสร้างไวยากรณ์ (structure) และการอ่าน (reading)

การตรวจสอบคุณภาพของข้อสอบก่อนนำไปใช้ พิจารณาความถูกต้อง เหมาะสม และความสอดคล้องตามกรอบของ CEFR โดยนักวิชาการที่มีชื่อเสียงด้านการวัดและประเมินผลทางภาษาของสหราชอาณาจักร การใช้โปรแกรม Lexitutor ในการตรวจสอบความเหมาะสมของระดับภาษา และ AWL list ตรวจสอบความเหมาะสมของการใช้คำศัพท์ทางวิชาการ

2. การตรวจสอบโครงสร้างข้อสอบ

จากการสัมภาษณ์กรรมการออกข้อสอบ STOU-EPT (E-Testing) ในประเด็นเกี่ยวกับ โครงสร้างของข้อสอบพบว่า โครงสร้างข้อสอบให้น้ำหนักความสำคัญกับการวัดสมรรถนะด้านการอ่าน (reading) เป็นหลัก มีจำนวนคำถาม 40 ข้อ เนื่องจากเป้าหมายหลักของกลุ่มผู้เข้าสอบ STOU-EPT (E-testing) เป็นการสอบเพื่อเข้าศึกษาต่อระดับบัณฑิตศึกษา และการสอบวัดสมรรถนะสำหรับนักศึกษาปริญญาตรี ซึ่งเป็นกลุ่มที่ใช้สมรรถนะด้านการอ่านภาษาอังกฤษเพื่อการเรียนรู้เป็นหลัก และสมรรถนะที่ให้น้ำหนักความสำคัญรองลงมา คือ โครงสร้างไวยากรณ์ (structure) มีจำนวนคำถาม 35 ข้อ และสมรรถนะด้านการฟัง (listening) มีจำนวนคำถาม 25 ข้อ ตามลำดับ การสร้างข้อสอบและการกำหนดน้ำหนักคะแนนที่ใช้ในการกำหนดระดับความสามารถแต่ละด้าน เป็นไปตามแนวของ Cambridge Assessment English ซึ่งเป็นสถาบันทางภาษาของมหาวิทยาลัยแคมบริดจ์ที่ได้รับการยอมรับในระดับสากล

ผลการตรวจสอบระดับของข้อสอบตามเกณฑ์ CEFR ตามความเห็นของผู้เชี่ยวชาญ พบว่า Structure เป็น Part ที่มีร้อยละของความสอดคล้องกับระดับของ CEFR สูงที่สุด ฉบับที่ข้อสอบที่มีความสอดคล้องสูงสุด คิดเป็นร้อยละ 94.3 ส่วน Listening มีความสอดคล้องรองลงมา คิดเป็นร้อยละ 56.0 ส่วน Reading มีข้อสอบที่อยู่ในระดับสูงกว่าเกณฑ์ CEFR คิดเป็นร้อยละ 82.5 ทั้งนี้ ผู้เชี่ยวชาญได้ให้คำแนะนำ และข้อเสนอแนะในการปรับปรุงคุณภาพข้อสอบ กล่าวโดยสรุปว่า รูปแบบข้อสอบมีความหลากหลายและน่าสนใจ แต่ควรมีการปรับปรุงในแต่ละ Part เพื่อให้สอดคล้องกับระดับของ CEFR โดยควรทำ Listening track ให้ช้าลง สำหรับข้อสอบที่วัดระดับ A2 ความเร็วควรอยู่ที่ ประมาณ 20 คำ/นาที และควรพิจารณาตัด Listening scripts/reading texts โดยพิจารณาเพิ่มข้อคำถามในแต่ละ Listening track/reading text และทำให้ task type ง่ายลง เพื่อเพิ่มความตรงในการวัด Listening Skill | Reading texts ควรเลือกคำที่สอดคล้องกับระดับ โดยเทียบจากตัวอย่างความยาก/ง่ายของภาษา ตามเกณฑ์ของ CEFR

3. การตรวจสอบความยากง่ายของข้อสอบ

จากการนำแบบทดสอบไปทดลองสอบกับผู้เข้าสอบต่างฉบับกัน รวมจำนวน 800 คน วิเคราะห์ค่าความยากและแปลผลค่าความยาก โดยการจัดช่วงค่าความยาก (p) เป็น 5 ช่วง ได้แก่ 0-0.19 อยู่ในระดับยากมาก 0.20 – 0.40 อยู่ในระดับค่อนข้างยาก 0.41 – 0.60 อยู่ในระดับยากง่ายพอเหมาะ 0.61 – 0.80 อยู่ในระดับ ค่อนข้างง่าย และ 0.81 – 1.00 อยู่ในระดับ ง่ายมาก ระดับที่จัดว่าข้อสอบอยู่ในเกณฑ์ใช้ได้ ได้แก่ ค่อนข้างยาก ยากง่ายพอเหมาะ ค่อนข้างง่าย

การวัดความสามารถทาง Listening พบว่า ข้อสอบทั้ง 4 ฉบับ อยู่ในเกณฑ์ที่ใช้ได้ คิดเป็นร้อยละ 88.00 – 100.00 โดยฉบับที่ 1 มีข้อสอบที่ใช้ได้สูงสุด คิดเป็นร้อยละ 100.00 รองลงมาคือ ข้อสอบฉบับที่ 3 คิดเป็นร้อยละ 96.00 และ ฉบับที่ 4 คิดเป็นร้อยละ 92.00 ตามลำดับ

การวัดความสามารถทาง Structure พบว่า ข้อสอบทั้ง 4 ฉบับ อยู่ในเกณฑ์ที่ใช้ได้ คิดเป็นร้อยละ 88.57 – 91.43 โดยฉบับที่ 1 2 และ 4 มีข้อสอบที่ใช้ได้สูงสุด คิดเป็นร้อยละ 91.43 รองลงมาคือ ข้อสอบฉบับที่ 3 คิดเป็นร้อยละ 88.57

การวัดความสามารถทาง Reading พบว่า ข้อสอบทั้ง 4 ฉบับ อยู่ในเกณฑ์ที่ใช้ได้ คิดเป็นร้อยละ 92.50 – 100.00 โดยฉบับที่ 3 มีข้อสอบที่ใช้ได้สูงสุด คิดเป็นร้อยละ 100.00 รองลงมาคือ ข้อสอบฉบับที่ 1 คิดเป็นร้อยละ 95.00 และ ฉบับที่ 4 คิดเป็นร้อยละ 97.50 ตามลำดับ

ผลสรุปจำนวนข้อสอบที่มีความยากระดับต่าง ๆ ดังตาราง 2

ตารางที่ 2 ผลสรุปจำนวนข้อสอบ ตามระดับความยากของแบบทดสอบฉบับที่ 1-4 ในภาพรวม

Part	รายการ	ฉบับ 1	ฉบับ 2	ฉบับ 3	ฉบับ 4
		(ข้อ)	(ข้อ)	(ข้อ)	(ข้อ)
Listening (25 ข้อ)	ยากมาก	-			2
	ค่อนข้างยาก	12	2	8	9
	ยากง่ายพอเหมาะ	8	7	11	10
	ค่อนข้างง่าย	5	13	5	4
	ง่ายมาก	-	3	1	-
	ข้อสอบที่อยู่ในเกณฑ์ใช้ได้ (%)	25 (100%)	22 (88.00%)	24 (96.00%)	23 (92.00%)
Structure (35 ข้อ)	ยากมาก	3	-	2	3
	ค่อนข้างยาก	11	9	15	15
	ยากง่ายพอเหมาะ	16	15	13	11
	ค่อนข้างง่าย	5	8	3	6
	ง่ายมาก	-	3	2	-
	ข้อสอบที่อยู่ในเกณฑ์ใช้ได้ (%)	32 (91.43%)	32 (91.43%)	31 (88.57%)	32 (91.43%)
Reading (40 ข้อ)	ยากมาก	2			1
	ค่อนข้างยาก	13	6	14	17
	ยากง่ายพอเหมาะ	20	16	24	19
	ค่อนข้างง่าย	5	15	2	3
	ง่ายมาก	-	3	-	-
	ข้อสอบที่อยู่ในเกณฑ์ใช้ได้ (%)	38 (95.00%)	37 (92.50%)	40 (100.00%)	39 (97.50%)

4. การตรวจสอบสัมประสิทธิ์ความเที่ยงของแบบทดสอบ

การคำนวณค่าสัมประสิทธิ์ความเที่ยงของแบบทดสอบแต่ละ Part ใช้สูตร KR-20 มีความเที่ยงอยู่ระหว่าง .350 ถึง .893 โดย Part 3 Reading มีความเที่ยงสูงสุดทุกฉบับ ค่าความเที่ยงอยู่ระหว่าง .853 ถึง .893 โดยแบบทดสอบฉบับที่ 3 มีความเที่ยงสูงสุด เท่ากับ .893 ส่วน Part 1 listening (Part I) มีค่าความเที่ยงต่ำสุด เท่ากับ .350 ผลการวิเคราะห์ดังตาราง 3

ตาราง 3 ค่าสัมประสิทธิ์ความเที่ยงของแบบทดสอบแต่ละ Part

แบบทดสอบ	Part 1 Listening				Part 2	Part 3
	Part I (8 ข้อ)	Part IIA (6 ข้อ)	Part IIB (6 ข้อ)	Part III (5 ข้อ)	Structure (35 ข้อ)	Reading (40 ข้อ)
ฉบับที่ 1	.709	.694	.808	.729	.843	.871
ฉบับที่ 2	.736	.772	.846	.753	.844	.890
ฉบับที่ 3	.350	.615	.879	.734	.853	.893
ฉบับที่ 4	.529	.675	.794	.607	.795	.853

สรุปและอภิปรายผล

จากการตรวจสอบคุณภาพข้อสอบ STOU-EPT (E-Testing) ตามกรอบอ้างอิงทางภาษาของสหภาพยุโรป (The Common European Framework of Reference for Languages -- CEFR) กล่าวโดยสรุปได้ว่ากระบวนการพัฒนาข้อสอบ STOU-EPT (E-Testing) มีการดำเนินงานเป็นไปตามขั้นตอน โดยเริ่มจากการส่งตัวแทนเข้ารับการฝึกอบรมการพัฒนาแบบทดสอบกับสถาบันที่มีความเชี่ยวชาญด้านการพัฒนาการเรียนรู้และการประเมินความสามารถด้านภาษาอังกฤษตามกรอบอ้างอิงทางภาษาของสหภาพยุโรป (CEFR) ณ Centre for Research in English Language Learning and Assessment University of Bedfordshire (University of Bedfordshire, 2019) เพื่อให้ผู้ออกข้อสอบมีความรู้ความเข้าใจอย่างแท้จริงและนำมาประยุกต์ใช้กับการสอบตามบริบทของมหาวิทยาลัยสุโขทัยธรรมมาธิราชได้ จากนั้นได้พัฒนากฎของแบบทดสอบ (Test specifications) ในแต่ละสมรรถนะตามเกณฑ์ของ CEFR ที่ผ่านการตรวจสอบความถูกต้องโดยผู้เชี่ยวชาญซึ่งเป็นวิทยากรอบรมเป็นผู้ตรวจสอบและให้ข้อเสนอแนะ รวมถึงการพิจารณาโครงสร้างของข้อสอบที่ผ่านความเห็นชอบจากผู้เชี่ยวชาญด้วยเช่นกัน แล้วจึงดำเนินการออกข้อสอบโดยตัวแทนคณะกรรมการออกข้อสอบที่มีประสบการณ์จากการอบรมกับสถาบันดังกล่าวเป็นผู้ถ่ายทอดแนวทางในการพัฒนาข้อสอบให้แก่คณะกรรมการออกข้อสอบและได้พัฒนาแบบทดสอบขึ้นจนแล้วเสร็จโดยมีกระบวนการในการตรวจสอบคุณภาพของข้อสอบก่อนนำไปใช้จากนักวิชาการด้านการวัดและประเมินผลทางภาษาของประเทศสหราชอาณาจักร และมีการใช้โปรแกรม Lexitutor ในการตรวจความเหมาะสมของระดับภาษา และ AWL list ตรวจสอบความเหมาะสมของการใช้คำศัพท์ทางวิชาการ

เมื่อพิจารณาในประเด็นเกี่ยวกับโครงสร้างข้อสอบ กล่าวได้ว่า โครงสร้างข้อสอบที่ผ่านการตรวจสอบโดยผู้เชี่ยวชาญด้านการวัดและประเมินผลทางภาษาของประเทศสหราชอาณาจักร มีความสอดคล้องกับโครงสร้างข้อสอบของสถาบันที่ดำเนินการเกี่ยวกับการทดสอบทางภาษาที่ได้รับการยอมรับในระดับสากล ได้แก่ TOEFL, IELTS และ TOEIC โดยมีความสอดคล้องในด้านการทดสอบการฟัง การอ่าน ส่วนการเขียน และการพูด มีการปรับให้เหมาะสมกับวัตถุประสงค์หลักของกลุ่มผู้เข้าสอบ STOU-EPT (E-testing) ที่มุ่งเน้นสอบเพื่อเข้าศึกษาต่อระดับบัณฑิตศึกษา และการสอบวัดสมรรถนะสำหรับนักศึกษาปริญญาตรี จึงเน้นที่การอ่านเป็นหลัก และให้ความสำคัญรองลงมาในเรื่องโครงสร้างไวยากรณ์ และการฟัง ตามลำดับ เพื่อให้สอดคล้องกับบริบทของมหาวิทยาลัยที่มุ่งเน้นให้บริการแก่นักศึกษาและผู้สนใจ ด้วยอัตราค่าลงทะเบียนที่ไม่สูงจนเกินไป จึงไม่มีการทดสอบการเขียนและการพูด ซึ่งมีค่าใช้จ่ายในการดำเนินการค่อนข้างสูง สอดคล้องกับ Luoma (2004) ที่ได้กล่าวว่า การพูดเป็นทักษะทางภาษาที่ประเมินยากที่สุด เนื่องจากจะต้องวัดโดยการปฏิสัมพันธ์ในขณะนั้น การกำหนดเกณฑ์การประเมิน และความเชี่ยวชาญของผู้ประเมินเป็นสิ่งที่สำคัญมาก เนื่องจากจะส่งผลต่อความถูกต้องและความน่าเชื่อถือของการทดสอบ และจากการให้สัมภาษณ์ของประธานคณะกรรมการออกข้อสอบได้ผ่านการพิจารณาจากผู้เชี่ยวชาญด้านการวัดและประเมินผลทางภาษาแล้วว่าสามารถจัดทำแบบทดสอบตามโครงสร้างดังกล่าวได้

ทั้งนี้ ในการตรวจสอบระดับของข้อสอบตามเกณฑ์ CEFR โดยผู้เชี่ยวชาญจากสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ตรวจสอบระดับของคำศัพท์ของข้อสอบแต่ละข้อจาก Cambridge Dictionary และตรวจสอบ grammar, topic, discourse markers โดยพิจารณาจาก core inventory พบว่า ข้อสอบส่วนมากมีระดับที่สูงกว่าเกณฑ์ CEFR โดยในด้านการฟัง (listening) มีประเด็นเกี่ยวกับความเร็วของบทสนทนาของข้อสอบในระดับ A2 ที่เร็วเกินไป ด้านการอ่าน (reading) มีประเด็นเกี่ยวกับเนื้อหาที่ถามบางประเด็นเป็นเรื่องไกลตัวไม่สอดคล้องระดับ A2 ด้านโครงสร้างไวยากรณ์ (structure) มีประเด็นเกี่ยวกับคำศัพท์บางคำที่สูงเกินไปสำหรับกลุ่ม A2

เมื่อนำผลการตอบข้อสอบมาวิเคราะห์ค่าความยากและอำนาจจำแนก พบว่า แต่ละฉบับมีบางข้อยากกว่าเกณฑ์ที่กำหนด ส่วนข้อที่ง่ายอยู่ในเกณฑ์ที่กำหนดเป็นส่วนใหญ่ ทำให้ค่าอำนาจจำแนกต่ำสุดถึงระดับติดลบ แม้มีจำนวนข้อไม่มากนัก แต่ก็เป็นการบ่งบอกได้ว่าข้อสอบนั้นยากเกินไป จะจำแนกความสามารถของผู้เข้าสอบได้น้อย และบางข้ออาจทำให้ผู้เข้าสอบกลุ่มเก่งสับสน จึงตอบข้อสอบได้ถูกน้อยกว่าผู้เข้าสอบกลุ่มอ่อน อีกทั้งรูปแบบข้อสอบที่มีความหลากหลายอาจส่งผลต่อความยากง่าย เช่น ข้อสอบแบบ ถูก-ผิด ผู้เข้าสอบมีโอกาสเดาได้มากกว่าการตอบรูปแบบอื่น ส่วนข้อสอบแบบจับคู่ หากผู้เข้าสอบเลือกคำตอบผิดไปข้อใดข้อหนึ่งจะมีผลกระทบต่อตัวเลือกอื่นด้วย จึงทำให้ได้รับผลวิเคราะห์ว่าเป็นข้อสอบที่ค่อนข้างยาก ซึ่งให้ผลที่สอดคล้องกับความเห็นของผู้เชี่ยวชาญจากสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ที่ว่าข้อสอบส่วนมากมีระดับที่สูงกว่าเกณฑ์ CEFR ซึ่งความคลาดเคลื่อนของระดับคุณภาพข้อสอบตามที่ตรวจสอบได้ มีความแตกต่างจากระดับคุณภาพข้อสอบที่ผู้ออกข้อสอบได้กำหนดไว้ ทำให้เกณฑ์การประเมินตามระดับของ CEFR อาจเกิดความคลาดเคลื่อนได้ โดยเฉพาะกลุ่มของ A2 ที่ข้อสอบบางข้อค่อนข้างยากกว่าระดับที่ควรจะเป็น จากงานวิจัยของ EF (2019) และซินจิตต์ อธิวรกุล และจิรดา วุฒนิยากร (2561) ที่พัฒนามาตรฐานการวัดทักษะภาษาอังกฤษ ซึ่งมีการกำหนดจุดตัดของคะแนน ที่เป็นกระบวนการเชื่อมโยงความเห็นของผู้เชี่ยวชาญและการแปลความหมายคะแนนเพื่อกำหนดจุดตัดของคะแนนแต่ละระดับ จะทำให้จุดตัดของคะแนนที่ได้มีความเหมาะสมมากยิ่งขึ้น จึงเป็นแนวทางที่การสอบ STOU-EPT (E-Testing) ควรมีการดำเนินการต่อ หลังจากที่มีการเพิ่มข้อสอบเข้าคลังข้อสอบจนมีจำนวนที่มากพอแล้ว

ส่วนความเที่ยงของแบบทดสอบในทุกด้าน ทุกฉบับ โดยภาพรวมอยู่ในเกณฑ์ที่ใช้ได้ แสดงว่าแบบสอบมีความสอดคล้องของผลการวัดระหว่างรายข้อกับแต่ละด้าน ทั้งนี้ ด้านการอ่าน (Reading) มีค่าความเที่ยงสูงสุดทั้ง 4 ฉบับ อาจเนื่องมาจากมีจำนวนข้อมากที่สุดเมื่อเทียบกับด้านอื่น ๆ ส่งผลให้ค่าสัมประสิทธิ์ความเที่ยงสูงตามไปด้วย แต่โดยภาพรวมของด้านอื่น ๆ ความเที่ยงยังคงเป็นไปตามเกณฑ์ ส่วน Part 2 listening (Part I) มีความเที่ยงต่ำที่สุด สอดคล้องกับความเห็นของผู้เชี่ยวชาญที่เห็นว่าคำศัพท์อยู่ในระดับสูงกว่าระดับความสามารถตามเกณฑ์ CEFR ที่กำหนด บางคำเป็นศัพท์เฉพาะที่ต้องมีความรู้พื้นฐานเกี่ยวกับบริบทของเรื่องนั้นมาก่อน ไม่เพียงแต่จับใจความจากการฟังเท่านั้น จึงทำให้ผลการตอบรายข้อและรายด้านสอดคล้องกันน้อยกว่าฉบับอื่น ๆ

ข้อเสนอแนะ

จากผลการพิจารณาของผู้เชี่ยวชาญที่มีความเห็นว่าข้อสอบหลายข้อโดยเฉพาะระดับ A2 มีความยากที่สูงกว่าเกณฑ์ของ CEFR ในระดับ A2 และผลการวิเคราะห์ความยาก และอำนาจจำแนกที่พบว่าหลายข้อมีค่าความยากต่ำกว่าเกณฑ์ อีกทั้งการที่ข้อสอบมีรูปแบบหลากหลายบางรูปแบบอาจไม่เหมาะสมและส่งผลกระทบต่อคำตอบของผู้เข้าสอบ จึงควรมีการปรับปรุงข้อสอบ STOU-EPT (E-Testing) ตามข้อเสนอแนะของผู้เชี่ยวชาญเพื่อให้สามารถวัดได้ตรงตามกรอบของ CEFR มากยิ่งขึ้น

จากผลการสัมภาษณ์ประธานคณะกรรมการออกข้อสอบในประเด็นเกี่ยวกับการพัฒนาแบบสอบ จะเห็นได้ว่าการสอบ STOU-EPT (E-Testing) มีจำนวน 400 ข้อ แบ่งเป็น 3 part ในการสอบแต่ละครั้งจะจัดฉบับใหม่โดยระบบคอมพิวเตอร์สามารถจัดได้ทั้งหมด 64 แบบ ซึ่งได้เริ่มใช้ตั้งแต่ พ.ศ. 2560 เป็นต้นมา ข้อสอบได้ถูกทวนใช้เป็นระยะเวลาพอสมควรแล้ว ประกอบ

กับการได้รับความนิยมมากยิ่งขึ้น มีจำนวนผู้เข้าสอบเพิ่มขึ้นในทุกปี จึงควรมีการเพิ่มข้อสอบในคลังเพื่อให้มีจำนวนที่มากพอในการจัดสอบแบบทดสอบ ซึ่งเป็นกระบวนการหนึ่งของการพัฒนาคุณภาพข้อสอบด้วยเช่นกัน

กิตติกรรมประกาศ

การวิจัยนี้ได้รับทุนอุดหนุนการวิจัย ประเภททุนอุดหนุนการวิจัยทางด้านการศึกษากิจการทางไกลจากสถาบันวิจัยและพัฒนา มหาวิทยาลัยสุโขทัยธรรมิกราช ประจำปี 2561

เอกสารอ้างอิง

- ชื่นจิตต์ อธิวรกุล และจิรดา วุฒยากร. (2561). การเทียบผลคะแนนแบบทดสอบสมรรถภาพภาษาอังกฤษแห่งมหาวิทยาลัยศรีนครินทรวิโรฒกับกรอบมาตรฐานอ้างอิงสากลของยุโรป. *Suranaree Journal of Social Science*, 12(2): 69-84.
- อลิสรา วาณิชดี และคณะ. (2551). รายงานการวิจัยเรื่อง พัฒนาแบบทดสอบวัดความสามารถด้านการใช้ภาษาอังกฤษโดยใช้สื่ออิเล็กทรอนิกส์ในระบบการศึกษาทางไกล. นนทบุรี: มหาวิทยาลัยสุโขทัยธรรมิกราช.
- Bradberry, T. (2016, January 19). *Why attitude is more important than IQ*. <https://www.forbes.com/sites/travisbradberry/2016/01/19/why-attitude-is-more-important-than-iq/?sh=2412c9d63bd0>
- Brown, L. M. (2006). Social Intelligence the New Science of Success. *Business Book Review*, 23(1), 1-11.
- Cherniss, C. & Goleman, D. (2001). *The emotionally intelligent workplace: how to select for, measure, and improve Emotional intelligence in individuals, groups, and organizations*. San Francisco: Jossey-Bass.
- Council of Europe. (2001). *Common European Framework of reference for languages: learning, teaching, assessment*. Cambridge University Press.
- EF. (2019). *EFSET Technical background report*. Retrieved April 24, 2019 from <https://www.efset.org/th/english-score/cefr/>
- Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. New York: Bantam Books.
- Hay Group. (2011). *The emotional and social competency inventory (ESCI) user guide*. Hay Group.
- Korn Ferry. (2017). *Emotional and social competency inventory (ESCI) research guide and technical manual*. Korn Ferry.
- Landry, L. (2019). *Why Emotional intelligence is important in leadership*. <https://online.hbs.edu/blog/post/emotional-intelligence-in-leadership>
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Morrison, N. (2013, December 30). *The eight characteristics of effective school leaders*. <https://www.forbes.com/sites/nickmorrison/2013/12/30/the-eight-characteristics-of-effective-school-leaders/#21a4bf069762>
- Patterson, S. E. (2008). Primal leadership: realizing the power of Emotional intelligence [review] / Goleman, D., Boyatzis, R., and McKee, A.. *Journal of applied Christian leadership*, 2(2), 76-80.
- The Consortium for Research on Emotional Intelligence in Organizations. (2012). *The emotional competence framework*. http://www.eiconsortium.org/pdf/emotional_competence_framework.pdf
- University of Bedfordshire. (2019). *Centre for research in english language learning and assessment*. Retrieved April 24, 2019 from <https://www.beds.ac.uk/crella>

