

การพัฒนาแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย

จตุพงศ์ โมรา บำรุง โตรัตน์ เสงี่ยม โตรัตน์ และ ปราณี นิลกรณ์

คณะศึกษาศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์ นครปฐม

DEVELOPMENT OF A UNIVERSITY ENGLISH PROFICIENCY TEST

Jatupong Mora, Bamrung Torut, Sa-ngiam Torut and Pranee Nilakorn

Faculty of Education, Silpakorn University, Sanamchandra Palace Campus, Nakhonpathom

E-mail address: mjatupong@tsu.ac.th

วันที่รับบทความ (Received) 30 มิถุนายน 2562

วันที่ได้รับบทความฉบับแก้ไข (Revised) 23 ธันวาคม 2564

วันที่ตอบรับบทความ (Accepted) 4 กรกฎาคม 2562

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์หลัก 2 ข้อ คือ 1) เพื่อสร้างและพัฒนาแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยให้เป็นแบบทดสอบมาตรฐานที่มีลักษณะเฉพาะคือ คุณประโยชน์ของแบบทดสอบเป็นไปตามเกณฑ์ที่กำหนด คุณภาพรายข้อและความเชื่อมั่นเป็นไปตามเกณฑ์ที่กำหนด และความตรงร่วมสมัย (Concurrent Validity) สูง เมื่อเทียบกับแบบทดสอบมาตรฐาน และ 2) เพื่อศึกษาความพึงพอใจของผู้เกี่ยวข้องต่อแบบทดสอบที่สร้างขึ้น กลุ่มตัวอย่างได้แก่อาสาสมัครที่เป็นนิสิตระดับปริญญาตรี มหาวิทยาลัยทักษิณ วิทยาเขตสงขลา ซึ่งกำลังศึกษาอยู่ในภาคเรียนที่ 2 ปีการศึกษา 2561 จำนวน 45 คน วิเคราะห์ข้อมูลโดยใช้สถิติพื้นฐาน การวิเคราะห์คุณภาพรายข้อและความเชื่อมั่น และการหาความตรงร่วมสมัยโดยใช้ Pearson's Product-Moment Correlation Coefficients ผลการวิจัยพบว่า

1) แบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยมีคุณภาพเบื้องต้นโดยรวมอยู่ในระดับดี ($\bar{X} = 4.26$, S.D. = .175) และมีคุณภาพรายด้าน อยู่ในระดับดีมาก 1 ด้าน คือ ด้านความเป็นธรรม (Fairness) ($\bar{X} = 4.66$, S.D. = .491) และระดับดี 6 ด้าน คือ 1) ด้านความตรงเชิงโครงสร้าง (Construct Validity) ($\bar{X} = 4.24$, S.D. = .121) 2) ด้านผลกระทบ (Impact) ($\bar{X} = 4.24$, S.D. = .164) 3) ด้านความสมจริง (Authenticity) ($\bar{X} = 4.20$, S.D. = .084) 4) ด้านความเชื่อมั่น (Reliability) ($\bar{X} = 4.19$, S.D. = .143) 5) ด้านการมีปฏิสัมพันธ์ (Interactiveness) ($\bar{X} = 4.18$, S.D. = .109) และ 6) ด้านความสะดวกใช้ (Practicality) ($\bar{X} = 4.14$, S.D. = .113)

2) แบบทดสอบปรนัย 150 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .55$) ค่าอำนาจจำแนกอยู่ในระดับจำแนกได้ดี ($r = .49$) และค่าความเชื่อมั่นระดับดี ($KR_{20} = .83$) ได้ข้อคำถามคุณภาพดี 122 ข้อ คิดเป็น 81.33 % และข้อคำถามที่ควรปรับปรุง 28 ข้อ คิดเป็น 18.67 % และแบบทดสอบอัตนัย 5 ข้อ มีค่าความยาก

ง่ายปานกลาง ($p = .55$) ค่าอำนาจจำแนกอยู่ในระดับจำแนกได้ดี ($r = .38$) และค่าความเชื่อมั่นระดับดี (Coefficient Alpha = .80) และเป็นข้อคำถามคุณภาพดีทั้ง 5 ข้อ คิดเป็น 100 %

3) ในภาพรวม แบบทดสอบทั้งฉบับมีค่าความตรงร่วมสมัย (Concurrent Validity) สูงกว่าเกณฑ์ ($r_{xy} = .90$) และมีความตรงร่วมสมัย (Concurrent Validity) ของแต่ละทักษะ เรียงตามลำดับ คือ ทักษะการเขียนและการพูด ($r_{xy} = .95$) การฟัง ($r_{xy} = .93$) การอ่าน ($r_{xy} = .87$) และการใช้ภาษา ($r_{xy} = .76$)

4) ในภาพรวม ผู้มีส่วนเกี่ยวข้องพึงพอใจแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยอยู่ในระดับมาก ($\bar{X} = 4.27$, S.D. = 0.770) และมีความพึงพอใจด้านต่าง ๆ เรียงตามลำดับ คือ 1) ด้านประโยชน์ของแบบทดสอบ ($\bar{X} = 4.40$, S.D. = 0.687) 2) ด้านคำสั่งของแบบทดสอบ ($\bar{X} = 4.36$, S.D. = 0.708) 3) ด้านการจัดการสอบ ($\bar{X} = 4.32$, S.D. = 0.842) 4) ด้านการตรวจให้คะแนน ($\bar{X} = 4.25$, S.D. = 0.784) 5) ด้านรูปแบบและการจัดลำดับความ ($\bar{X} = 4.22$, S.D. = 0.813) และ 6) ด้านเนื้อหาสาระของแบบทดสอบ ($\bar{X} = 4.19$, S.D. = 0.786)

คำสำคัญ: การพัฒนาแบบทดสอบ แบบทดสอบวัดความสามารถภาษาอังกฤษ แบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย

Abstract

Two main objectives of this research were 1) to construct and develop a university English proficiency test to be a standardized one with three specified characteristics: test usefulness was at a specified level, item analysis indexes and reliability were at a specified level, and there was high concurrent validity when compared the scored with a standardized test, and 2) to study the stake holders' satisfaction towards the test. The sample consisted of 45 Bachelor's degree students studying in the second semester of 2019 academic year at Thaksin University, Songkhla Campus. Basic statistics, item analysis and reliability coefficients, and Pearson's product moment correlation coefficients were computed to analyze the data. It was found that:

1) overall, test usefulness of UEPT was at a good level ($\bar{X} = 4.26$, S.D. = .175), and Fairness ($\bar{X} = 4.66$, S.D. = .491) was rated at the best level, while the other 6 qualities were at the good level: Construct Validity ($\bar{X} = 4.24$, S.D. = .121), Impact ($\bar{X} = 4.24$, S.D. = .164), Authenticity ($\bar{X} = 4.20$, S.D. = .084), Reliability ($\bar{X} = 4.19$, S.D. = .143), Interactiveness ($\bar{X} = 4.18$, S.D. = .109), and Practicality ($\bar{X} = 4.14$, S.D. = .113)

2) 150 objective test items had moderate difficulty index ($p = .55$), their discrimination was at a good level ($r = .49$), and the reliability was at a good level ($KR_{20} = .83$); 122 test items

(81.33%) were collected for future use and 28 items (18.67%) needed improvement or reconstruction. 5 subjective test items also had moderate difficulty index ($p = .55$), their discrimination was at a good level ($r = .38$), and the reliability coefficient was at a good level (Cronbach Alpha = .80) and all items were collected.

3) overall, the concurrent validity of UEPT was higher than the set value ($r_{xy} = .90$) and the concurrent validity of each test section could be put in this order: Writing and Speaking ($r_{xy} = .95$), Listening ($r_{xy} = .93$), Reading ($r_{xy} = .87$), and Language Use ($r_{xy} = .76$).

4) overall, the stake holders' satisfaction towards UEPT was at a much satisfied level ($\bar{X} = 4.27$, S.D. = 0.770), and the satisfaction towards each part could be put in this order: 1) test usefulness ($\bar{X} = 4.40$, S.D. = 0.687), 2) test instructions ($\bar{X} = 4.36$, S.D. = 0.708), 3) test administration ($\bar{X} = 4.32$, S.D. = 0.842), 4) Scoring ($\bar{X} = 4.25$, S.D. = 0.784), 5) test format and arrangement ($\bar{X} = 4.22$, S.D. = 0.813), and 6) test content ($\bar{X} = 4.19$, S.D. = 0.786).

Keywords: Test Development, English Proficiency Test, University English Proficiency Test

บทนำ

ในปัจจุบัน การรับสมัครงานบางตำแหน่งมักจะขอให้ผู้สมัครแสดงหลักฐานผลการสอบวัดความสามารถภาษาอังกฤษมาตรฐานสากลแนบพร้อมหลักฐานประกอบการสมัครงานอื่น ๆ ที่ได้ระบุไว้ในประกาศรับสมัครงาน หรือในบางตำแหน่งงาน เช่น โครงการรับสมัครครูคืนถิ่นของ สำนักงานคณะกรรมการการอุดมศึกษา (สำนักงานคณะกรรมการการอุดมศึกษา, 2561) ได้กำหนดให้ผู้สมัครที่ผ่านการคัดเลือกนำผลการสอบวัดความสามารถภาษาอังกฤษมายื่นก่อนวันบรรจุหรือภายในระยะเวลาที่หน่วยงานกำหนด และนอกจากนี้ ในวันที่ 12 เมษายน 2559 สำนักงานคณะกรรมการการอุดมศึกษา (2559) ได้ออกประกาศ เรื่อง นโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา ที่ระบุให้สถาบันอุดมศึกษาพิจารณาจัดให้นิสิตนักศึกษาทุกคน ทดสอบความรู้ตามแบบทดสอบมาตรฐานระดับอุดมศึกษาที่สถาบันสร้างขึ้น หรือที่เห็นสมควรจะนำมาใช้วัดสมรรถภาพทางภาษาอังกฤษ (English Proficiency) โดยสามารถเทียบเคียงผลกับ Common European Framework of Reference for Languages (CEFR) หรือมาตรฐานอื่น เพื่อให้ทราบระดับความสามารถของนิสิตนักศึกษาแต่ละคน และสถาบันอุดมศึกษาอาจพิจารณานำผลการทดสอบความรู้ทางภาษาอังกฤษบันทึกในใบรับรองผลการศึกษาหรือจัดทำเป็นประกาศนียบัตร โดยเริ่มตั้งแต่ปีการศึกษา 2559 เป็นต้นไป

มหาวิทยาลัยทักษิณตอบรับกระแสสังคมที่ให้ความสำคัญกับการมีผลการทดสอบวัดความสามารถภาษาอังกฤษของผู้สมัครงาน และขานรับนโยบายการยกระดับมาตรฐานภาษาอังกฤษของผู้เรียน โดยผู้บริหารได้กำหนดนโยบายเกี่ยวกับการทดสอบวัดความสามารถของผู้เรียนไว้ในแผนยุทธศาสตร์การพัฒนา

มหาวิทยาลัยทักษิณ พ.ศ. 2558 – 2567 โดยกำหนดไว้ในประเด็นยุทธศาสตร์ที่ 1 การพัฒนาความเป็นเลิศในการผลิตบัณฑิตคุณภาพระดับสากล และความเป็นเลิศในการจัดการศึกษาระดับนานาชาติ ตามตัวบ่งชี้ TSU02 นิสิตที่ผ่านเกณฑ์สมรรถนะสากลด้านภาษา และได้นิยาม สมรรถนะด้านภาษา ว่าเป็นสมรรถนะทักษะสากลด้านภาษาอังกฤษ โดยมีเกณฑ์การทดสอบตามมาตรฐานที่มหาวิทยาลัยกำหนด หรือมีผลการทดสอบ Placement Test ของโปรแกรม English Discoveries Online ระดับ CEFR B1 (ระดับ Intermediate 1 สำหรับนิสิตระดับปริญญาตรี หรือระดับ Intermediate 2 สำหรับนิสิตระดับบัณฑิตศึกษา) และผลการทดสอบมาตรฐานด้านภาษาของนิสิต สามารถใช้นับซ้ำสำหรับจำนวนนิสิตที่ผ่านเกณฑ์ได้จนกระทั่งถึงนิตจบการศึกษา (มหาวิทยาลัยทักษิณ, 2559) จากการกำหนดเกณฑ์ดังกล่าว มีผลให้คณะ สาขา และหลักสูตรต่าง ๆ เริ่มต้นผลักดันให้นิสิตเห็นความสำคัญ และทดลองสอบวัดความสามารถภาษาอังกฤษมากขึ้น และด้วยมหาวิทยาลัยทักษิณยังไม่มีแบบทดสอบวัดความสามารถภาษาอังกฤษเป็นของตนเอง จึงทำให้นิสิตต้องเสียค่าใช้จ่าย ๆ ในการสมัครสอบที่ค่อนข้างสูงและอาจสร้างภาระแก่ผู้ปกครองได้ นอกจากนี้การสอบวัดความสามารถภาษาอังกฤษด้วยแบบทดสอบมาตรฐาน อาจจะไม่เหมาะสมกับระดับพื้นฐานความรู้ทางภาษาของนิสิต โดยส่วนใหญ่ทั้งด้านบริบทและเนื้อหาสาระของแบบทดสอบ มากกว่าการสร้างแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยขึ้นใช้เอง นอกจากจะช่วยลดค่าใช้จ่ายในการสมัครสอบของนิสิตแล้ว เนื้อหาสาระและบริบทแบบทดสอบสามารถควบคุมให้อยู่ในระดับเดียวกับพื้นฐานความรู้ของผู้เข้าสอบโดยส่วนใหญ่ได้

ปัจจุบันนี้การใช้เครื่องมือวัดความสามารถด้านภาษาอังกฤษระดับมหาวิทยาลัยยังไม่อิงมาตรฐานสากล เพราะแบบทดสอบที่ใช้เป็นแบบทดสอบที่สร้างขึ้นเพื่อใช้ภายในแต่ละมหาวิทยาลัย แบบทดสอบบางชนิดวัดเพียงบางทักษะ ไม่สะท้อนความสามารถในการใช้ภาษาอังกฤษในขั้นการสื่อสาร แบบทดสอบที่มีอยู่ส่วนใหญ่ใช้เพื่อคัดเลือกบุคคลเข้าศึกษาต่อในระดับบัณฑิตศึกษา และเป็นแบบทดสอบวัดความสามารถก่อนจบการศึกษาสำหรับนักศึกษาระดับบัณฑิตศึกษามากกว่าที่จะให้นักศึกษาระดับปริญญาตรี สอบด้วย และที่สำคัญแบบทดสอบดังกล่าวยังไม่ได้เทียบคุณภาพกับแบบทดสอบมาตรฐานสากล ทำให้นักศึกษาไม่สามารถใช้ผลการสอบเพื่ออ้างอิงความสามารถภาษาอังกฤษในระดับสากลได้ สำหรับแบบทดสอบวัดความสามารถภาษาอังกฤษมาตรฐานสากลที่ใช้กันอยู่ เช่น TOEIC, TOEFL, และ IELTS นั้นได้มีการเทียบเคียงคุณภาพระหว่างกันแล้ว เพื่อให้สามารถอ้างอิงผลการสอบแทนกันได้ และในปัจจุบันแบบทดสอบมาตรฐานทุกชนิดควรได้รับการเทียบคุณภาพกับกรอบอ้างอิงความสามารถด้านภาษาของสภายุโรป (CEFR) ซึ่งเป็นเกณฑ์ที่ได้รับการยอมรับและเป็นที่รู้จักกันอย่างกว้างขวาง ดังนั้นหากมีการพัฒนาแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยที่สามารถเทียบเคียงคุณภาพกับแบบทดสอบมาตรฐานสากล ได้รับการเทียบเคียงคุณภาพกับกรอบอ้างอิงความสามารถด้านภาษาของสภายุโรป (CEFR) แล้ว จะทำให้นิสิตที่ผ่านการสอบดังกล่าวมีความสามารถภาษาอังกฤษเสมือนผ่านการสอบวัดความสามารถด้วยแบบทดสอบมาตรฐานสากลและความสามารถภาษาอังกฤษของบัณฑิตจะได้รับการยอมรับในระดับสากลอีกด้วย (Arunee Wiriyaichitra, n.d.; Pan, 2009; Paltridge, 2007; Gao, 2010; Jin, 2011; Arkoudis, et al 2009; The European Association for Quality Language Services, 2009)

งานวิจัยที่ผ่านมาด้านการทดสอบความสามารถภาษาอังกฤษระดับมหาวิทยาลัยยังมีช่องว่างที่ควรได้รับการเติมเต็ม 3 ประการ คือ 1) แบบทดสอบที่สร้างขึ้นวัดความสามารถของผู้เข้าสอบไม่ครบทุกทักษะ ไม่สะท้อนความสามารถในการใช้ภาษาเพื่อการสื่อสารได้อย่างครบถ้วน 2) การเลือกใช้แบบทดสอบวัดความสามารถภาษาอังกฤษในระดับสากลมาวัดความสามารถด้านภาษาอังกฤษของนักศึกษาไทย อาจเกิดความลำเอียงด้านวัฒนธรรมได้ เพราะแบบทดสอบในระดับสากลอาจใช้บริบทของเนื้อหาสาระข้อสอบที่ไม่เป็นที่คุ้นเคยของนักศึกษาไทย และ 3) ยังไม่มีการวิเคราะห์สาระแบบทดสอบที่สร้างขึ้นจากกรอบอ้างอิงความสามารถในการเรียนภาษาของสภายุโรป (CEFR) ดังนั้น แบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยที่ผู้วิจัยได้พัฒนา มีคุณสมบัติเด่น 3 ประการ คือ 1) เป็นแบบทดสอบที่วัดความสามารถของผู้เข้าสอบครบทุกทักษะ สามารถสะท้อนความสามารถในการใช้ภาษาอังกฤษเพื่อการสื่อสารได้อย่างแท้จริง 2) เนื้อหาสาระของแบบทดสอบไม่มีความลำเอียงด้านวัฒนธรรมและเหมาะสมกับผู้เข้าสอบที่เป็นชาวไทย และ 3) มีสาระแบบทดสอบสัมพันธ์กับกรอบอ้างอิงความสามารถในการเรียนภาษาของสภายุโรป (CEFR) โดยสามารถสรุปเป็นวัตถุประสงค์การวิจัย 2 ข้อ คือ

1) เพื่อสร้างและพัฒนาแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย (University English Proficiency Test) ให้เป็นแบบทดสอบมาตรฐานที่มีลักษณะเฉพาะ คือ

- คุณประโยชน์ของแบบทดสอบเป็นไปตามเกณฑ์ที่กำหนด
- คุณภาพรายข้อและความเชื่อมั่นเป็นไปตามเกณฑ์ที่กำหนด
- ความตรงร่วมสมัย (Concurrent Validity) สูง เมื่อเทียบกับแบบทดสอบมาตรฐาน

2) เพื่อศึกษาความพึงพอใจของผู้เกี่ยวข้องต่อแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย (University English Proficiency Test)

วิธีการดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเพื่อพัฒนาแบบทดสอบ เก็บรวบรวมข้อมูลโดยมีรายละเอียดดังนี้

ประชากรและกลุ่มตัวอย่าง

ประชากรในการศึกษาครั้งนี้ เป็นนิสิตระดับปริญญาตรี ภาคปกติ ชั้นปีที่ 1-4 และกำลังศึกษาอยู่ในภาคเรียนที่ 2 ปีการศึกษา 2561 จำนวน 8,167 คน ใน 5 คณะ คือ 1) คณะมนุษยศาสตร์และสังคมศาสตร์ 2) คณะศึกษาศาสตร์ 3) คณะศิลปกรรมศาสตร์ 4) คณะเศรษฐศาสตร์และบริหารธุรกิจ และ 5) คณะนิติศาสตร์ ส่วนกลุ่มทดลองใช้แบบทดสอบและกลุ่มตัวอย่างในการศึกษาครั้งนี้ได้มาโดยการรับสมัครนิสิตที่มีความสนใจสอบวัดความสามารถภาษาอังกฤษ กลุ่มทดลองใช้แบบทดสอบ หมายถึง ผู้เข้าสอบในครั้งที่ 1-2 ของการทดลองใช้แบบทดสอบที่สร้างขึ้น และกลุ่มตัวอย่าง หมายถึง ผู้เข้าสอบในครั้งที่ 3 ซึ่งเป็นการใช้แบบทดสอบฉบับจริงหลังจากได้ปรับปรุงคุณภาพแล้ว ผู้วิจัยดำเนินการทดลองและเก็บข้อมูลระหว่างเดือนมกราคม – มีนาคม 2562

ตัวแปรที่จะใช้ในการวิจัย ประกอบด้วย

1) คะแนนที่ได้จากการทดสอบทักษะการฟัง การใช้ภาษา การอ่าน การเขียนและการพูดของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย (University English Proficiency Test) และแบบทดสอบวัดความสามารถภาษาอังกฤษมาตรฐาน (แบบทดสอบปรนัยใช้คะแนนจากแบบทดสอบ TOEFL และแบบทดสอบอัตนัยใช้คะแนนจากแบบทดสอบ IELTS)

2) ความพึงพอใจของผู้เชี่ยวชาญ อาจารย์ผู้สอนภาษาอังกฤษ และผู้เข้าสอบต่อแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย (University English Proficiency Test) ที่ประเมินโดยใช้แบบประเมินความพึงพอใจของผู้เกี่ยวข้องต่อแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย (University English Proficiency Test)

เครื่องมือที่ใช้ในการวิจัย มี 5 ชนิด ได้แก่

1) แบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย ซึ่งวัดความสามารถผู้เข้าสอบ 5 ทักษะ คือ การฟัง การใช้ภาษา การอ่าน การเขียนและการพูด โดยมีรายละเอียด คือ

แบบทดสอบการฟัง วัดความสามารถในการฟังสาระในชีวิตประจำวันและเป็นที่คุ้นเคย เป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก จำนวน 50 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .41$) ค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = .44$) และค่าความเชื่อมั่นระดับยอมรับได้ ($KR_{20} = .74$)

แบบทดสอบการใช้ภาษา วัดความสามารถในการเข้าใจภาษาเขียนมาตรฐานและระบุข้อผิดพลาดในการใช้ภาษา เป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .50$) ค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = .48$) และค่าความเชื่อมั่นระดับยอมรับได้ ($KR_{20} = .71$)

แบบทดสอบการอ่าน วัดความสามารถในการอ่านสาระในชีวิตประจำวันและเชิงวิชาการ เป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก จำนวน 60 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .45$) ค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = .54$) และค่าความเชื่อมั่นระดับดี ($KR_{20} = .86$)

แบบทดสอบการเขียน วัดความสามารถในการเขียนเชิงวิชาการ ให้ผู้เข้าสอบเขียนตอบตามภาระงาน จำนวน 2 ข้อ โดยข้อที่ 1 เป็นการเขียนสรุปความ (Summary writing) และข้อที่ 2 เป็นการเขียนแสดงความคิดเห็น (Argumentative writing) มีค่าความยากง่ายระดับปานกลาง ($p = .51$) ค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = .67$) และค่าสัมประสิทธิ์ความเชื่อมั่นระดับดี (Cronbach Alpha = .82)

แบบทดสอบการพูด วัดความสามารถในการพูดสื่อสารข้อมูลส่วนตัวและแสดงความคิดเห็นเชิงวิชาการ ให้ผู้เข้าสอบตอบคำถามและพูดแสดงความคิดเห็นตามภาระงาน จำนวน 3 ข้อ โดยข้อ 1 เป็นการพูดสั้น ๆ เกี่ยวกับข้อมูลของตนเอง ข้อ 2 พูดตามบัตรภาระงาน และ ข้อ 3 พูดแสดงความคิดเห็นที่เกี่ยวข้องกับข้อ 2 มีค่าความยากง่ายระดับปานกลาง ($p = .48$) ค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = .85$) และค่าสัมประสิทธิ์ความเชื่อมั่นระดับดี (Cronbach Alpha = .85)

2) แบบทดสอบวัดความสามารถภาษาอังกฤษมาตรฐาน มีโครงสร้างเหมือนกับแบบทดสอบที่ผู้วิจัยสร้างขึ้น ใช้สอบเพื่อนำคะแนนมาหาค่าความตรงร่วมสมัย (Concurrent Validity) ของแบบทดสอบ

ประกอบด้วย 2 ชนิด คือ แบบทดสอบ TOEFL ใช้วัดทักษะการฟัง การใช้ภาษา และการอ่าน และแบบทดสอบ IELTS ใช้วัดทักษะการเขียนและการพูด

3) แบบตรวจให้คะแนนแบบทดสอบอัตนัย มี 2 ชนิด คือ แบบตรวจให้คะแนนทักษะการเขียนและแบบตรวจให้คะแนนทักษะการพูด แบบตรวจให้คะแนนทักษะเขียนพัฒนาตามแนวคิดของ Brown (2004) และ Cambridge English (2015) ซึ่งประกอบด้วยเกณฑ์ให้คะแนน 4 ด้าน คือ 1) เนื้อหาสาระ (Content) 2) การจัดระบบความคิด (Organization) 3) ภาษา (Language) และ 4) กลไกทางภาษา (Mechanics) โดยแต่ละด้านมีระดับคะแนนตั้งแต่ 1-5 มีค่าความเชื่อมั่นของผู้ให้คะแนนหลายคนระดับดีมาก (Inter-rater Reliability = .99) และแบบตรวจให้คะแนนทักษะการพูด พัฒนาตามแนวคิดของ Brown (2004), Carroll (1983), and Cambridge English (2015) ซึ่งประกอบด้วยเกณฑ์ให้คะแนน 4 ด้าน คือ 1) การสื่อสารและปริมาณ (Communication and Size) 2) ความคล่องแคล่ว (Fluency) 3) ความถูกต้องของภาษา (Accuracy) และ 4) ความสามารถที่จะสื่อสารให้เข้าใจ (Comprehensibility) โดยแต่ละด้านมีระดับคะแนนตั้งแต่ 1-5 มีค่าความเชื่อมั่นระหว่างผู้ให้คะแนนหลายคนระดับดีมาก (Inter-rater Reliability = .99)

4) แบบประเมินคุณภาพเบื้องต้นของแบบทดสอบ เป็นแบบมาตราส่วนประมาณค่า (Rating Scale) แบ่งระดับความคิดเห็นเป็น 5 ระดับ จำนวน 1 ฉบับ ประเมินคุณภาพ 7 ด้าน คือ 1) ความตรงเชิงโครงสร้าง (Construct Validity) 2) ความเชื่อมั่น (Reliability) 3) ความสมจริง (Authenticity) 4) การมีปฏิสัมพันธ์ (Interactiveness) 5) ผลกระทบ (Impact) 6) ความสะดวกใช้ (Practicality) และ 7) ความเป็นธรรม (Fairness) รวมเป็นข้อคำถามจำนวน จำนวน 20 ข้อ มีค่าดัชนีความสอดคล้องเท่ากับ 1.00 ทุกข้อ

5) แบบประเมินความพึงพอใจของผู้เกี่ยวข้องต่อแบบทดสอบวัดความสามารถภาษาอังกฤษ เป็นแบบมาตราส่วนประมาณค่า (Rating Scale) แบ่งระดับความพึงพอใจเป็น 5 ระดับ จำนวน 1 ฉบับ ประเมินความพึงพอใจ 6 ด้าน คือ คำสั่ง เนื้อหา รูปแบบและการจัดลำดับความ การตรวจให้คะแนน ประโยชน์ และการจัดสอบ จำนวน 39 ข้อ มีค่าดัชนีความสอดคล้องเท่ากับ 1.00 ทุกข้อ

การเก็บรวบรวมข้อมูล ดำเนินการดังนี้

1) ทดสอบโดยใช้แบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย ใช้เวลาสอบ 3 ชั่วโมง 20 นาที และให้ผู้เข้าสอบทำแบบประเมินความพึงพอใจของผู้เกี่ยวข้องต่อแบบทดสอบวัดความสามารถภาษาอังกฤษ

2) ตรวจและให้คะแนนแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย วิเคราะห์แบบทดสอบรายข้อและหาค่าความเชื่อมั่นของแบบทดสอบ และคัดเลือกผู้เข้าสอบที่ได้คะแนนอยู่ในกลุ่มสูง ปานกลาง และต่ำ กลุ่มละ 20 % ของจำนวนผู้เข้าสอบทั้งหมดมาทดสอบการพูด โดยมีระยะเวลาห่างจากการสอบปรนัยและการเขียน 5 วัน (เลือกกลุ่มละ 20 % เนื่องจากผู้เข้าสอบมีจำนวนจำกัดและให้ได้เห็นความต่างของคะแนนในแต่ละกลุ่มที่เลือกมา)

3) ให้คะแนนแบบทดสอบพูดและวิเคราะห์แบบทดสอบรายข้อและหาค่าความเชื่อมั่นของแบบทดสอบ

4) ทดสอบโดยใช้แบบทดสอบมาตรฐาน โดยแบบทดสอบปรนัยใช้แบบทดสอบ TOEFL และแบบทดสอบอัตนัยใช้แบบทดสอบ IELTS โดยดำเนินการสอบ 2 สัปดาห์หลังจากการสอบชุดแรก

5) ตรวจสอบให้คะแนนแบบทดสอบมาตรฐาน จัดทำค่าสถิติพื้นฐาน รวมคะแนน และคัดเลือกผู้เข้าสอบที่ได้คะแนนอยู่ในกลุ่มสูง ปานกลาง และต่ำ กลุ่มละ 20 % ของจำนวนผู้เข้าสอบทั้งหมดมาทดสอบการพูดโดยมีระยะเวลาห่างจากการสอบปรนัยและการเขียน 5 วัน

6) ให้คะแนนแบบทดสอบพูดและจัดทำคะแนนสอบ

7) หาค่าความตรงร่วมสมัยของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยและแบบทดสอบมาตรฐานทั้งฉบับและแต่ละทักษะ

สถิติที่ใช้ในการวิเคราะห์ข้อมูลได้แก่

- 1) คะแนนสูงสุด คะแนนต่ำสุด คะแนนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน (สถิติพื้นฐาน)
- 2) ความยากง่าย อำนาจจำแนก (การวิเคราะห์แบบทดสอบรายข้อ)
- 3) ความเชื่อมั่นแบบทดสอบปรนัย (KR_{20}) และความเชื่อมั่นของแบบทดสอบอัตนัย (Cronbach Alpha Coefficient)
- 4) ความเชื่อมั่นของการให้คะแนนแบบทดสอบอัตนัยแบบ Inter-Rater Reliability (The Spearman-Brown)
- 5) ความตรงร่วมสมัย (Concurrent Validity) คำนวณโดยใช้สัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson's Product-Moment Correlation Coefficient)

ผลการวิจัยและการอภิปรายผล

ผลการวิจัยจะนำเสนอพร้อมการอภิปรายตามคุณสมบัติของแบบทดสอบ 4 ข้อ คือ

- 1) มีคุณภาพเบื้องต้นตามเกณฑ์คุณลักษณะของแบบทดสอบที่ดี

ในภาพรวมแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยมีคุณภาพเบื้องต้นอยู่ในระดับดี ($\bar{X} = 4.26$, S.D. = .175) และเมื่อพิจารณารายด้าน แบบทดสอบมีคุณภาพอยู่ในระดับดีมาก 1 ด้าน คือ ด้านความเป็นธรรม (Fairness) ($\bar{X} = 4.66$, S.D. = .491) และมีคุณภาพอยู่ในระดับดี 6 ด้าน คือ 1) ด้านความตรงเชิงโครงสร้าง (Construct Validity) ($\bar{X} = 4.24$, S.D. = .121) 2) ด้านผลกระทบ (Impact) ($\bar{X} = 4.24$, S.D. = .164) 3) ด้านความสมจริง (Authenticity) ($\bar{X} = 4.20$, S.D. = .084) 4) ด้านความเชื่อมั่น (Reliability) ($\bar{X} = 4.19$, S.D. = .143) 5) ด้านการมีปฏิสัมพันธ์ (Interactiveness) ($\bar{X} = 4.18$, S.D. = .109) และ 6) ด้านความสะดวกใช้ (Practicality) ($\bar{X} = 4.14$, S.D. = .113) โดยมีรายละเอียดในการอภิปราย ดังนี้

ด้านความเป็นธรรม (Fairness) ได้รับคะแนนคุณภาพสูงสุด อาจเป็นเพราะในขั้นการกำหนดเนื้อหาสาระของแบบทดสอบ ผู้วิจัยได้พิจารณา คัดเลือกหัวข้อหรือสาระของแบบทดสอบทุกทักษะให้มีความเป็นธรรมต่อผู้เข้าสอบทุกคน ปราศจากหัวข้อและข้อคำถามที่เกี่ยวข้องกับความได้เปรียบเสียเปรียบอันเนื่องมาจากเพศ เชื้อชาติ ศาสนา และฐานะทางเศรษฐกิจและสังคมของผู้เข้าสอบ และรวมถึงไม่ให้มีหัวข้อที่

นำไปสู่ความรุนแรงหรือส่งเสริมให้เกิดความแตกแยก แบ่งพรรคแบ่งฝ่ายในสังคม (Whiting, 2009; Weyant & Chisholm, 2014; Munson & Parton, 2013) ในขณะเดียวกัน ผู้วิจัยได้คัดเลือกหัวเรื่องหรือสาระแบบทดสอบให้เป็นที่ยอมรับและผู้อ่านอาจจะได้พบทั้งในการดำรงชีวิตปัจจุบันหรือในอนาคต ในแบบทดสอบการฟัง การอ่าน การเขียนและการพูด เนื้อหาสาระจะเกี่ยวข้องกับการใช้ภาษาอังกฤษในชีวิตประจำวัน และการใช้ภาษาอังกฤษเชิงวิชาการ ที่ผู้อ่านจะต้องใช้ในการทำงานหรือการศึกษาต่อในอนาคต จึงกล่าวได้ว่าแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยมีทั้งความจริง (Authenticity) และการมีปฏิสัมพันธ์ (Interactiveness) เพราะมีเนื้อหาสาระของแบบทดสอบเกี่ยวข้องกับการใช้ภาษาอังกฤษทั้งในชีวิตประจำวันและในอนาคตของผู้อ่าน (Bachman & Palmer, 1996; Brown, 2004) ซึ่งเหมือนกับแนวทางการหาคุณภาพของแบบทดสอบในการวิจัยของ Chen (2009) ที่พบว่าแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยของประเทศจีนทั้ง 2 ชุด มีความจริง (Authenticity) พอสมควร และแบบทดสอบชุดใหม่มีความจริง (Authenticity) สูงกว่าชุดเดิม เพราะมีการเพิ่มการฟังเชิงวิชาการเข้ามาในแบบทดสอบทักษะฟัง มีหัวเรื่องหลากหลายและเป็นที่ยอมรับของผู้อ่าน จึงทำให้แบบทดสอบชุดใหม่มีความจริง (Authenticity) และมีปฏิสัมพันธ์ (Interactiveness) กับผู้อ่านมากกว่าแบบทดสอบชุดเดิม ส่วนในทักษะการเขียนยังไม่มีปรับเปลี่ยนรูปแบบการทดสอบและยังไม่มีทดสอบภาระงานที่ผู้อ่านสามารถนำไปใช้ได้ในอนาคต อย่างไรก็ตาม แบบทดสอบอรรถนัยในการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดภาระงานทั้งในการเขียนและการพูดให้เป็นสิ่งที่เกี่ยวข้องกับการใช้ภาษาในอนาคตของผู้อ่าน

ด้านความตรงเชิงโครงสร้าง (Construct Validity) แบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย ทดสอบความสามารถครบทั้ง 5 ด้าน คือ การฟัง การอ่าน การพูด และการเขียน และในแต่ละด้านมีการทดสอบมากกว่า 1 ภาระงาน โดยได้แสดงเนื้อหาสาระและวิธีการทดสอบไว้ในตารางกำหนดเนื้อหาสาระแบบทดสอบ (Test Specifications) อย่างละเอียด ทำให้แบบทดสอบชุดนี้วัดความสามารถได้ครบถ้วนและมีภาระงานหลากหลาย ซึ่งเป็นการกำหนดโครงสร้างแบบทดสอบที่คล้ายกับงานวิจัยของ Chen (2009) ที่พบว่า แบบทดสอบ CET-4 ชุดใหม่ 1) มีตารางแสดงเนื้อหาสาระและวิธีการทดสอบอย่างละเอียด 2) ทดสอบทักษะต่าง ๆ ด้วยภาระงานที่หลากหลาย เช่น ในการสอบการฟัง มีการเพิ่ม Long Conversations และ Compound Dictations เข้ามา ต่างจากชุดเดิมที่มีเพียง Short Conversations และ Listening Strategies อย่างไรก็ตาม แบบทดสอบการฟังที่ผู้วิจัยสร้างขึ้นในการวิจัยครั้งนี้ มีทั้ง Listening Strategies, Short Conversations, Long Conversations และ Monologues ที่เป็น Talks หรือ Lectures นอกจากนี้ Chen (2009) พบว่า แบบทดสอบ CET-4 ชุดใหม่มีการทดสอบคำศัพท์และไวยากรณ์โดยใช้บริบทมากขึ้น และในการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดโครงสร้างแบบทดสอบให้มีการทดสอบคำศัพท์และไวยากรณ์ ซึ่งอยู่ในด้านการใช้ภาษา

ด้านผลกระทบ (Impact) แบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยมีคุณภาพด้านผลกระทบอยู่ในระดับมาก ($\bar{X} = 4.24$, S.D. = .164) เพราะแบบทดสอบชุดนี้สะท้อนแนวโน้มปัจจุบันที่กระแสด้านการทดสอบวัดความสามารถภาษาอังกฤษเป็นเรื่องที่พูดถึงกันมากและยังมีความสำคัญต่อการ

ได้งานทำของบัณฑิต แบบทดสอบชุดนี้จึงมีประโยชน์ต่อผู้เข้าสอบทั้งในด้านการทดลองวัดความสามารถภาษาอังกฤษและสร้างประสบการณ์สำหรับการทดสอบภาษาอังกฤษในอนาคต (Bachman & Palmer, 1996; Brown, 2004) ซึ่งเป็นแนวทางหาคุณภาพแบบทดสอบที่ใช้ในงานวิจัยของ Chen (2009) ที่พบว่าแบบทดสอบชุดใหม่มีผลกระทบต่อรูปแบบและเนื้อหาในการเรียนรู้ของผู้เข้าสอบ และกลายเป็นแรงจูงใจ ทำให้นักศึกษาสนใจการเรียนรู้ทั้งในเชิงปริมาณและคุณภาพ แต่การสอบทำให้นักศึกษาอยู่ในภาวะกดดันมากขึ้นด้วย สำหรับผลกระทบต่ออาจารย์ผู้สอน พบว่ามีการปรับเปลี่ยนในเนื้อหาและวิธีการสอน อาจารย์มีแนวโน้มที่จะสอนให้สอดคล้องกับการสอบมากขึ้น เนื่องจากแบบทดสอบชุดใหม่มีการเพิ่มสัดส่วนในการวัดทักษะฟังและพูดมากกว่าเดิม และอาจารย์ได้ปรับการสอนให้ไปทางการสื่อสารมากขึ้นด้วย สำหรับผลกระทบต่อระบบการศึกษาและสังคม แบบทดสอบ CET-4 ได้ส่งผลให้ผู้เข้าสอบคิดถึงความสำเร็จ ความรู้และความสามารถทั้งของรายบุคคลและสังคมมากยิ่งขึ้น สำหรับในบริบทการสอบวัดความสามารถภาษาอังกฤษในประเทศไทย อาจจะยังไม่มีผลกระทบต่อฝ่ายต่าง ๆ อย่างเห็นได้ชัดเหมือนในประเทศจีน แต่มีแนวโน้มที่จะมีผลกระทบชัดเจนขึ้นในอนาคตอันใกล้นี้ ทั้งในระดับนโยบายของผู้บริหาร การจัดการเรียนการสอนโดยผู้สอน และผู้เรียน เนื่องจากเริ่มมีการกำหนดให้ผู้เรียนแสดงผลการสอบวัดความสามารถภาษาอังกฤษก่อนการจบการศึกษาและยื่นประกอบการสมัครงานมากขึ้น (Jobsdb.com, 2015; Wallstreet English, 2015; British Council, 2015)

ด้านความสะดวกใช้ (Practicality) แบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย มีคุณภาพด้านความสะดวกใช้อยู่ในระดับดี ($\bar{X} = 4.14$, S.D. = .113) ทั้งในด้านการสร้างแบบทดสอบ และนำแบบทดสอบไปใช้ ซึ่งเป็นแนวทางหาคุณภาพแบบทดสอบที่ใช้ในงานวิจัยของ Chen (2009) ที่ระบุว่าแบบทดสอบ CET-4 ของประเทศจีนไม่พบปัญหาทั้งในการพัฒนา การบริหาร และการตรวจให้คะแนน เนื่องจากมีการใช้งานมาเกือบ 20 ปี สำหรับปัญหาการใช้งานแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยที่ผู้วิจัยสร้างขึ้น ยังไม่ปรากฏปัญหาที่ชัดเจน เนื่องจากเป็นการเริ่มต้นใช้กับกลุ่มตัวอย่างที่มีจำนวนจำกัด การนำไปใช้จริงกับกลุ่มตัวอย่างขนาดใหญ่ในอนาคตอาจจะเจอปัญหาด้านการใช้งานที่จะต้องนำมาพัฒนาและปรับปรุงระบบการทดสอบให้ดีขึ้นต่อไป

2) มีคุณภาพรายข้อและความเชื่อมั่นของแบบทดสอบเป็นไปตามเกณฑ์

คุณภาพรายข้อและความเชื่อมั่นของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยทั้ง 2 ชนิด คือ แบบทดสอบปรนัย และแบบทดสอบอัตนัยโดยมีรายละเอียด ดังนี้

แบบทดสอบปรนัยซึ่งประกอบด้วยการฟัง การใช้ภาษา และการอ่าน มีข้อคำถามทั้งหมด 150 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .55$) ค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = .49$) และค่าความเชื่อมั่นระดับดี ($KR_{20} = .83$) มีข้อคำถามคุณภาพดี 122 ข้อ คิดเป็น 81.33 % และข้อคำถามที่ควรปรับปรุง 28 ข้อ คิดเป็น 18.67 % และสามารถแยกคุณภาพรายทักษะได้ คือ แบบทดสอบการฟัง จำนวน 50 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .50$) ค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = .49$) และค่าความเชื่อมั่นระดับดี ($KR_{20} = .83$) มีข้อคำถามคุณภาพดี 38 ข้อ คิดเป็น 76 % และข้อคำถามที่ควรปรับปรุง 12 ข้อ คิด

เป็น 24 % แบบทดสอบการใช้ภาษา จำนวน 40 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .54$) ค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = .52$) และค่าความเชื่อมั่นระดับดี ($KR_{20} = .82$) มีข้อคำถามคุณภาพดี 36 ข้อ คิดเป็น 90 % และข้อคำถามที่ควรปรับปรุง 4 ข้อ คิดเป็น 10 % และแบบทดสอบการอ่าน จำนวน 60 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .60$) ค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = .46$) และค่าความเชื่อมั่นระดับดี ($KR_{20} = .83$) มีข้อคำถามคุณภาพดี 48 ข้อ คิดเป็น 80 % และข้อคำถามที่ควรปรับปรุง 12 ข้อ คิดเป็น 20 %

แบบทดสอบอัตนัยซึ่งประกอบด้วย การเขียนและการพูด มีข้อคำถามทั้งหมด 5 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .55$) ค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = .38$) และค่าความเชื่อมั่นระดับดี (Coefficient Alpha = .80) และเป็นข้อคำถามคุณภาพดีทั้ง 5 ข้อ คิดเป็น 100 % และสามารถแยกคุณภาพรายทักษะได้ คือ แบบทดสอบการเขียน จำนวน 2 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .60$) ค่าอำนาจจำแนกระดับจำแนกได้ปานกลาง ($r = .26$) และค่าความเชื่อมั่นระดับยังน่าสงสัย (Coefficient Alpha = .61) และเป็นข้อคำถามคุณภาพดีทั้ง 2 ข้อ คิดเป็น 100 % และแบบทดสอบการพูด จำนวน 3 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .50$) ค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = .49$) และค่าความเชื่อมั่นระดับดีมาก (Coefficient Alpha = .99) และเป็นข้อคำถามคุณภาพดีทั้ง 3 ข้อ คิดเป็น 100 %

ด้านคุณภาพรายข้อ

จากผลการวิจัยในด้านคุณภาพรายข้อของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย พบว่า แบบทดสอบปรนัยซึ่งประกอบด้วย การฟัง การใช้ภาษา และการอ่าน มีข้อคำถามทั้งหมด 150 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .55$) ซึ่งใช้วิธีการหาคุณภาพรายข้อของแบบทดสอบเหมือนกับในงานวิจัยของศิริพร พงษ์สุรพิพัฒน์และคณะ (2544) ที่ได้วิเคราะห์คุณภาพของแบบทดสอบ CU-TEP 3 ฉบับ และพบว่าแบบทดสอบปรนัยมีค่าความยากง่ายระดับปานกลาง ($p = 0.48 - 0.53$) สำหรับแบบทดสอบอัตนัยซึ่งประกอบด้วย การเขียนและการพูด และมีข้อคำถามทั้งหมด 5 ข้อ มีค่าความยากง่ายระดับปานกลาง ($p = .55$) ส่วนแบบทดสอบอัตนัยของศิริพร พงษ์สุรพิพัฒน์และคณะ (2544) พบว่าแบบทดสอบอัตนัยข้อ 1 และ 3 ของทุกฉบับยากเกินไป ($FV = 0.14 - 0.15$) แต่ข้อ 2 ของทุกฉบับมีค่าความยากง่ายพอใช้ได้

ในการพิจารณาค่าอำนาจจำแนกของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย พบว่า แบบทดสอบปรนัยมีค่าอำนาจจำแนกระดับจำแนกได้ดี ($r = 0.49$) ซึ่งใช้วิธีการหาค่าอำนาจจำแนกเหมือนกับในงานวิจัยของศิริพร พงษ์สุรพิพัฒน์และคณะ (2544) ที่พบว่าแบบทดสอบปรนัย มีค่าอำนาจจำแนกดีพอใช้ ($R_{bis} 0.34 - 0.38$) ค่าอำนาจจำแนกของแบบทดสอบอัตนัยในงานวิจัยครั้งนี้อยู่ในระดับจำแนกได้ดี ($r = .38$) สำหรับในงานวิจัยของศิริพร พงษ์สุรพิพัฒน์และคณะ (2544) พบว่าแบบทดสอบอัตนัยข้อ 2 และ 3 ของฉบับที่ 1 ข้อ 1 และ 3 ของฉบับที่ 2 และ ข้อ 1, 2, 3 ของฉบับที่ 3 มีค่าอยู่ในเกณฑ์ดี ส่วนข้ออื่น ๆ มีค่าอยู่ในเกณฑ์ที่ยอมรับได้

แบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยมีความง่ายอยู่ในระดับปานกลาง และค่าอำนาจจำแนกอยู่ในเกณฑ์ดี อาจมีสาเหตุมาจาก 3 ประเด็นดังนี้

1) การคัดเลือกเนื้อหาสาระของแบบทดสอบ เนื้อหาสาระแบบทดสอบของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยทุกทักษะ ผ่านการคัดเลือกให้เหมาะสมโดยผู้วิจัยและตรวจสอบอีกครั้งโดยผู้เชี่ยวชาญ เพื่อให้แบบทดสอบมีเนื้อหาสาระเหมาะสมกับผู้เข้าสอบทุกคน ไม่ง่ายหรือไม่ยากจนเกินไป และสอดคล้องกับระดับความสามารถของผู้เข้าสอบ (Genesee & Upshur, 1996)

2) การเพิ่มหรือลดข้อคำถาม หลังจากการทดลองใช้แบบทดสอบครั้งที่ 1 และ 2 ผู้วิจัยได้ตัดข้อคำถามที่มีคุณภาพรายข้อไม่ตรงตามเกณฑ์ออก เช่น ยาก หรือง่ายเกินไป ไม่มีอำนาจจำแนก หรือจำแนกได้ไม่ดี หรืออำนาจจำแนกเป็นลบ ทำให้เหลือข้อคำถามที่มีคุณภาพดีสำหรับนำมาใช้ในการเก็บข้อมูลจริงกับกลุ่มตัวอย่าง (Genesee & Upshur, 1996)

3) การวิเคราะห์ความลำเอียง หลังจากการทดลองใช้แบบทดสอบครั้งที่ 2 ผู้วิจัยได้วิเคราะห์ความลำเอียงของข้อคำถามของแบบทดสอบปรนัย ในกลุ่มของผู้เข้าสอบสายวิทยาศาสตร์และศิลปศาสตร์ และได้ตัดข้อคำถามที่มีค่าความลำเอียงสูงเกินไปออกจากชุดคำถาม คงไว้เฉพาะข้อคำถามที่มีคุณภาพดีและมีค่าความลำเอียงต่ำมาใช้ในการเก็บข้อมูลจริงกับกลุ่มตัวอย่าง (วรภรณ์ ประจงเศรษฐ์, 2561)

ด้านความเชื่อมั่น

ความเชื่อมั่นของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย แบ่งออกได้เป็น 2 ชนิด คือ ความเชื่อมั่นของแบบทดสอบปรนัย และอัตนัย ดังรายละเอียดต่อไปนี้

จากผลการวิจัยพบว่า แบบทดสอบปรนัยซึ่งประกอบด้วยการฟัง การใช้ภาษา และการอ่าน มีข้อคำถามทั้งหมด 150 ข้อ มีค่าความเชื่อมั่นระดับดี ($KR_{20} = .83$) มีข้อคำถามคุณภาพดี 122 ข้อ คิดเป็น 81.33 % และข้อคำถามที่ควรปรับปรุง 28 ข้อ คิดเป็น 18.67 % โดยในการวิจัยครั้งนี้ได้ใช้วิธีการหาค่าความเชื่อมั่นของแบบทดสอบปรนัยคล้ายกับในงานของอัจฉรา วงศ์โสธร และศิริพร พงษ์สุรพิพัฒน์ (2544) ซึ่งทำการวิจัยด้านการสร้างและพัฒนาแบบทดสอบภาษาอังกฤษเพื่อพัฒนาแบบทดสอบสมรรถภาพทั่วไปทางภาษาอังกฤษด้านการฟัง การอ่านและการเขียน ได้แบบทดสอบปรนัยวัดทักษะการฟัง อ่าน และเขียน จำนวน 4 ฉบับที่มีค่าความเชื่อมั่นอยู่ในเกณฑ์ดี – ดีมาก ($KR_{20} = .854 - .910$) และได้ข้อสอบแบบปรนัยวัดสมรรถภาพทั่วไปทางภาษาอังกฤษที่มีคุณภาพดีมาก – พอใช้ จำนวน 471 ข้อ และพบว่า แบบทดสอบฉบับที่ 3 มีค่าความเชื่อมั่นสูงที่สุด ($KR_{20} = 0.92$) และฉบับที่ 2 มีค่าต่ำที่สุด ($KR_{20} = 0.89$) และได้ใช้วิธีการหาค่าความเชื่อมั่นของแบบทดสอบคล้ายกับในงานของ Ito (2005) ที่ต้องการหาค่าความเชื่อมั่นของข้อคำถามในแบบทดสอบ JFSAT (The Joint First Achievement Test) ได้ค่าสัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบเท่ากับ .853

แบบทดสอบอัตนัยซึ่งประกอบด้วยกรเขียนและการพูด มีข้อคำถามทั้งหมด 5 ข้อ มีค่าความเชื่อมั่นระดับดี (Coefficient Alpha = .80) และเป็นข้อคำถามคุณภาพดีทั้ง 5 ข้อ คิดเป็น 100% โดยในการวิจัยครั้งนี้ได้ใช้วิธีการหาค่าความเชื่อมั่นของแบบทดสอบอัตนัยคล้ายกับงานวิจัยของอัจฉรา วงศ์โสธร และศิริพร

พงษ์สุรพัฒน์ (2544) ซึ่งได้แบบทดสอบอัตนัยวัดทักษะการเขียนที่มีค่าความเชื่อมั่นอยู่ในเกณฑ์ดี – ดีมาก ($R_{xy} = .594 - .840$) และได้ข้อทดสอบแบบอัตนัยวัดสมิทธิภาพทางการเขียนที่มีคุณภาพดี – พอใช้ จำนวน 12 ข้อ และได้ค่าความเชื่อมั่นของแบบทดสอบอัตนัย ฉบับที่ 1 มีค่าความเชื่อมั่นสูงสุด (0.940) และฉบับที่ 3 มีค่าความเชื่อมั่นต่ำที่สุด (0.726)

เมื่อพิจารณาความเชื่อมั่นรายทักษะของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย ทั้ง 5 ทักษะ สามารถแบ่งค่าความเชื่อมั่นได้ 3 ระดับ คือ 1) ทักษะที่มีค่าความเชื่อมั่นระดับดีมาก คือ ทักษะการพูด (Coefficient Alpha = .99) 2) ทักษะที่มีค่าความเชื่อมั่นระดับดี คือ ทักษะการฟัง ทักษะการอ่าน ($KR_{20} = .83$) และการใช้ภาษา ($KR_{20} = .82$) และ 3) ทักษะที่มีค่าความเชื่อมั่นระดับยังน่าสงสัย คือ ทักษะการเขียน (Coefficient Alpha = .61) โดยการวิจัยครั้งนี้ได้ใช้วิธีการหาค่าความเชื่อมั่นรายทักษะคล้ายกับในงานวิจัยของ LCRT Consulting (2013) ที่วิเคราะห์ความเชื่อมั่นของแบบทดสอบ MELA (The Michener English Language Assessment) 4 ทักษะ คือ การฟัง การพูด การอ่าน และการเขียน โดยแบบทดสอบ MELA มีค่าความเชื่อมั่นด้านการฟัง เท่ากับ .89 การพูดเท่ากับ .94 การอ่านเท่ากับ .85 และการเขียนเท่ากับ .88

แบบทดสอบด้านการเขียนของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยมีค่าความเชื่อมั่นต่ำ อาจมีผลมาจาก 3 ปัจจัย คือ 1) ความสามารถด้านการเขียนของผู้เข้าสอบ จากผลการวิจัยที่เป็นคะแนนด้านการเขียน พบว่า ผู้เข้าสอบที่ทำคะแนนทักษะเขียนได้ส่วนใหญ่เป็นผู้ที่มีคะแนนรวมอยู่ในกลุ่มสูง ซึ่งถือว่ามีความสามารถด้านทักษะเขียนอยู่บ้างแล้ว และมักจะเป็นนิสิตวิชาเอกภาษาอังกฤษและนิสิตที่มีวิชาเอกด้านวิทยาศาสตร์ ในทางกลับกัน ผู้เข้าสอบสายศิลป์บางกลุ่ม เช่น กลุ่มวิชาเอกด้านธุรกิจหรือภาษาอื่น ๆ ที่ไม่ใช้ภาษาอังกฤษเลือกที่จะไม่ตอบแบบทดสอบการเขียน อาจเป็นเพราะแบบทดสอบยากเกินความสามารถของผู้เข้าสอบส่วนใหญ่ (Bachman & Palmer, 1996) 2) ความไม่คุ้นเคยกับภาระงานสอบ ในการจัดการเรียนการสอนวิชาภาษาอังกฤษพื้นฐาน นิสิตทั่วไปอาจไม่มีประสบการณ์ในการฝึกเขียนบรรยายหรืออธิบายที่มีความยาวเกินระดับประโยค แต่นิสิตวิชาเอกภาษาอังกฤษได้ฝึกเขียนในรายวิชาการเขียนเบื้องต้น การเขียนย่อหน้า และการเขียนเรียงความ จึงกล่าวได้ว่า ผู้เข้าสอบส่วนใหญ่ขาดประสบการณ์หรือไม่คุ้นเคยกับภาระงานสอบในแบบทดสอบเขียน (Bachman & Palmer, 1996) และ 3) จำนวนผู้เข้าสอบทำแบบทดสอบเขียน จากสาเหตุในข้อ 1 และ 2 ส่งผลให้จำนวนผู้เข้าสอบที่ตอบแบบทดสอบได้มีจำนวนจำกัด และทำให้ความเชื่อมั่นของแบบทดสอบด้านการเขียนอยู่ในระดับยังน่าสงสัย

3) มีความตรงร่วมสมัย (Concurrent Validity) สูง เมื่อเทียบคะแนนกับแบบทดสอบมาตรฐาน

ความตรงร่วมสมัยของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย คือ ความสัมพันธ์ระหว่างคะแนนของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยและแบบทดสอบมาตรฐาน ซึ่งแบ่งได้ 2 ชนิด คือ ความสัมพันธ์โดยรวมของคะแนนแต่ละทักษะ และความสัมพันธ์แต่ละตอนย่อยหรือภาระงานย่อยในแต่ละทักษะ โดยการวิจัยครั้งนี้ได้กำหนดความสัมพันธ์ระหว่างคะแนนของแบบทดสอบวัดความสามารถภาษาอังกฤษและแบบทดสอบมาตรฐานว่า ควรมีค่าสัมประสิทธิ์สหสัมพันธ์ (Pearson's

Product-Moment Correlation Coefficient) 0.75 ขึ้นไป และสามารถสรุปความสัมพันธ์ระหว่างคะแนนในแต่ละทักษะได้ดังนี้

ทักษะการเขียน การพูดและการฟังมีค่าความสัมพันธ์โดยรวมของคะแนนอยู่ในระดับสูงมากโดยค่าความสัมพันธ์ของคะแนนทักษะการเขียนและการพูดอยู่ในระดับสูงมาก ($r_{xy} = .95$) ทักษะการฟังอยู่ในระดับสูงมาก ($r_{xy} = .93$) โดยตอนย่อยของทักษะการเขียนมีค่าความสัมพันธ์ของคะแนนอยู่ในระดับสูง-สูงมาก โดย Task A (การเขียนสรุปความ) อยู่ในระดับสูง ($r_{xy} = .88$) และ Task B (การเขียนแสดงความคิดเห็น) อยู่ในระดับสูงมาก ($r_{xy} = .90$) และทุกตอนย่อยของทักษะการพูดมีค่าความสัมพันธ์ของคะแนนอยู่ในระดับสูงมาก โดย Task A (การพูดสั้น ๆ เกี่ยวกับข้อมูลของตนเอง) และ Task B (การพูดตามบัตรภาระงาน) อยู่ในระดับสูงมาก ($r_{xy} = .93$) และ Task C (การพูดแสดงความคิดเห็นที่เกี่ยวข้องกับ Task B) อยู่ในระดับสูงมาก ($r_{xy} = .94$) และตอนย่อยของทักษะการฟังมีค่าความสัมพันธ์ของคะแนนอยู่ในระดับสูงมากและต่ำ โดยภาระงาน A มีค่าความสัมพันธ์อยู่ในระดับสูงมาก ($r_{xy} = .91$) ภาระงาน B อยู่ในระดับต่ำ ($r_{xy} = .37$) และภาระงาน C อยู่ในระดับปานกลาง ($r_{xy} = .65$)

ทักษะการอ่านมีค่าความสัมพันธ์โดยรวมของคะแนนอยู่ในระดับสูง ($r_{xy} = .87$) ส่วนทักษะการใช้ภาษามีค่าความสัมพันธ์โดยรวมของคะแนนอยู่ในระดับค่อนข้างสูง ($r_{xy} = .76$) และตอนย่อยของทักษะการใช้ภาษาทั้งตอน A ($r_{xy} = .59$) และ B ($r_{xy} = .71$) มีความสัมพันธ์ของคะแนนต่ำกว่าเกณฑ์ ($r_{xy} \geq .75$)

ความตรงร่วมสมัย (Concurrent Validity) ของแบบทดสอบทั้งฉบับ

ผลการวิจัยพบว่า ความตรงร่วมสมัย (Concurrent Validity) ของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยและแบบทดสอบมาตรฐานทั้งฉบับ มีค่า $r_{xy} = .90$ ซึ่งสูงกว่าเกณฑ์ที่กำหนดไว้ ($r_{xy} \geq 0.75$) โดยการวิจัยครั้งนี้ใช้วิธีการหาค่าความตรงร่วมสมัยของแบบทดสอบทั้งฉบับคล้ายกับในงานวิจัย 4 งาน คือ 1) งานวิจัยของ Wullur (2011) ที่หาความตรงร่วมสมัย (Concurrent Validation) ของคะแนนของแบบทดสอบจัดเข้าชั้นเรียน (Placement Test) ของ Universitas Klabat ที่พัฒนาขึ้นใหม่กับแบบทดสอบจัดเข้าชั้นเรียนที่มีอยู่แล้ว และพบว่าแบบทดสอบทั้ง 2 ชนิด มีความสัมพันธ์กันอย่างสูง โดยมีค่ามีความตรงร่วมสมัย $r_{xy} = .964$ 2) งานวิจัยของ Cheng (2013) ซึ่งหาความสัมพันธ์ระหว่างคะแนนของแบบทดสอบ AST Exam English และ The GMRT พบว่าแบบทดสอบทั้ง 2 ชนิด มีความสัมพันธ์กันในเชิงบวก $r_{xy} = .801$ 3) งานวิจัยของ Riazzi (2014) ซึ่งหาความสัมพันธ์ระหว่างคะแนนของ PTE Academic และ IELTS Academic ของผู้เข้าสอบ พบความสัมพันธ์สูงสุดมาจากคะแนนรวมของแบบทดสอบทั้ง 2 ชนิด ($r_{xy} = .824$) และ 4) งานวิจัยของ Zhou (2004) ซึ่งหาความสัมพันธ์ระหว่างคะแนนของแบบทดสอบภาษาอังกฤษระดับมหาวิทยาลัยในประเทศจีน 2 ฉบับ คือ แบบทดสอบ CET-6 (The National College English Tests-Band 6) และแบบทดสอบ TEM-4 (The Test for English Majors-4) พบว่าคะแนนของแบบทดสอบทั้ง 2 ฉบับ มีความสัมพันธ์กันอย่างมีนัยสำคัญ ค่าความสัมพันธ์ $r_{xy} = .712$

ความตรงร่วมสมัย (Concurrent Validity) ของแบบทดสอบแต่ละทักษะ

ผลการวิจัยพบว่า ค่าความตรงร่วมสมัย (Concurrent Validity) ของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยและแบบทดสอบมาตรฐานแต่ละทักษะ สามารถจัดอันดับได้ คือ การเขียนและการพูด ($r_{xy} = .95$) การฟัง ($r_{xy} = .93$) การอ่าน ($r_{xy} = .87$) และการใช้ภาษา ($r_{xy} = .76$) ซึ่งสูงกว่าเกณฑ์ที่กำหนดไว้ ($r_{xy} \geq 0.75$) ทุกทักษะ โดยการวิจัยครั้งนี้ใช้วิธีการหาค่าความตรงร่วมสมัยของแบบทดสอบแต่ละทักษะคล้ายกับในงานวิจัย 2 งาน คือ 1) งานวิจัยของ Weir, Chan, and Nakatsuhara (2013) ที่ศึกษาความตรงร่วมสมัย (Concurrent Validity) ของแบบทดสอบการอ่านและการเขียนของ the Advanced Level GEPT พบความสัมพันธ์ระดับปานกลางถึงมากระหว่างคะแนนรวมในทักษะการเขียนของ GEPT และคะแนนการสอบเขียนของ IELTS โดยมีค่าความสัมพันธ์ $r_{xy} = .432$ ($p < .01$) และระหว่างคะแนนการเขียนของ GEPT กับคะแนนการอ่านและการเขียนของแบบทดสอบ IELTS โดยมีค่าความสัมพันธ์ $r_{xy} = .503$ ($p < .01$) และ 2) งานวิจัยของศิริพร พงษ์สุรพิพัฒน์ และคณะ (2544) ที่วิเคราะห์ความสัมพันธ์ระหว่างคะแนนจากแบบทดสอบ CU-TEP และแบบทดสอบ TOEFL พบว่า คะแนนจากแบบทดสอบ CU-TEP ฉบับที่ 3 มีความสัมพันธ์กับคะแนนจากแบบทดสอบ TOEFL สูงที่สุดในแต่ละทักษะ ยกเว้นด้านการเขียนแบบปรนัย ซึ่งมีค่าน้อยกว่าฉบับที่ 1 เล็กน้อยแต่ก็ยังมีค่าความสัมพันธ์อยู่ในระดับสูง ($r_{xy} = 0.84$) ในขณะที่คะแนนของแบบทดสอบฉบับที่ 2 มีค่าความสัมพันธ์น้อยกว่าฉบับอื่น ๆ ทุกทักษะ

เหตุผลที่แบบทดสอบด้านการใช้ภาษาของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยมีค่าความตรงร่วมสมัย (Concurrent Validity) ต่ำกว่าทักษะอื่น ๆ น่าจะมีสาเหตุมาจากหัวเรื่อง (Topical Theme) และความซับซ้อนของประโยค (Text Complexity) ไม่เท่ากัน กล่าวคือ แบบทดสอบด้านการใช้ภาษาของ TOEFL มีหัวเรื่องและคำศัพท์ที่ยากเกินระดับความสามารถของผู้เข้าสอบ แม้ว่าจะวัดทักษะเดียวกัน ทำให้ดูเหมือนว่าแบบทดสอบด้านการใช้ภาษาของ TOEFL ยากกว่าแบบทดสอบด้านการใช้ภาษาของแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย ซึ่งหัวเรื่องเป็นที่คุ้นเคยและใช้คำศัพท์ที่เหมาะสมกับระดับความสามารถของผู้เข้าสอบมากกว่า

4) ผู้เกี่ยวข้องมีความพึงพอใจต่อแบบทดสอบที่สร้างขึ้นในระดับมาก

ในภาพรวมผู้มีส่วนเกี่ยวข้องพึงพอใจแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยอยู่ในระดับมาก ($\bar{X} = 4.27$, S.D. = 0.770) และมีความพึงพอใจด้านต่าง ๆ เรียงตามลำดับ คือ 1) ด้านประโยชน์ของแบบทดสอบ ($\bar{X} = 4.40$, S.D. = 0.687) 2) ด้านคำสั่งของแบบทดสอบ ($\bar{X} = 4.36$, S.D. = 0.708) 3) ด้านการจัดการสอบ ($\bar{X} = 4.32$, S.D. = 0.842) 4) ด้านการตรวจให้คะแนน ($\bar{X} = 4.25$, S.D. = 0.784) 5) ด้านรูปแบบและการจัดลำดับความ ($\bar{X} = 4.22$, S.D. = 0.813) และ 6) ด้านเนื้อหาสาระของแบบทดสอบ ($\bar{X} = 4.19$, S.D. = 0.786)

ข้อเสนอแนะ

การนำแบบทดสอบชุดนี้ไปใช้ในอนาคต ควรดำเนินการดังนี้

1. ก่อนนำแบบทดสอบฉบับนี้ไปใช้จริง ผู้วิจัยควรปรับปรุงจำนวนข้อคำถามให้ได้ครบตามจำนวนที่ต้องการ เพราะในขณะนี้ข้อคำถามที่ได้คุณภาพแล้วบางส่วน คือ ด้านการฟังได้ข้อคำถามคุณภาพดี 30 ข้อ (76%) ต้องสร้างเพิ่ม 12 ข้อ (24%) ด้านการใช้ภาษา 36 ข้อ (90%) ต้องสร้างเพิ่ม 4 ข้อ (10%) และด้านการอ่าน 48 ข้อ (80%) ต้องสร้างเพิ่ม 12 ข้อ (20%) และทดลองใช้แบบทดสอบชุดนี้อีกครั้งกับกลุ่มตัวอย่างที่มีจำนวนมากกว่าเดิม โดยให้ฝ่ายวิชาการ มหาวิทยาลัยทักษิณ เป็นผู้ประสานงานและดำเนินการในการจัดสอบ
2. หลังจากนำแบบทดสอบที่มีอยู่ไปใช้ และหาคุณภาพได้คงที่ตามเกณฑ์แล้ว ควรพัฒนาแบบทดสอบคู่ขนานชุดอื่น ๆ เพื่อให้มีแบบทดสอบไว้ใช้มากกว่า 1 ฉบับ และควรจัดเก็บข้อคำถามแยกตามทักษะย่อย ตามระดับคุณภาพ ไว้ในคลังข้อสอบ
3. ควรพัฒนาแบบทดสอบออนไลน์ในทักษะต่าง ๆ เพื่อให้บัณฑิตได้ฝึกวัดความสามารถภาษาอังกฤษด้วยตนเอง จนมีความคุ้นเคยกับการะงานสอบ ก่อนที่จะไปทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยหรือแบบทดสอบมาตรฐานสากล

การวิจัยในอนาคต

จากผลการวิจัยและการอภิปรายผล การวิจัยในอนาคตควรเกี่ยวข้องกับประเด็นต่อไปนี้

- 1) จากการอภิปรายผลการวิจัย พบว่า ผู้เข้าสอบส่วนใหญ่ไม่ทำแบบทดสอบด้านการเขียน อาจจะเป็นเพราะแบบทดสอบยากเกินความสามารถหรือไม่คุ้นเคยกับการะงานเขียน และเนื่องจากภาระงานเขียนในแบบทดสอบเป็นภาระงานที่ผู้เข้าสอบสามารถพบได้ในการสอบมาตรฐานในอนาคต และเป็นการวัดความสามารถการเขียนที่ใช้กับแบบทดสอบมาตรฐาน จึงควรดำเนินงานวิจัยที่เกี่ยวข้องกับการพัฒนาทักษะการเขียนของนิสิตปริญญาตรีสาขาวิชาอื่นๆ ที่ไม่ใช่เอกภาษาอังกฤษ เพื่อช่วยพัฒนาความสามารถด้านการเขียนภาษาอังกฤษของนิสิตเหล่านั้น
- 2) จากผลการวิจัย พบว่า ทักษะการฟังเป็นภาระงานสอบที่ผู้เข้าสอบยังไม่คุ้นเคย หากจะเตรียมนิสิตให้สามารถมีพัฒนาการด้านการฟังที่ดี ควรศึกษาและทำการวิจัยเพื่อพัฒนาทักษะการฟังของนิสิตปริญญาตรี มหาวิทยาลัยทักษิณ
- 3) เนื่องจากคะแนนเฉลี่ยของการทำแบบทดสอบได้ทั้ง 2 ฉบับ อยู่ในระดับปานกลาง แสดงถึงระดับความสามารถด้านภาษาอังกฤษของผู้เข้าสอบโดยส่วนใหญ่ที่ยังไม่ถึงเกณฑ์มาตรฐาน ควรมีการวิจัยในอนาคตเพื่อหาความสอดคล้องระหว่างเนื้อหาสาระที่สอนในวิชาภาษาอังกฤษพื้นฐานกับความสามารถในการทำแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัยหรือแบบทดสอบวัดความสามารถภาษาอังกฤษมาตรฐานสากล

4) ควรดำเนินการวิจัยต่อ ด้วยการเพิ่มขยายกลุ่มตัวอย่างไปยังสถาบันอุดมศึกษาอื่น ๆ เพื่อพัฒนาแบบทดสอบวัดความสามารถภาษาอังกฤษระดับมหาวิทยาลัย (University English Proficiency Test) ให้เป็นแบบทดสอบมาตรฐาน (Standardized Test) เพื่อการใช้งานในอนาคต

เอกสารอ้างอิง

- มหาวิทยาลัยทักษิณ. (2559). *แผนยุทธศาสตร์การพัฒนามหาวิทยาลัยทักษิณ พ.ศ. 2558 – 2567*. สืบค้น 28 พฤษภาคม 2562, จาก http://planning.tsu.ac.th/main/files_sec3/
- วารกรณ์ ประจงเศรษฐ์. (2561). *การวิเคราะห์ความลำเอียงของข้อสอบ*. สืบค้น 15 มกราคม 2562, จาก <https://btu6053200011mci.blogspot.com/2018/06/8.html>
- ศิริพร พงษ์สุรพิพัฒน์, กรองแก้ว กรรณสูตร, และสุภาณี ชินวงศ์. (2544). ความตรงร่วมสมัย: การเปรียบเทียบคะแนนสอบแบบทดสอบสมิทธิภาพทั่วไปทางภาษาอังกฤษของจุฬาลงกรณ์มหาวิทยาลัย (CU-TEP) กับคะแนนแบบทดสอบ TOEFL. *วารสารภาษาปริทัศน์*, 2544(19), 32-48.
- สำนักงานคณะกรรมการการศึกษา. (2559). *ประกาศคณะกรรมการการอุดมศึกษา เรื่อง นโยบายการยกระดับมาตรฐานภาษาอังกฤษในสถาบันอุดมศึกษา*. สืบค้น 28 พฤษภาคม 2562, จาก http://www.mua.go.th/users/bhes/front_home/Data%20Bhes_2559/04052559.pdf
- สำนักงานคณะกรรมการการศึกษา. (2561). *ประกาศคณะกรรมการการอุดมศึกษากระทรวงศึกษาธิการ เรื่อง การคัดเลือกเข้าร่วมโครงการผลิตครูเพื่อพัฒนาท้องถิ่น ปี พ.ศ. 2561*. สืบค้น 28 พฤษภาคม 2562, จาก https://file.job.thai.com/prakad/niets201711/niets201711_5.pdf
- อัจฉรา วงศ์โสธร ,และศิริพร พงษ์สุรพิพัฒน์. (2544). *การวิจัยเพื่อพัฒนาข้อทดสอบสมิทธิภาพทั่วไปทางภาษาอังกฤษ*. สืบค้น 6 พฤศจิกายน 2557, จาก <http://cuir.car.chula.ac.th/handle/123456789/757>
- Arkoudis, Sophie., et.al. (2009). *The Impact of English Language Proficiency and Workplace Readiness on the Employment Outcomes of Tertiary International Students*. Retrieved September 12, 2014, from https://internationaleducation.gov.au/research/Publications/Documents/ELP_Full_Report.pdf
- Arunee Wiriyaichitra. (n.d.). *English Language Teaching and Learning in Thailand in this Decade*. Retrieved September 12, 2014, from <http://apecknowledgebank.org/resources/downloads/English>
- Bachman, Lyle F. & Palmer, Adrian S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

- Brown, H. Douglas. (2004). *Language Assessment: Principles and Classroom Practices*. New York: Pearson Education, Inc.
- Cambridge English. (2015). *Cambridge English First: Information for Candidates*. Retrieved December 10, 2018, from <https://www.cambridgeenglish.org/br/Images/181310-first-information-for-candidates-2015-document.pdf>
- Carroll, Brendan J. (1983). *Testing Communicative Performance*. Great Britain: A Wheaton & Co. Ltd., Exeter.
- Chen, Lan. (2009). *Assessing the Test Usefulness: A Comparison Between the Old and the New College English Test Band 4 (CET-4) in China*. Retrieved April 17, 2019, from <http://www.diva-portal.org/smash/get/diva2:228454/FULLTEXT01.pdf>
- Cheng, Tzung-yu. (2013). *Correlating the Gates-MacGinitie Reading Tests with the University Entrance English Examinations Held in Taiwan*. Retrieved November 11, 2015, from <http://files.eric.ed.gov/fulltext/ED544082.pdf>
- Cohen, Andrew D. (1994). *Assessing Language Ability in the Classroom*. Boston, Mass.: Heinle & Heinle Publishers.
- Gao, Fen. (2010). What's Wrong with Current Chinese College English Assessment System? Reform or Not?. *International Education Studies*, 3(1), 34-37.
- Genesee, Fred., & Upshur, John A. (1996). *Classroom-Based Evaluation in Second Language Education*. Cambridge: Cambridge University Press.
- Heaton, J.B. (1975). *Writing English Language Tests*. London: Longman.
- Ito, Akihiro. (2005). A Validation Study on the English language test in a Japanese Nationwide University Entrance Examination. *The Asian EFL Journal*. Retrieved February 12, 2016, from http://www.asian-efl-journal.com/June_05_ai.pdf
- Jin, Yan. (2011). Fundamental Concerns in High-Stakes Language Testing: The Case of the College English Test. *Pan-Pacific Association of Applied Linguistics*, 15(2), 71-83.
- Jobsdb.com. (2015). 3 เหตุผลที่คนทำงานไม่สนใจภาษาอังกฤษไม่ได้แล้ว. สืบค้น 21 ธันวาคม 2558, จาก <http://th.jobsdb.com/th-th/articles/>
- LCRT Consulting. (2013). *MELA Validation Project: Final Report*. Retrieved November 11, 2015, from http://www.themela.com/assets/documents/MELAVValidationProjectFinal_Report_2013.pdf
- Madsen, Harold S. (1983). *Techniques in Testing*. Oxford: Oxford University Press.

- Munson, Robin & Parton, Cinda. (2013). *Bias and Fairness in State Testing*. Retrieved June 10, 2016, from <http://www.k12.wa.us/LegisGov/2013documents/BiasAndFairnessInStateTestingDec2013.pdf>
- Paltridge, Brian. (2007). Beyond the text: A textography of Chinese College English writing. *University of Sydney Papers in TESOL*, 2(2), 149-165.
- Pan, Yi-Ching. (2009). Test Impact: English Certification Exit Requirements in Taiwan. *TEFLIN Journal*, 20(2), 120-139.
- Riazi, Mehdi. (2014). *Research Note; Concurrent and Predictive Validating of Pearson Test of English Academic (PTE Academic)*. Retrieved: November 11, 2015, from http://pearsonpte.com/wp-content/uploads/2014/07/Riazi_M_2014.pdf
- 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 September 12, 2014, from http://clients.squareeye.net/uploads/eaquals2011/documents/EAQUALS_Calibrating_tests_to_the_CEFR.pdf
- Weir, Cyril J., Chan, Sathena H. C., & Nakatsuhara, Fumiyo. (2013). Examining the criterion-related validity of the GEPT advanced reading and writing tests: Comparing GEPT with IELTS and real-life academic performance. *LTTC-GEPT Research Reports*. Retrieved November 12, 2014, from https://www.ltcc.ntu.edu.tw/academics/ptresearch/Wu_and_Wu%20_2010_.pdf
- Weyant, Katie., & Chisholm, Amanda. (2014). Safeguarding Fairness Principles through the Test Development Process: A Tale of Two Organizations. *Cambridge English: Research Notes*, 55, 3-6. Retrieved June 10, 2016, from <http://www.cambridgeenglish.org/images/164055-research-notes-55-document.pdf>
- Whiting, Gilman. (2009). *Cultural Bias in Testing*. Retrieved June 10, 2016, from <http://www.education.com/reference/article/cultural-bias-in-testing/>
- Zhou, Yuemei. (2004). *Comparability Study of Two National EFL Tests (CET-6 and TEM-4) in China*. Retrieved November 11, 2015, from http://www.asiatefl.org/main/download_pdf.php?i=4&c=1391758451