

Chulalongkorn University

## Chula Digital Collections

---

Chulalongkorn University Theses and Dissertations (Chula ETD)

---

2019

การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนี  
ลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความ  
ก้าวหน้า

รังสิมากรณ์ หนูน้อย  
คณะครุศาสตร์

Follow this and additional works at: <https://digital.car.chula.ac.th/chulaetd>



Part of the [Educational Assessment, Evaluation, and Research Commons](#)

---

### Recommended Citation

หนูน้อย, รังสิมากรณ์, "การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า" (2019). *Chulalongkorn University Theses and Dissertations (Chula ETD)*. 9081.  
<https://digital.car.chula.ac.th/chulaetd/9081>

This Thesis is brought to you for free and open access by Chula Digital Collections. It has been accepted for inclusion in Chulalongkorn University Theses and Dissertations (Chula ETD) by an authorized administrator of Chula Digital Collections. For more information, please contact [ChulaDC@car.chula.ac.th](mailto:ChulaDC@car.chula.ac.th).

การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญ  
สูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า



น.ส.รังสิมาภรณ์ หนูน้อย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต

สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2562

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

DEVELOPMENT OF HYBRID COMPUTERIZED ADAPTIVE TESTING SYSTEM BY MAXIMUM  
PRIORITY INDEX METHOD AND PROGRESSIVE-RESTRICTED STANDARD ERROR METHOD



A Dissertation Submitted in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy in Educational Measurement and Evaluation

Department of Educational Research and Psychology

FACULTY OF EDUCATION

Chulalongkorn University

Academic Year 2019

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์
โดย	แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธี
สาขาวิชา	ค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	น.ส.รังสิมาภรณ์ หนูน้อย
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	การวัดและประเมินผลการศึกษา
	รองศาสตราจารย์ ดร.โชติกา ภาษีผล
	ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต

..... คณบดีคณะครุศาสตร์  
(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ  
(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก  
(รองศาสตราจารย์ ดร.โชติกา ภาษีผล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม  
(ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

..... กรรมการ  
(รองศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง)

..... กรรมการ  
(อาจารย์ ดร.สิวะโชติ ศรีสุทธิยากร)

..... กรรมการภายนอกมหาวิทยาลัย  
(อาจารย์ ดร.รังสรรค์ มณีเล็ก)

รังสิมาภรณ์ หนูน้อย : การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับ  
ความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า . ( DEVELOPMENT OF HYBRID  
COMPUTERIZED ADAPTIVE TESTING SYSTEM BY MAXIMUM PRIORITY INDEX METHOD AND  
PROGRESSIVE-RESTRICTED STANDARD ERROR METHOD) อ.ที่ปรึกษาหลัก : รศ. ดร.โชติกา ภาษีผล, อ.ที่  
ปรึกษาร่วม : ศ. ดร.ศิริชัย กาญจนวาสี

การศึกษาค้นคว้าครั้งนี้มีวัตถุประสงค์ เพื่อพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า และเพื่อประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดโดยการประเมินแบบอิงมาตรฐาน การประเมินฮิวริสติก และการประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบ โดยงานวิจัยครั้งนี้ศึกษาภายใต้ทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ใช้ข้อสอบแบบฝึกพัฒนาทักษะภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ของสถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จำนวน 640 ข้อ มีการตรวจสอบคุณภาพข้อสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 1,109 คน และทดลองใช้ระบบทดสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 74 คน เครื่องมือที่ใช้ในการประเมินคุณภาพระบบ ประกอบด้วยแบบประเมินอิงมาตรฐาน แบบประเมินฮิวริสติก และแบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด สถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ ความถี่ ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความแปรปรวนและความโด่ง

ผลการวิจัยสรุปได้ดังนี้

1. การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า ซึ่งทดสอบผ่านระบบออนไลน์ ระบบที่พัฒนาขึ้นประกอบด้วย 3 ส่วน ได้แก่ 1) รูปแบบของระบบการทดสอบ 2) การทดสอบ ซึ่งประกอบไปด้วย จุดเริ่มต้นการทดสอบ การประมาณค่าความสามารถผู้สอบด้วยวิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) การคัดเลือกข้อสอบด้วยวิธีดัชนีลำดับความสำคัญสูงสุด (MPI) การควบคุมการใช้ข้อสอบซ้ำด้วยวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (PR-SE) การยุติการทดสอบ ( $\leq 0.3$ ) และ 3) การรายงานผลการทดสอบ

2. ผลการประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ด้วยการประเมินแบบอิงมาตรฐานภาพรวมระบบการทดสอบมีคุณภาพอยู่ในระดับมากที่สุด ( $M = 4.59$ ,  $SD = 0.35$ ) ผลการประเมินระบบแบบฮิวริสติกภาพรวมระบบการทดสอบมีคุณภาพอยู่ในระดับมาก ( $M = 4.42$ ,  $SD = 0.33$ ) และผลการประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ภาพรวมนักเรียนมีความพึงพอใจระดับมากที่สุดต่อระบบการทดสอบ ( $M = 3.91$ ,  $SD = 0.50$ )

สาขาวิชา การวัดและประเมินผลการศึกษา

ปีการศึกษา 2562

ลายมือชื่อนิสิต .....

ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

# # 5784220627 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION

KEYWORD: HYBRID COMPUTERIZED ADAPTIVE TESTING SYSTEM, MAXIMUM PRIORITY INDEX METHOD,  
PROGRESSIVE-RESTRICTED STANDARD ERROR METHOD

Rungsimaporn Noonoi : DEVELOPMENT OF HYBRID COMPUTERIZED ADAPTIVE TESTING SYSTEM BY  
MAXIMUM PRIORITY INDEX METHOD AND PROGRESSIVE-RESTRICTED STANDARD ERROR METHOD.

Advisor: Assoc. Prof. SHOTIGA PASIPHOL, Ph.D. Co-advisor: Prof. SIRICHAI KANJANAWASEE, Ph.D.

This research aims to develop the hybrid computerized adaptive testing system by Maximum priority Index method and Progressive-restricted standard error method and evaluate quality of the hybrid computerized adaptive testing system by standard evaluation, heuristic evaluation and satisfaction evaluation. This research was studied within the concept of three-parameter item response model. The development of hybrid computerized adaptive testing system used the English skills development test for Grade 9 students, from English Language Institute, office of the basic education commission, was 640 items and examining the quality of items was 1,109 students in Grade 9 and 74 students in grade 9 tested the system. The tools comprised of standard evaluation scale, heuristic evaluation scale and satisfaction evaluation scale for hybrid computerized adaptive testing system. Data were analyzed by using frequency, percentage, mean, standard deviation, skewness and kurtosis.

The research finding were as follows:

1. The development of the hybrid computerized adaptive testing system by Maximum priority Index method and Progressive-restricted standard error method is online testing system. The system had three stages: 1) Testing system design 2) Test delivery, including initial item, parameter estimation by maximum likelihood estimation method (MLE), item selection by Maximum priority index method (MPI), item exposure control by Progressive-restricted standard error method (PR-SE) and test termination ( $\leq 0.3$ ), and 3) Test score report. Accompanying manuals of the CAT system were provided for users and system administrators.

2. The hybrid computerized adaptive testing system was standard evaluation the results showed that the quality of the system reached the highest levels ( $M = 4.59$ ,  $SD = 0.35$ ), heuristic evaluation the results showed that the quality of the system reached the high levels ( $M = 4.42$ ,  $SD = 0.33$ ) and the students were satisfied the results showed that the quality of the system reached the high levels ( $M = 3.91$ ,  $SD = 0.50$ ).

Field of Study: Educational Measurement and  
Evaluation

Student's Signature .....

Academic Year: 2019

Advisor's Signature .....

Co-advisor's Signature .....

## กิตติกรรมประกาศ

การทำวิทยานิพนธ์ฉบับนี้ ผู้วิจัยได้รับความรู้และคำแนะนำต่าง ๆ ด้วยความกรุณาอย่างยิ่งจาก รองศาสตราจารย์ ดร.โชติกา ภาชีผล อาจารย์ที่ปรึกษาวิทยานิพนธ์ และ ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ผู้ให้ความรู้ ข้อคิดเห็น และคำแนะนำในการพัฒนาระบบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ ด้วยความละเอียดถี่ถ้วนจนงานวิจัยฉบับนี้เสร็จสมบูรณ์

ขอขอบพระคุณ รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง และ ดร.สวีสัย ศรีสุทธียากร กรรมการสอบวิทยานิพนธ์ ดร.รังสรรค์ มณีเล็ก กรรมการภายนอก ที่ช่วยให้ข้อเสนอแนะอันเป็นประโยชน์ต่อการปรับปรุงวิทยานิพนธ์ให้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณสถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ในการอนุเคราะห์แบบฝึกทักษะภาษาอังกฤษมาใช้ในงานวิทยานิพนธ์ฉบับนี้ ขอขอบคุณผู้อำนวยการ ครู และนักเรียนโรงเรียนบุญวาทย์วิทยาลัย โรงเรียนสุนารีวิทยา โรงเรียนเศรษฐบุตรบำเพ็ญ โรงเรียนแก่ง "วิทยสถาวร" โรงเรียนท่ามะกาวิทยาคม โรงเรียนสตรีภูเก็ต และโรงเรียนพิชัยรัตนาคาร ที่อนุเคราะห์ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูล

ขอขอบพระคุณบัณฑิตวิทยาลัยจุฬาลงกรณ์มหาวิทยาลัยที่มอบ “ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย” จากกองทุนรัชดาภิเษกสมโภช สำหรับการจัดทำวิทยานิพนธ์ฉบับนี้

ขอขอบพระคุณ บิดา มารดาและน้องสาว ผู้ที่คอยให้กำลังใจและและความช่วยเหลือในทุก ๆ ด้านแก่ผู้วิจัย ขอขอบคุณรุ่นพี่ รุ่นน้อง และเพื่อนร่วมรุ่นสาขาวิชาการวัดและประเมินผลการศึกษา ผู้ที่คอยให้กำลังใจและคอยช่วยเหลือผู้วิจัยด้วยดีเสมอมา

รังสิมาภรณ์ หนูน้อย

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ค
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ช
สารบัญภาพ .....	ญ
บทที่ 1 บทนำ.....	1
ความเป็นมาและความสำคัญของการศึกษา .....	1
คำถามการวิจัย .....	8
วัตถุประสงค์ของการวิจัย.....	9
ขอบเขตของการวิจัย.....	9
คำจำกัดความที่ใช้ในการวิจัย.....	10
ประโยชน์ที่ได้รับ.....	13
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง .....	14
ตอนที่ 1 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (Computer Adaptive Testing) .....	14
ตอนที่ 2 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด .....	42
ตอนที่ 3 การประเมินผลระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์.....	67
ตอนที่ 4 มาตรฐานและตัวชี้วัดภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 .....	70
ตอนที่ 5 กรอบแนวคิดการวิจัย.....	80
บทที่ 3 วิธีดำเนินการวิจัย .....	83
ระยะที่ 1 การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด.....	83



ระยะที่ 2 การประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด .....	94
บทที่ 4 ผลการวิเคราะห์ข้อมูล .....	102
ระยะที่ 1 ผลการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด .....	102
ระยะที่ 2 ผลการประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด...	137
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ .....	143
สรุปผลการวิจัย.....	144
อภิปรายผลการวิจัย .....	147
ข้อเสนอแนะ .....	151
ภาคผนวก.....	154
ภาคผนวก ก รายนามผู้เชี่ยวชาญ .....	155
ภาคผนวก ข ผลการวิเคราะห์คุณภาพของข้อสอบรายข้อ (IRT) ตามทฤษฎีการตอบสนองข้อสอบ .....	157
ภาคผนวก ค เครื่องมือที่ใช้ในการวิจัย .....	177
ภาคผนวก ง คู่มือการใช้งานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด .....	180
บรรณานุกรม.....	203
ประวัติผู้เขียน .....	211

## สารบัญตาราง

ตาราง 2.1	ความแตกต่างระหว่างการทดสอบแบบประเพณีนิยมและการทดสอบที่เหมาะสม	16
ตาราง 2.2	จำแนกวิธีการควบคุมการใช้ข้อสอบซ้ำและจุดเด่นและข้อจำกัดของแต่ละวิธี	34
ตาราง 2.3	จุดแข็งและจุดอ่อนของการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายข้อ	50
ตาราง 2.4	จุดแข็งและจุดอ่อนของการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบหลายขั้นตอน	53
ตาราง 2.5	การเปรียบเทียบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายข้อ การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด (Hybrid design)	64
ตาราง 2.6	ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 1 ภาษาเพื่อการสื่อสาร มาตรฐาน ต 1.1	73
ตาราง 2.7	ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 1 ภาษาเพื่อการสื่อสาร มาตรฐาน ต 1.2	75
ตาราง 2.8	ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 1 ภาษาเพื่อการสื่อสาร มาตรฐาน ต 1.3	76
ตาราง 2.9	ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 2 ภาษาและวัฒนธรรม มาตรฐาน ต 2.1	76
ตาราง 2.10	ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 2 ภาษาและวัฒนธรรม มาตรฐาน ต 2.2	77
ตาราง 2.11	ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 3 ภาษากับความสัมพันธ์กับกลุ่มสาระการเรียนรู้อื่น มาตรฐาน ต 3.1	77
ตาราง 2.12	ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 4 ภาษากับความสัมพันธ์ กับชุมชนและโลก มาตรฐาน ต 4.1	78
ตาราง 2.13	ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 4 ภาษากับความสัมพันธ์ กับชุมชนและโลก มาตรฐาน ต 4.2	78
ตาราง 3.1	ผังข้อสอบจำแนกตามตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้ ภาษาต่างประเทศ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551	84
ตาราง 3.2	การดำเนินการทดสอบกับกลุ่มตัวอย่าง	89

ตาราง 3.3	ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินอิงมาตรฐาน.....	95
ตาราง 3.4	ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินอิวิริสติก.....	98
ตาราง 3.5	ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินความพึงพอใจของผู้ใช้งาน ต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด.....	99
ตาราง 4.1	จำนวนข้อสอบแบบฝึกทักษะภาษาอังกฤษ.....	103
ตาราง 4.2	ผลการเปรียบเทียบความสอดคล้องกลมกลืนระหว่างโมเดล 1PL, 2PL, และ 3PL ....	104
ตาราง 4.3	ตัวอย่างข้อสอบที่ผ่านการคัดเลือกและคัดออกของด้าน Conversation .....	105
ตาราง 4.4	ตัวอย่างข้อสอบที่ผ่านการคัดเลือกและคัดออกของด้าน Grammar .....	106
ตาราง 4.5	ตัวอย่างข้อสอบที่ผ่านการคัดเลือกและคัดออกของด้าน Reading.....	107
ตาราง 4.6	ตัวอย่างข้อสอบที่ผ่านการคัดเลือกและคัดออกของด้าน Miscellaneous .....	108
ตาราง 4.7	ค่าสถิติพื้นฐานของพารามิเตอร์ของข้อสอบ .....	108
ตาราง 4.8	จำนวนข้อสอบที่ผ่านการคัดเลือก .....	108
ตาราง 4.9	ค่าสถิติพื้นฐานของพารามิเตอร์ของข้อสอบที่ผ่านการคัดเลือก.....	109
ตาราง 4.10	ข้อมูลพื้นฐานของนักเรียนที่ทดลองใช้ระบบการทดสอบปรับเหมาะ ด้วยคอมพิวเตอร์แบบไฮบริด.....	136
ตาราง 4.11	ค่าสถิติพื้นฐานของความสามารถด้านทักษะภาษาอังกฤษของนักเรียน ระดับชั้นมัธยมศึกษาชั้นปีที่ 3 .....	136
ตาราง 4.12	ค่าสถิติพื้นฐานของคะแนนทักษะภาษาอังกฤษ .....	137
ตาราง 4.13	ผลการตรวจสอบคุณภาพของระบบการทดสอบในการประเมินอิงมาตรฐาน.....	138
ตาราง 4.14	ผลการตรวจสอบคุณภาพของระบบการทดสอบในการประเมินอิวิริสติก.....	140
ตาราง 4.15	การปรับปรุงแก้ไขตามข้อเสนอแนะของผู้เชี่ยวชาญ .....	141
ตาราง 4.16	ผลการศึกษาความพึงพอใจของผู้ใช้ที่มีต่อระบบการทดสอบปรับเหมาะด้วย คอมพิวเตอร์แบบไฮบริด .....	142

## สารบัญภาพ

ภาพ 2.1	โมเดลการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional IRT Model) (Y.-Y. Cheng, Wang, & Ho, 2008).....	17
ภาพ 2.2	ลักษณะของโมเดลการตอบสนองข้อสอบแบบพหุมิติระหว่างข้อสอบ A (Between-Items MIRT Model) และโมเดลการตอบสนองข้อสอบแบบพหุมิติภายในข้อสอบ B (Within-Items MIRT Model) (Y.-Y. Cheng et al., 2008) .....	18
ภาพ 2.3	ตัวอย่างขั้นตอนและกฎเส้นทางของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) (Brossman & Guille, 2014).....	52
ภาพ 2.4	ตัวอย่างโมเดล การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST): 2-stage โมเดลการออกแบบ 1 – 4 และ 3-stage โมเดลการออกแบบ 1 – 4 (Wang et al., 2016).....	53
ภาพ 2.5	แผนภาพการทำงานของการทำงานของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด.....	65
ภาพ 2.6	กรอบแนวคิดการวิจัยการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า.....	82
ภาพ 3.1	ตัวอย่างข้อสอบด้าน Conversation.....	85
ภาพ 3.2	ตัวอย่างข้อสอบด้าน Grammar .....	86
ภาพ 3.3	ตัวอย่างข้อสอบด้าน Reading.....	87
ภาพ 3.4	ตัวอย่างข้อสอบด้าน Miscellaneous .....	88
ภาพ 3.5	ตัวอย่างกระดาษคำตอบที่ใช้ในการทดสอบ .....	90
ภาพ 3.6	ขั้นตอนการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด .....	101
ภาพ 4.1	ชุดข้อสอบสำหรับการทดสอบขั้นแรกและขั้นที่สอง.....	109
ภาพ 4.2	ชุดข้อสอบสำหรับการสอบขั้นที่ 3 .....	110
ภาพ 4.3	ตัวอย่างข้อสอบที่บันทึกลงในฐานข้อมูล.....	111
ภาพ 4.4	การเข้าระบบของผู้จัดการระบบ .....	112
ภาพ 4.5	หน้าหลักของระบบการทดสอบ .....	113

ภาพ 4.6	การลงทะเบียน.....	113
ภาพ 4.7	การเข้าระบบของผู้ใช้งาน.....	114
ภาพ 4.8	หน้าจอหลัก.....	115
ภาพ 4.9	หน้าจอหลักระบบการทดสอบ.....	116
ภาพ 4.10	การลงทะเบียนผู้เข้าสอบ.....	117
ภาพ 4.11	หน้าจอคู่มือการใช้งานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด.....	117
ภาพ 4.12	การเข้าระบบการทดสอบ.....	118
ภาพ 4.13	หน้าจอแสดงรายละเอียดเกี่ยวกับผู้เข้ารับการทดสอบ.....	118
ภาพ 4.14	หน้าจอแก้ไขข้อมูลส่วนตัว.....	119
ภาพ 4.15	หน้าจอแสดงประวัติการทำข้อสอบ.....	120
ภาพ 4.16	หน้าจอแสดงรายละเอียดคำชี้แจง.....	121
ภาพ 4.17	หน้าจอข้อสอบ.....	121
ภาพ 4.18	หน้าจอรายงานผลการทดสอบ.....	122
ภาพ 4.19	แบบประเมินความพึงพอใจต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด.....	123
ภาพ 4.20	การเข้าระบบของผู้จัดการระบบ.....	124
ภาพ 4.21	หน้าจอหลักของผู้จัดการระบบ.....	125
ภาพ 4.22	การจัดการข้อมูลโรงเรียน.....	125
ภาพ 4.23	ตัวอย่างชุดข้อสอบสำหรับการทดสอบขั้นแรกและขั้นที่สอง.....	126
ภาพ 4.24	การเพิ่มข้อสอบ.....	127
ภาพ 4.25	ตัวอย่างข้อสอบที่จะแก้ไข.....	128
ภาพ 4.26	ตัวอย่างการแก้ไขข้อสอบ.....	129
ภาพ 4.27	ตัวอย่างการแก้ไขข้อมูลของสมาชิก.....	130
ภาพ 4.28	หน้าจอแสดงรายละเอียดคำชี้แจง.....	133

ภาพ 4.29	หน้าจอข้อสอบ .....	133
ภาพ 4.30	แบบประเมินความพึงพอใจต่อระบบการทดสอบปรับเหมาะ ด้วยคอมพิวเตอร์แบบไฮบริด.....	134
ภาพ 4.31	รายงานผลการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด .....	135



## บทที่ 1

### บทนำ

#### ความเป็นมาและความสำคัญของการศึกษา

การทดสอบปรับเหมาะเป็นการทดสอบที่ใช้ในการสอบที่มีความเหมาะสมกับความสามารถของผู้สอบ โดยลักษณะเด่นของการทดสอบนี้มุ่งที่จะวัดและคัดเลือกข้อสอบที่มีระดับความยากให้เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคนเพื่อให้สามารถประมาณค่าระดับความสามารถได้อย่างมีประสิทธิภาพ โดยมีหลักเกณฑ์ที่ว่าถ้าผู้สอบทำข้อสอบที่ยากได้ถูกต้องแล้วข้อง่ายก็น่าจะทำได้ถูกต้องหรือถ้าทำข้อสอบที่ง่ายผิดแล้วข้อที่ยากก็อาจจะผิดด้วยจึงไม่จำเป็นต้องทำข้อสอบข้อนั้น ดังนั้นในการทดสอบครั้งหนึ่งๆ ผู้สอบแต่ละคนไม่จำเป็นต้องทำข้อสอบเหมือนกันทุกข้อแต่จะขึ้นอยู่กับผลการตอบข้อสอบในแบบทดสอบของแต่ละคน ในการทดสอบจะให้ผู้สอบเริ่มทำข้อสอบที่มีความยากระดับปานกลางแล้วพิจารณาจากคำตอบของผู้สอบ ถ้าผู้สอบตอบถูกก็จะได้ตอบข้อสอบที่มีความยากเพิ่มขึ้น แต่ถ้าผู้สอบตอบผิดจะได้ข้อสอบที่มีความยากลดลง และในการทำข้อสอบข้อถัดไปจะปฏิบัติเช่นเดียวกันจนสิ้นสุดการสอบ การทดสอบปรับเหมาะเป็นตัวอย่างที่แสดงถึงประสิทธิภาพที่เหนือกว่าการทดสอบดั้งเดิมแบบเส้นตรงหรือการจำกัดรูปแบบข้อสอบ ซึ่งในการทดสอบจะใช้ข้อสอบน้อยกว่าในการตัดสินสอดคล้องกับการทดสอบปรับเหมาะที่ใช้จำนวนข้อสอบเท่ากับการทดสอบเส้นตรงแบบดั้งเดิม แต่ให้ผลการประมาณค่าความสามารถของผู้สอบที่แม่นยำกว่า (Drasgow, Luecht, & Bennett, 2006)

ปัจจุบันได้มีการนำเทคโนโลยีทางคอมพิวเตอร์มาใช้ในการทดสอบ ซึ่งเทคโนโลยีทางคอมพิวเตอร์ได้เข้ามาช่วยเสริมสร้างประสิทธิภาพในการตรวจข้อสอบ การวิเคราะห์ข้อสอบ การตัดเกรด และรายงานผล ซึ่งมีการใช้กันอย่างกว้างขวาง รวมทั้งยังมีการนำมาใช้สร้างคลังข้อสอบ (Item Bank) การทดสอบที่ใช้คอมพิวเตอร์ที่รู้จักกันดีคือ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (Computer Adaptive) ซึ่งเป็นหารทดสอบที่มีการคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ โดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ในการประมาณค่าความสามารถที่แท้จริงของผู้สอบ (ศิริชัย กาญจนวาสี, 2555)

การทดสอบปรับเหมาะเป็นวิธีการประเมินตามลำดับที่เลือกข้อคำถามจากคลังข้อสอบเพื่อให้ตรงกับระดับความสามารถของผู้สอบในการบริหารการทดสอบ สองวิธีได้รับความนิยมมากที่สุดในการออกแบบการทดสอบตามลำดับได้แก่ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบเลือกข้อสอบเป็นรายข้อ และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบเลือกข้อสอบเป็นชุดหลายขั้นตอน (MST) สองวิธีการนี้ได้ถูกนำไปใช้ในการวัดหลายงาน เช่น การทดสอบการศึกษา การรายงาน

การรักษาผู้ป่วย และการวัดคุณภาพชีวิต ตัวอย่างการดำเนินงานขนาดใหญ่ของการทดสอบ  
ปรับเหมาะด้วยคอมพิวเตอร์แบบดั้งเดิม ได้แก่ ACCUPLACER, GMAT และ ASVB และตัวอย่าง  
การดำเนินงานของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ได้แก่ GRE,  
AICPA และการวัดความก้าวหน้าทางวิชาการ (Luecht & Sireci, 2011)

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์ ในงานวิจัยนี้ขอเรียกว่า การทดสอบปรับเหมาะด้วย  
คอมพิวเตอร์แบบรายข้อ ใช้ข้อสอบเพียง 1 ข้อ ในการประมาณค่าความสามารถของผู้สอบและเลือก  
ข้อถัดไปจากค่าความสามารถที่ได้จากการประมาณค่าในข้อแรก ถ้าทำถูก ข้อถัดไปจะมีความยากขึ้น  
ถ้าทำผิดข้อถัดไปจะมีความยากลดลง ส่วนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน  
(MST) ใช้ชุดข้อสอบ โดยให้ผู้สอบทำข้อสอบทั้งหมด ในการประมาณค่าความสามารถของผู้สอบ เลือก  
ชุดข้อสอบข้อถัดไปโดยดูจากค่าความสามารถในการประมาณค่าในขั้นแรก โดยเลือกชุดที่มีความยาก  
ใกล้เคียงกับผู้สอบ ถ้าผู้สอบมีความสามารถต่ำจะเลือกชุดข้อสอบที่ง่ายลง ถ้าผู้สอบมีความสามารถ  
สูงจะเลือกชุดข้อสอบที่ยากขึ้น อย่างไรก็ตามการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลาย  
ขั้นตอน (MST) แตกต่างกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ ในจำนวนของข้อ  
คำถามที่ใช้ในการทดสอบตามเวลาที่กำหนด โดยปกติการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบหลายขั้นตอน (MST) จะกำหนดกลุ่มข้อสอบไว้ล่วงหน้าตามระดับความยาก ตัวอย่างเช่น ง่าย  
ปานกลาง และยาก และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ถูกมองว่า  
เป็นโมดูลที่ใช้ทดสอบการปรับเหมาะ แม้ว่าวัตถุประสงค์หลักของการทดสอบปรับเหมาะด้วย  
คอมพิวเตอร์แบบรายข้อ และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) คือ  
การประเมินประสิทธิภาพความสามารถของผู้สอบกับข้อคำถามน้อยกว่าที่จำเป็นต้องใช้ในการทดสอบ  
กระดาษและดินสอ แต่ทั้งสองวิธีก็มีจุดแข็งและจุดอ่อนที่ส่งผลต่อการประมาณค่าความสามารถ

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ มีจุดแข็งในเชิงทฤษฎีเป็นการทดสอบที่  
ผู้เข้ารับการทดสอบได้รับการจัดข้อสอบที่มีความยากตามระดับความสามารถของตน โดยที่ข้อสอบไม่  
ง่ายหรือยากจนเกินไปเหมือนกับการทดสอบแบบดั้งเดิมที่ใช้กระดาษในการทดสอบ และจำนวน  
ข้อสอบที่ใช้ในการทดสอบมีจำนวนลดลงโดยที่ไม่สูญเสียความแม่นยำในการประมาณ  
ค่าความสามารถของผู้เข้ารับการทดสอบ ขั้นตอนการเลือกข้อสอบจะเลือกข้อสอบแต่ละข้อภายหลัง  
การประมาณค่าความสามารถของผู้สอบจากคลังข้อสอบด้วยกระบวนการทางคอมพิวเตอร์ โดยอยู่บน  
พื้นฐานของผลการตอบข้อสอบจากข้อก่อนหน้า ซึ่งเป็นการควบคุมคุณภาพที่มีประสิทธิภาพเนื่องจาก  
สามารถลดความคลาดเคลื่อนที่เกิดจากคนในการจัดการข้อสอบได้ โดยมีการควบคุมเงื่อนไขต่างๆ  
เช่น ความครอบคลุมของเนื้อหา (content coverage) และอัตราการเปิดเผยของข้อสอบ  
(item exposure) เป็นต้น (Zheng & Chang, 2015; ทศน์ศิริรินทร์ สว่างบุญ ศิริเดช สุชีวะ ศิริชัย  
กาญจนวาสี และ Muraki, 2555) สำหรับจุดแข็งของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์



แบบหลายขั้นตอน (MST) มีอยู่ด้วยกันหลายประการคือ มีการจัดข้อสอบเป็นชุด กระบวนการในการเลือกข้อสอบเป็นชุดจะกระทำตามลำดับขั้น (stage) เท่านั้น ไม่ได้ใช้กระบวนการเลือกข้อสอบภายในขั้น ดังนั้นจึงอนุญาตให้ผู้สอบสามารถย้อนหรือข้ามข้อสอบภายในขั้นที่กำลังดำเนินการทดสอบได้ เพื่อทำการทบทวนหรืออาจเปลี่ยนคำตอบ ซึ่งต่างจากการทดสอบแบบรายข้อ ซึ่งผู้เข้ารับการทดสอบไม่สามารถกลับไปแก้ไขข้อสอบก่อนหน้าได้ ดังนั้น MST จึงทำให้ผู้เข้ารับการทดสอบรู้สึกเครียดและกังวลลดลงในระหว่างการทดสอบ นอกจากนี้กระบวนการในการจัดการทดสอบแบบ MST ยังอนุญาตให้ผู้พัฒนาแบบทดสอบสามารถกำกับติดตามและควบคุมคุณภาพของชุดแบบทดสอบเนื่องจากชุดแบบทดสอบจะถูกเตรียมขึ้นก่อนการทดสอบ ดังนั้นผู้เชี่ยวชาญในการพัฒนาการทดสอบสามารถทบทวนชุดแบบทดสอบจากหลากหลายมุมมอง อาทิเช่น การควบคุมความเท่าเทียมของเนื้อหา (content balance) ความเท่าเทียมกันของการเฉลยคำตอบ (answer key) เป็นต้น (Zheng, Nozawa, Gao, & Chang, 2012) นอกจากนี้ในการประมาณค่าความสามารถของผู้สอบเพื่อคัดเลือกชุดข้อสอบที่เหมาะสมกับความสามารถของผู้สอบจะทำการประมาณค่าจากชุดของข้อสอบในขั้นแรก ซึ่งต่างจากการทดสอบปรับเหมาะแบบรายข้อ จึงทำให้ลดความผิดพลาดจากการประมาณค่าความสามารถผู้สอบจากข้อสอบข้อแรกเพียงข้อเดียว ซึ่งหากเกิดความผิดพลาดในการประมาณค่าความสามารถจากข้อสอบข้อแรกจะทำให้ผู้เข้ารับการทดสอบอาจให้เวลานานกว่าปกติในการทดสอบ (Zheng & Chang, 2015) ด้านจุดอ่อนของการทดสอบปรับเหมาะหลายขั้นตอน คือต้องมีการจัดเตรียมชุดข้อสอบ (panel) และโมดูล (module) ของข้อสอบไว้ล่วงหน้า ซึ่งใช้คนในการดำเนินการ ดังนั้นหากเป็นการทดสอบที่มีขนาดใหญ่อาจเกิดความผิดพลาดในการจัดชุดข้อสอบได้

Wang, Lin, Chang, and Douglas (2016) ) ได้นำจุดแข็งในการดำเนินการของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มาศึกษาจนนำไปสู่การพัฒนาการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เพื่อเพิ่มประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ โดยการรวมกรอบแนวคิดของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อเข้าด้วยกัน ซึ่งเป็นเพียงกรอบแนวคิดและเป็นการศึกษากระบวนการต้นแบบ ยังไม่เคยมีการนำกรอบแนวคิดดังกล่าวมาใช้ในการทดสอบจริง โดยนำวิธีการประมาณค่าความสามารถและวิธีการคัดเลือกข้อสอบในการออกแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) มาใช้ในการทดสอบ เช่น สารสนเทศสูงสุดของฟิชเชอร์ (Maximum Fisher Information: MFI) และการประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood Estimate: MLE) สามารถนำมาใช้ในขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดได้ สำหรับขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จะมี 2 ขั้นตอน หรือ 3 ขั้นตอนนั้นขึ้นอยู่กับข้อเสนอนั้น และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST)

มีการลดความยาวของข้อสอบ โดยมีการศึกษาวิจัยเรื่องความยาวของโมดูลจะขึ้นอยู่กับขั้นตอนของเส้นทางที่ยาวซึ่งสามารถให้ประมาณการที่แม่นยำมากขึ้น (Hambleton & Xing, 2006; Jodoin, Zenisky, & Hambleton, 2006; Patsula, 1999) และในการทดสอบปรับเหมาะด้วยแบบไฮบริดมีการดำเนินการออกแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) โดยมีการกำหนดเส้นทางและลดความยาวโมดูลเพื่อตอบสนองต่อวัตถุประสงค์ในการประมาณค่าความสามารถที่ถูกต้อง สำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อจะใช้วิธี automatic test assembly (ATA) ในการตรวจสอบ

แนวคิดการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดนั้น จะเริ่มต้นการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ด้วยชุดข้อสอบที่กำหนดขึ้น (initial item) (MST ขั้นที่ 1) ที่มีความยากระดับปานกลาง จากนั้นระบบจะทำการประมาณค่าความสามารถของผู้สอบจากการตอบข้อสอบในชุดแรก และระบบจะคัดเลือกข้อสอบชุดถัดไปที่มีความยากใกล้เคียงกับความสามารถของผู้สอบ (MST ขั้นที่ 2) ในขั้นตอนนี้จะมีการแบ่งโมดูลข้อสอบออกเป็น 3 ระดับ ได้แก่ ระดับง่าย ระดับปานกลาง และระดับยาก เมื่อผู้สอบทำชุดข้อสอบตามระดับความสามารถแล้ว ระบบจะประมาณค่าความสามารถของผู้สอบจากการตอบข้อสอบในชุดที่ 2 เพื่อคัดเลือกข้อสอบถัดไปโดยการสุ่มข้อสอบทีละข้อที่ใกล้เคียงความสามารถของผู้สอบ ซึ่งเป็นส่วนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ จนถึงเกณฑ์การยุติการทดสอบ สำหรับกระบวนการเตรียมชุดข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จะมีการจัดเตรียมชุดข้อสอบไว้ล่วงหน้า และการเตรียมข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อจะใช้คอมพิวเตอร์ในการคัดเลือกข้อสอบแต่ละข้อ โดยมีการควบคุมอัตราการใช้ข้อสอบซ้ำ (exposure rates) และคุณสมบัติอื่นๆ ของกระบวนการในการเลือกข้อสอบ (Zheng & Chang, 2015)

ในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ จะมีองค์ประกอบที่สำคัญ ได้แก่ จุดเริ่มต้นการทดสอบ, การประมาณค่าความสามารถ, การคัดเลือกข้อสอบ, การควบคุมการใช้ข้อสอบซ้ำ และการยุติการทดสอบ ซึ่งการคัดเลือกข้อสอบนับว่าเป็นหัวใจสำคัญของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (Weiss & Kingsbury, 1984) วิธีการเลือกข้อสอบที่ดีต้องพิจารณาประสิทธิภาพสูงสุดในการประมาณค่าความสามารถให้ใกล้เคียงค่าความสามารถจริงของผู้สอบ รวมทั้งพิจารณาการจัดสมดุลของเนื้อหาและการควบคุมการเปิดเผยข้อสอบหรือการควบคุมการใช้ข้อสอบซ้ำไปพร้อมๆ กันด้วย ซึ่งการควบคุมสมดุลเนื้อหาเพื่อต้องการให้แน่ใจว่าแบบทดสอบที่แตกต่างกันในการจัดให้ผู้สอบแต่ละคนครอบคลุมสัดส่วนเนื้อหาเหมือนกัน สามารถอ้างอิงไปยังการวัดผู้สอบคนอื่นๆ ในลักษณะเดียวกัน และการควบคุมการใช้ข้อสอบซ้ำเพื่อไม่ต้องการให้ข้อสอบบางข้อนำไปใช้บ่อยเกินไปจนรู้จักกันไปในกลุ่มผู้สอบ

สำหรับวิธีการคัดเลือกข้อสอบ (item selection Method) วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) เสนอโดย Y. Cheng and Chang (2009) เพื่อใช้ควบคุมเงื่อนไขบังคับอย่างเข้มงวดในการคัดเลือกข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ จะพิจารณาความแปรปรวนของค่าสารสนเทศโดยการนำดัชนีไปเป็นตัวคูณให้กับค่าสารสนเทศของข้อสอบ ดังนั้นในการคัดเลือกข้อสอบแทนที่จะพิจารณาจากค่าสารสนเทศเพียงอย่างเดียววิธีการนี้จะนำการลำดับความสำคัญมาพิจารณาร่วมด้วย และสามารถปรับให้เข้ากับข้อบังคับที่ไม่ใช่ทางสถิติที่หลากหลาย และสามารถควบคุมได้หลายอย่างพร้อมๆ กัน เช่น ความสมดุลของเนื้อหา การควบคุมการใช้ข้อสอบซ้ำ ความสมดุลของคำตอบ วิธีดัชนีลำดับความสำคัญสูงสุด (MPI) สามารถนำไปปรับใช้กับโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ที่มีอยู่ในปัจจุบันได้ สะดวก โดยไม่จำเป็นต้องปรับค่าน้ำหนักความสัมพันธ์ระหว่างข้อบังคับและสารสนเทศข้อสอบ

การดำเนินงานโดยทั่วไปของการทดสอบที่มีการบริหารการทดสอบโดยใช้คอมพิวเตอร์ ข้อสอบในคลังข้อสอบจะถูกนำกลับมาใช้ใหม่ในการทดสอบครั้งต่อไปข้อสอบที่ถูกนำออกมาใช้บ่อยครั้งเกินไป (Overexposed Times) จนเป็นที่รู้จักทั่วไปในกลุ่มผู้สอบจะนำไปสู่ปัญหา ในเรื่องความปลอดภัยของแบบสอบ (Test Security) ซึ่งเกิดจากการร่วมใช้ข้อสอบ (Item Sharing) ระหว่างผู้สอบ ผู้สอบสามารถตอบข้อสอบถูกต้องได้โดยไม่ต้องใช้ความสามารถที่แท้จริงของตนเอง เนื่องจากผู้สอบอาจได้รับข้อมูลเกี่ยวกับแบบสอบและล่วงรู้ข้อสอบก่อน (Item Pre-Knowledge) จากผู้ที่เข้ารับการทดสอบก่อนหน้า คะแนนที่สังเกตได้ของผู้สอบที่รู้ข้อสอบก่อนจะขาดความถูกต้อง ไม่สะท้อนความสามารถที่แท้จริงของผู้สอบ ในขณะที่การทดสอบแบบดั้งเดิม การร่วมใช้ข้อสอบไม่เป็นประเด็นปัญหาสำคัญ เนื่องจากการทดสอบแบบดั้งเดิมส่วนมากมีการจัดการทดสอบตามช่วงเวลา ที่แน่นอนและข้อสอบที่ใช้ในการทดสอบแต่ละครั้งส่วนใหญ่ไม่ได้นำกลับมาใช้ใหม่ ซึ่งตรงกันข้ามกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (Chang & Zhang, 2002; Revuelta & Ponsoda, 1998; Way, 1998) เพื่อลดการคุกคามของ การร่วมใช้ข้อสอบระหว่างผู้สอบซึ่งมีผลต่อความน่าเชื่อถือของการวัดโดยตรง ดังนั้นการควบคุมการใช้ข้อสอบซ้ำ (Item Exposure Control) ต้องถูกกำหนดรวมเข้าในการคัดเลือกข้อสอบ (S.-Y. Chen, Ankenmann, & Spray, 2003; Davis & Dodd, 2005; Eggen, 2001; Leung, Chang, & Hau, 2002; Revuelta & Ponsoda, 1998; Stocking & Lewis, 1998, 2002)

Stocking and Lewis (1998) กล่าวว่าในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์กับความสามารถของผู้สอบจะต้องมีการควบคุมอัตราการใช้ข้อสอบซ้ำ เพื่อป้องกันผู้สอบที่สามารถจำแนกข้อสอบได้ ทำให้การประมาณค่าความสามารถของผู้สอบมีความถูกต้องแม่นยำ วิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PR-SE) พัฒนาโดย McClarty, Sperling, and Dodd (2006) โดยนำอัตราส่วนของ

ความคลาดเคลื่อนมาตรฐาน (SE) มาใช้ในกฎการหยุด โดยการใช้ความคลาดเคลื่อนมาตรฐาน (SE) ไม่ได้ขึ้นอยู่กับตำแหน่งตามลำดับแต่ใช้ความคลาดเคลื่อนมาตรฐาน เพื่อตรวจสอบองค์ประกอบ การสุ่ม หรือ สารสนเทศที่มีอิทธิพลต่อการถ่วงน้ำหนัก เช่นเดียวกับขั้นตอนการสุ่ม ขั้นตอนการทำงาน ร่วมกันง่ายกว่าวิธีการแบบเงื่อนไข วิธี PR-SE มีจุดแข็งในการดำเนินการบริหารจัดการเมื่อ เปรียบเทียบกับขั้นตอนการควบคุมการเปิดเผยข้อสอบวิธีอื่นๆ และการทดสอบปรับเหมาะ ด้วยคอมพิวเตอร์แบบกำหนดความยาวคงที่ และ แบบความยาวแปรผัน เนื่องจากจะกำหนดค่า ความคลาดเคลื่อนมาตรฐาน (SE) ของการประมาณค่าความสามารถผู้สอบ

เนื่องจาก วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) และ วิธีค่า ความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PR-SE) พัฒนาขึ้นภายใต้บริบทของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ ดังนั้นเพื่อขยายองค์ ความรู้ใหม่ผู้วิจัยจึงสนใจนำวิธีการคัดเลือกข้อสอบทั้ง 2 วิธีมาพัฒนาต่อในบริบทของการทดสอบปรับ เหมาะด้วยคอมพิวเตอร์แบบไฮบริดเพื่อให้ประโยชน์ในการประมาณค่าความสามารถที่ถูกต้องแม่นยำ

ในสังคมโลกปัจจุบัน การเรียนรู้ภาษาต่างประเทศมีความสำคัญและจำเป็นอย่างยิ่งใน ชีวิตประจำวัน เนื่องจากเป็นเครื่องมือสำคัญในการติดต่อสื่อสาร การศึกษา การแสวงหาความรู้ การประกอบอาชีพ การสร้างความเข้าใจเกี่ยวกับวัฒนธรรมและวิถีทัศน์ของชุมชนโลก และตระหนัก ถึงความหลากหลายทางวัฒนธรรมและมุมมองของสังคมโลก นำมาซึ่งมิตรไมตรีและความร่วมมือกับ ประเทศต่างๆ ช่วยพัฒนาผู้เรียนให้มีความเข้าใจตนเองและผู้อื่นดีขึ้น เรียนรู้และเข้าใจความแตกต่าง ของภาษาและวัฒนธรรม ขนบธรรมเนียมประเพณี การคิด สังคม เศรษฐกิจ การเมือง การปกครอง มีเจตคติที่ดีต่อการใช้ภาษาต่างประเทศ และใช้ภาษาต่างประเทศเพื่อการสื่อสารได้ รวมทั้งเข้าถึงองค์ ความรู้ต่างๆ ได้ง่ายและกว้างขึ้น และมีวิสัยทัศน์ในการดำเนินชีวิต ภาษาต่างประเทศที่เป็นสาระ การเรียนรู้พื้นฐาน ซึ่งกำหนดให้เรียนตลอดหลักสูตรการศึกษาขั้นพื้นฐาน คือ ภาษาอังกฤษ ส่วนภาษาต่างประเทศอื่น เช่น ภาษาฝรั่งเศส เยอรมัน จีน ญี่ปุ่น อาหรับ บาลี และภาษากลุ่มประเทศ เพื่อนบ้าน หรือภาษาอื่นๆ ให้อยู่ในดุลยพินิจของสถานศึกษาที่จะจัดทำรายวิชาและจัดการเรียนรู้ตาม ความเหมาะสม (กระทรวงศึกษาธิการ, 2552)

ภาษาอังกฤษเป็นภาษาต่างประเทศที่นิยมใช้กันอย่างแพร่หลายมากที่สุด เนื่องจากเป็น ภาษากลางที่ใช้สื่อความหมายไปเกือบทั่วโลก ทั้งในฐานะที่เป็นเครื่องมือที่จะเข้าถึงแหล่ง วิทยาการต่างๆ และค้นคว้าความรู้ใหม่ๆ ซึ่งปัจจุบัน มีคนทั่วโลกใช้ภาษาอังกฤษเป็นหลักใน การสื่อสารถึงจำนวน 2,000 ล้านคน หรือ 1 ใน 3 ของประชากรของโลก ดังนั้น จึงเป็นความจำเป็น อย่างยิ่งที่จะส่งเสริมให้ประชากรไทยได้เรียนรู้ภาษาอังกฤษในระดับที่จะสื่อสารได้ เป็นเครื่องมือใน การแสวงหาความรู้ และการประกอบอาชีพ ตลอดจนการเจรจาต่อรองสำหรับการแข่งขัน ด้านเศรษฐกิจและสังคมในเวทีสากล (กระทรวงศึกษาธิการ, 2553)

ประเทศไทยได้มีความพยายามที่จะเพิ่มขีดความสามารถในการใช้ภาษาอังกฤษ ด้วยการประกาศนโยบายให้นักเรียนเรียนภาษาอังกฤษเป็นภาษาต่างประเทศที่ 1 โดยกำหนดให้มีการสอนตั้งแต่ระดับประถมศึกษาเป็นต้นไป รวมทั้งการส่งเสริมให้มีการเปิดโรงเรียนนานาชาติ และโรงเรียนสองภาษา ซึ่งเป็นโรงเรียนที่จัดการเรียนการสอนตามหลักสูตรกระทรวงศึกษาธิการ โดยใช้ภาษาอังกฤษเป็นสื่อในการจัดการเรียนการสอน แม้จะมีมาตรการดังกล่าวแต่ยังไม่ส่งผลต่อความสามารถในการใช้ภาษาอังกฤษเท่าที่ควร กระทรวงศึกษาธิการมีนโยบายที่จะปฏิรูป การเรียนการสอนภาษาอังกฤษทั้งระบบ โดยใช้วิธีการปรับเปลี่ยนกระบวนทัศน์และพัฒนาการเรียนการสอนภาษาอังกฤษแบบสื่อสาร การสร้างความเสมอภาคในโอกาสการเรียนรู้ภาษาอังกฤษและการสร้างบรรยากาศและเพิ่มโอกาสในการเรียนรู้ภาษาอังกฤษนอกห้องเรียนและได้เสนอ "แผนการพัฒนาการศึกษาของกระทรวงศึกษาธิการ ฉบับที่ 12 (พ.ศ. 2560 – 2564) เพื่อยกระดับภาษาอังกฤษให้นักเรียนสามารถใช้ภาษาอังกฤษเพื่อการสื่อสารในชีวิตประจำวันได้ โดยการสร้างแรงจูงใจในการเรียนภาษา ปรับชั่วโมงการเรียนภาษาให้มากขึ้น เน้นการเรียนการสอนสำหรับนำไปใช้สื่อสารจริง มากกว่าหลักไวยากรณ์ จัดทำแอปพลิเคชันทางการศึกษา รวมทั้งพัฒนาครูภาษาอังกฤษ (กระทรวงศึกษาธิการ, 2559)

ในช่วงที่ผ่านมาประเทศไทยได้มีความพยายามที่จะพัฒนาความรู้ความสามารถในการใช้ภาษาอังกฤษโดยใช้ภาษาอังกฤษเป็นสื่อในการจัดการเรียนการสอน แต่ยังไม่ส่งผลต่อความสามารถในการใช้ภาษาอังกฤษของนักเรียน ดังจะเห็นได้จาก ผลการทดสอบทางการศึกษาระดับชาตินำขั้นพื้นฐาน หรือ โอเน็ต (O-NET) ปีการศึกษา 2562 ผลการประเมินด้านผลสัมฤทธิ์ทางการเรียน วิชาภาษาอังกฤษ ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 พบว่ามีคะแนนผลสัมฤทธิ์อยู่ในระดับต่ำ กล่าวคือ มีคะแนนเฉลี่ยระดับประเทศ 33.25 และเมื่อเปรียบเทียบกับวิชาอื่นๆ พบว่า ผลการประเมินด้านผลสัมฤทธิ์ทางการเรียนวิชาภาษาอังกฤษอยู่ในระดับต่ำสุด โดยการประเมินผลสัมฤทธิ์ทางการเรียนวิชาภาษาอังกฤษเมื่อจำแนกตามสาระพบว่าผลการประเมินผลสัมฤทธิ์ทางการเรียนวิชาภาษาอังกฤษ สาระที่ 1 ภาษาเพื่อการสื่อสาร มีค่าเฉลี่ย 33.75 และ สาระที่ 2 ภาษาและวัฒนธรรม มีค่าเฉลี่ย 29.55 (สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน), 2562)

งานวิจัยของ ญาณิกา ลุนราศรี (2557) ที่ศึกษาเรื่อง ผลกระทบของการทดสอบทางการศึกษาระดับชาตินำขั้นพื้นฐานที่มีต่อการเรียนภาษาอังกฤษตามการรับรู้ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 พบว่า นักเรียนมีความรู้สึกรัดกักงวลในการสอบ นักเรียนส่วนใหญ่ใช้ทักษะการจำเป็นกลยุทธ์ที่ใช้ในการเรียนเป็นบางครั้งและรู้สึกกลัวว่าจะได้คะแนนสอบโอเน็ตต่ำเป็นประจำ และนักเรียนส่วนใหญ่เห็นด้วยกับการสอบโอเน็ตที่ใช้วัดความสามารถทางภาษาอังกฤษและเนื้อหาในข้อสอบเน้นทักษะการคิดวิเคราะห์ อย่างไรก็ตามนักเรียนส่วนใหญ่ไม่แน่ใจว่าเนื้อหาในข้อสอบออกตรงตามเนื้อหาในหนังสือเรียน สถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ตระหนักถึง

ความสำคัญในการพัฒนาแบบฝึกทักษะภาษาอังกฤษ ชั้นมัธยมศึกษาปีที่ 3 ว่าจะเป็นเครื่องมือที่สำคัญหนึ่งที่มีส่วนช่วยให้ผู้เรียนมีความคุ้นเคยรูปแบบข้อสอบ คำสั่งในข้อสอบและการฝึกฝนการประเมิน ส่งผลต่อการยกระดับผลสัมฤทธิ์การทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) ดังนั้นสถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จึงพัฒนาแบบฝึกทักษะภาษาอังกฤษ ระดับชั้นมัธยมศึกษาปีที่ 3 เพื่อเป็นประโยชน์สำหรับสำนักงานเขตพื้นที่การศึกษา ศึกษาานิเทศก์และครูผู้สอนในการนำไปใช้ เพื่อก่อให้เกิดการพัฒนาอย่างต่อเนื่อง

จากที่กล่าวมาข้างต้นทำให้ผู้วิจัยสนใจที่จะพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) ในการคัดเลือกข้อสอบ และ วิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PRSE) ในการควบคุมการใช้ข้อสอบซ้ำ เนื่องจากที่ผ่านมาเป็นเพียงกรอบแนวคิดและเป็นการศึกษากระบวนการต้นแบบ ยังไม่เคยมีการนำกรอบแนวคิดดังกล่าวมาใช้ในการทดสอบจริง เพื่อให้ผลการทดสอบมีความชัดเจนมากขึ้นและเห็นแนวทางในการนำมาปฏิบัติจริง ผู้วิจัยจึงพัฒนาการทดสอบโดยใช้ข้อสอบจากคลังข้อสอบแบบฝึกพัฒนาทักษะภาษาอังกฤษ สถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ และเก็บรวบรวมข้อมูลจริงจากนักเรียนชั้นมัธยมศึกษาปีที่ 3 ซึ่งการศึกษาครั้งนี้จะเป็นเครื่องมือที่มีส่วนช่วยให้นักเรียนพัฒนาทักษะภาษาอังกฤษ และเป็นแนวทางในการนำการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดมาใช้ในการทดสอบขนาดใหญ่

### คำถามการวิจัย

1. ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า มีลักษณะเป็นอย่างไร
2. ผลประเมินระบบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีลักษณะเป็นอย่างไร

## วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์ดังนี้

1. เพื่อพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้การคัดเลือกข้อสอบวิธีลำดับความสำคัญสูงสุด และการควบคุมอัตราการใช้ข้อสอบซ้ำด้วยวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า
2. เพื่อประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยการประเมินแบบอิงมาตรฐาน การประเมินฮิวริสติก และการประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบ

## ขอบเขตของการวิจัย

1. การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เป็นการนำแนวคิดการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ (CAT) มาประยุกต์ใช้ร่วมกัน ซึ่งเป็นเพียงกรอบแนวคิดและเป็นการศึกษากระบวนการต้นแบบ ยังไม่เคยมีการนำกรอบแนวคิดดังกล่าวมาใช้ในการทดสอบจริง ในการศึกษาครั้งนี้จะพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด จากการถอดขั้นตอนการทำงาน โดยใช้วิธีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) ในการคัดเลือกข้อสอบ และ วิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PR-SE) ในการควบคุมการใช้ข้อสอบซ้ำ นำมาพัฒนาเป็นระบบสำหรับการทดสอบแบบฝึกทักษะภาษาอังกฤษสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3

2. ข้อสอบที่นำมาใช้เป็นแบบฝึกพัฒนาทักษะภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ของสถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ตั้งแต่ปี พ.ศ. 2556 – 2559 จำนวน 640 ข้อ ที่มีการเผยแพร่ข้อสอบพร้อมเฉลยผ่านเว็บไซต์สถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ มีรูปแบบของคำตอบเป็นแบบ 4 ตัวเลือก โดยมีการมุ่งวัดความรู้ทางภาษาอังกฤษ ซึ่งประกอบด้วย 4 ด้าน ได้แก่ ด้าน Conversation, ด้าน Grammar and Vocabulary, ด้าน Reading Comprehension และ ด้าน Miscellaneous ตามตัวชี้วัดและสาระการเรียนรู้แกนกลางกลุ่มสาระการเรียนรู้ภาษาต่างประเทศ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ข้อสอบผ่านการตรวจสอบโดยครูผู้เชี่ยวชาญทางด้านภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน และตรวจสอบและเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบ เพื่อยืนยันคุณภาพของข้อสอบ พิจารณาจากค่าสถิติ G2 (deviance statistic) ค่า AIC (Akaike Information Criterion) ค่า BIC (Bayesian information criterion) และตรวจสอบคุณภาพของข้อสอบรายข้อ

ตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ มีเกณฑ์ในการคัดเลือกข้อสอบเข้าคลัง คือ ค่าความยาก (b) มีค่า -4 ถึง 4 ค่าอำนาจจำแนก (a) มีค่า 0 ถึง  $\infty$  และ ค่าการเดา (c) มีค่าไม่เกิน 0.50 ดำเนินการวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ของข้อสอบด้วยโปรแกรม R ด้วยชุดคำสั่ง ltm เป็นคำสั่งสำเร็จรูปที่ใช้ในการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (Item response theory: IRT)

3. ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด สามารถใช้งานผ่านระบบออนไลน์ โดยเป้าหมายของการพัฒนาระบบเพื่อเป็นเครื่องมือในการฝึกทักษะภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3 และสามารถนำผลการทดสอบมาใช้พัฒนาระดับความสามารถทางทักษะภาษาอังกฤษของตนเอง

4. การประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีการประเมิน 3 ส่วน ได้แก่ ส่วนที่ 1 การประเมินแบบอิงมาตรฐาน ได้แก่ ความมีประโยชน์ ความเป็นไปได้ ความเหมาะสม และความถูกต้อง ส่วนที่ 2 การประเมินแบบฮิวริสติก (heuristic evaluation) ได้แก่ ความชัดเจนของระบบ (Visibility of system status) ความสอดคล้องของระบบและการใช้งานจริง (Match between system and the real world) การควบคุมและอิสระของผู้ใช้งาน (User control and freedom) ความสอดคล้องและมาตรฐาน (Consistency and standards) การป้องกันความผิดพลาด (Error prevention) การจำข้อมูลมากกว่าการเรียกคืนข้อมูล (Recognition rather than recall) ความยืดหยุ่นและประสิทธิภาพการใช้งาน (Flexibility and efficiency of use) การออกแบบที่สวยงาม (Aesthetic and minimalist design) การช่วยเหลือโดยการกู้คืนระบบจากข้อผิดพลาด (Help users recognize, diagnose, and recover from errors) และการช่วยเหลือและคู่มือ (Help and documentation) และส่วนที่ 3 การประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ได้แก่ 1) ภาพรวมของระบบ 2) หน้าจอของระบบ 3) การเรียนรู้ และ 4) ความสามารถของระบบ

### คำจำกัดความที่ใช้ในการวิจัย

1. ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด (Hybrid Computerized Adaptive Testing System) หมายถึง ชุดขององค์ประกอบที่รวมกรอบแนวคิดของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ (CAT) เข้าด้วยกัน ประกอบด้วย 3 ส่วน ได้แก่ 1) รูปแบบของระบบการทดสอบ ที่มีการกำหนดสิทธิ์ในการเข้าถึงในการใช้งานของระบบ โดยแบ่งออกเป็น ผู้จัดการระบบ (admin) และผู้ใช้งาน (user) และคู่มือการใช้ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด 2) การทดสอบ ซึ่งประกอบไปด้วย จุดเริ่มต้นการทดสอบ การประมาณค่าความสามารถ



ของผู้สอบด้วยวิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (maximum likelihood estimation: MLE) การคัดเลือกข้อสอบด้วยวิธีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) การควบคุมการใช้ข้อสอบซ้ำด้วยวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PR-SE) การยุติการทดสอบ โดยการกำหนดค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ น้อยกว่าหรือเท่ากับ 0.30 และ 3) การรายงานผลการทดสอบที่มีการรายงานผลการตอบข้อสอบและกราฟแสดงความสัมพันธ์ของการตอบข้อสอบ

**2. การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด (Hybrid Computer Adaptive Testing)** หมายถึง การทดสอบปรับเหมาะด้วยคอมพิวเตอร์ที่รวมแนวคิดการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ (CAT) มาใช้ร่วมกัน โดยมีการบริหารการสอบ ดังนี้ เริ่มต้นการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ด้วยชุดข้อสอบที่กำหนดขึ้น (MST ขั้นที่ 1) ที่มีความยากระดับปานกลาง จากนั้นจะทำการประมาณค่าความสามารถของผู้สอบจากการตอบข้อสอบของผู้สอบในชุดแรกและระบบจะคัดเลือกข้อสอบชุดถัดไปที่มีความยากใกล้เคียงกับความสามารถของผู้สอบ (MST ขั้นที่ 2) ในขั้นตอนนี้จะมีการแบ่งโมดูลข้อสอบออกเป็น 3 ระดับ ได้แก่ ระดับง่าย ระดับปานกลาง และระดับยาก เมื่อผู้สอบทำชุดข้อสอบตามระดับความสามารถแล้ว ระบบจะประมาณค่าความสามารถของผู้สอบจากการตอบข้อสอบในชุดที่ 2 เพื่อคัดเลือกข้อสอบถัดไปโดยการสุ่มข้อสอบที่ละข้อที่ใกล้เคียงความสามารถของผู้สอบ ซึ่งเป็นส่วนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ จนถึงเกณฑ์การยุติการทดสอบ

**3. การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ (computerized adaptive testing)** หมายถึง การทดสอบที่มีการคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ ซึ่งเป็นการใช้ระบบคอมพิวเตอร์ในการสอบเพื่อประเมินความสามารถของผู้สอบแต่ละคน มีการคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ เริ่มต้นจากข้อสอบข้อที่ยากง่ายปานกลาง และเมื่อทำถูกจะได้ข้อสอบข้อถัดไปที่ยากขึ้นเรื่อยๆ โดยกระบวนการสอบจะดำเนินการไปจนกระทั่งประเมินความสามารถของผู้สอบได้จนค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) จนถึงเกณฑ์การยุติการทดสอบ

**4. การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (Multi stage testing: MST)** หมายถึง การทดสอบที่มีการกำหนดชุดของข้อสอบที่ถูกเลือกตามความสามารถของผู้สอบก่อนที่จะทำข้อสอบชุดนั้น โดยมีการประมาณค่าความสามารถของผู้สอบมาจากโมดูลของข้อสอบในแต่ละขั้น (Stage) และในแต่ละโมดูลจะมีข้อสอบที่มีความยากใกล้เคียงกัน

**5. การคัดเลือกข้อสอบ** หมายถึง การคัดเลือกชุดข้อสอบและข้อสอบรายข้อที่เหมาะสมกับความสามารถของผู้สอบ พิจารณาจากผลการตอบข้อสอบที่ผ่านมาของผู้สอบแต่ละคน หลังจากนั้นจะทำการประมาณค่าความสามารถของผู้สอบจากผลการตอบข้อสอบ เพื่อทำการคัดเลือกข้อสอบถัดไปที่มีความยากและอำนาจจำแนกที่เหมาะสมกับระดับความสามารถของผู้สอบ

**6. วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI)** หมายถึง วิธีการคัดเลือกข้อสอบจากพิจารณาค่าสารสนเทศควบคู่กับลำดับความสำคัญ การคัดเลือกข้อสอบจะพิจารณาจากผลคูณของค่าสารสนเทศของข้อสอบกับดัชนีความสำคัญ ข้อใดให้ผลคูณมากที่สุด ข้อนั้นจะได้รับการคัดเลือกให้นำไปใช้กับผู้สอบ

**7. การควบคุมการใช้ข้อสอบซ้ำ** หมายถึง วิธีการในการวางแผนป้องกันการใช้ข้อสอบแต่ละข้อในคลังข้อสอบถูกนำมาใช้บ่อยครั้งเกินไป

**8. วิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted error: PR-SE)** หมายถึง วิธีการควบคุมการใช้ข้อสอบโดยคำนวณน้ำหนักข้อสอบแต่ละข้อจากอัตราส่วนของความคลาดเคลื่อนมาตรฐานกฎการหยุด เพื่อตรวจสอบองค์ประกอบการสุ่มหรือสารสนเทศที่มีอิทธิพลต่อการถ่วงน้ำหนัก สุ่มเลือกจำนวนข้อสอบ กับตำแหน่งความสัมพันธ์ของข้อสอบ และสารสนเทศของข้อสอบ

**9. อัตราการใช้ข้อสอบซ้ำสูงสุด** หมายถึง ตัวเลขที่แสดงถึงค่าเป้าหมายที่ตั้งขึ้นเพื่อนำไปใช้วางแผนป้องกันการใช้ข้อสอบแต่ละข้อในคลังข้อสอบบ่อยครั้งเกินไป โดยการวิจัยครั้งนี้กำหนดอัตราการใช้ข้อสอบสูงสุดไว้ที่ 20 เปอร์เซ็นต์

**10. คลังข้อสอบ** หมายถึง แหล่งรวบรวมข้อสอบที่มีคุณภาพ ซึ่งผ่านกระบวนการวิเคราะห์ข้อสอบตามทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) ใช้รูปแบบโลจิสติก 3 พารามิเตอร์ (3 Parameter Logistic Model : 3 PLM) ประกอบด้วย ค่าพารามิเตอร์ความยากของข้อสอบ (b) มีค่าตั้งแต่ -4 ถึง 4 ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง  $\infty$  และ ค่าพารามิเตอร์การเดาของข้อสอบ (c) มีค่าไม่เกิน 0.50 จัดเก็บไว้บนฐานข้อมูลอย่างเป็นระบบและมีใช้คอมพิวเตอร์ช่วยในการนำข้อสอบมาใช้

**11. เกณฑ์การยุติการทดสอบ** หมายถึง ข้อกำหนดเพื่อใช้ในการสิ้นสุดการทดสอบ ซึ่งกำหนดโดยใช้ระดับความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถผู้สอบน้อยกว่าหรือเท่ากับ 0.3

**12. การประเมินคุณภาพระบบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด** หมายถึง ผลการประเมินความสามารถในการทำงานของระบบการทดสอบจากผู้เชี่ยวชาญและผู้ใช้งานระบบ ซึ่งมีการประเมิน 3 ส่วน ได้แก่ การประเมินแบบอิงมาตรฐาน การประเมินแบบ

ฮิวริสติก และการประเมินความพึงพอใจของผู้ใช้ที่มีต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดโดยผู้ใช้งาน มีการประเมิน ดังนี้

12.1 การประเมินแบบอิงมาตรฐาน มีการประเมิน 4 ด้าน ได้แก่ 1) ความมีประโยชน์ 2) ความเป็นไปได้ 3) ความเหมาะสม และ 4) ความถูกต้อง

12.2 การประเมินแบบฮิวริสติก มีการประเมิน 10 ด้าน ได้แก่ 1) ความชัดเจนของระบบ 2) ความสอดคล้องของระบบและการใช้งานจริง 3) การควบคุมและอิสระของผู้ใช้งาน 4) ความสอดคล้องและมาตรฐาน 5) การป้องกันความผิดพลาด 6) การจำข้อมูลมากกว่าการเรียกคืนข้อมูล 7) ความยืดหยุ่นและประสิทธิภาพการใช้งาน 8) การออกแบบที่สวยงาม 9) การช่วยเหลือโดยการกู้คืนระบบจากข้อผิดพลาด และ 10) การช่วยเหลือและคู่มือ

12.3 การประเมินความพึงพอใจของผู้ใช้ที่มีต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดโดยผู้ใช้งาน มีการประเมิน 4 องค์ประกอบ ได้แก่ 1) ภาพรวมของระบบ 2) หน้าจอของระบบ 3) การเรียนรู้ และ 4) ความสามารถของระบบ

## ประโยชน์ที่ได้รับ

การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดังกล่าวดำเนินความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้าได้สารสนเทศที่เป็นประโยชน์ทั้งด้านวิชาการและด้านการปฏิบัติ ดังนี้

### 1. ประโยชน์ด้านวิชาการ

1.1 ได้แนวทางในการจัดทำคลังข้อสอบ การวิเคราะห์ข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบที่มีการตรวจสอบโมเดลที่เหมาะสม เพื่อให้การประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบมีประสิทธิภาพมากที่สุด

1.2 ได้แนวทางการศึกษาวิธีการพัฒนาเงื่อนไขอื่น ๆ ที่นำมาใช้กับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เช่น การสร้างคลังข้อสอบ การคัดเลือกข้อสอบเริ่มต้น การประมาณค่าความสามารถ และเกณฑ์การยุติการทดสอบ เป็นต้น

### 2. ประโยชน์ด้านการปฏิบัติ

2.1 ครู อาจารย์และบุคลากรทางการศึกษาสามารถนำระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด รายวิชาภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 เป็นเครื่องมือในการฝึกทักษะภาษาอังกฤษของนักเรียน และสามารถนำผลการทดสอบมาใช้พัฒนาระดับความสามารถทางทักษะภาษาอังกฤษในตัวผู้เรียนเป็นรายบุคคลได้

2.2 สามารถนำแนวทางในการจัดทำระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ไปประยุกต์ใช้กับในการพัฒนาการทดสอบประเภทอื่นๆ ได้

## บทที่ 2

### เอกสารและงานวิจัยที่เกี่ยวข้อง

เนื้อหาในส่วนนี้ผู้วิจัยขอนำเสนอผลการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า ซึ่งผู้วิจัยได้นำเสนอเนื้อหาออกเป็น 5 ตอนดังนี้

#### ตอนที่ 1 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์

- 1.1 ความเป็นมาของการทดสอบปรับเหมาะ
- 1.2 หลักการของการทดสอบปรับเหมาะ
- 1.3 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์
- 1.4 ประเภทของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์
- 1.5 องค์ประกอบของระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์
- 1.6 งานวิจัยที่เกี่ยวข้องกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์

#### ตอนที่ 2 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

- 2.1 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายชื่อ
- 2.2 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน
- 2.3 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

#### ตอนที่ 3 การประเมินระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์

#### ตอนที่ 4 มาตรฐานและตัวชี้วัดภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3

#### ตอนที่ 5 กรอบแนวคิดการวิจัย

โดยแต่ละตอนมีรายละเอียดดังต่อไปนี้

#### ตอนที่ 1 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (Computer Adaptive Testing)

##### 1.1 ความเป็นมาของการทดสอบปรับเหมาะ

การทดสอบปรับเหมาะกับความสามารถของผู้สอบถูกนำมาใช้ครั้งแรกกับการทดสอบเชาว์ปัญญาด้วยแบบสอบของบินเน็ต (Binet) ในปี ค.ศ. 1908 โดยผู้ที่ดำเนินการทดสอบสามารถใช้ดุลยพินิจถึงระดับความยากของข้อสอบเริ่มต้น ถ้าข้อสอบเริ่มต้นยากเกินไป หรือง่ายเกินไปสำหรับผู้สอบ ผู้ดำเนินการทดสอบสามารถเลือกใช้ข้อสอบที่ง่ายลงหรือยากขึ้นได้ตามความเหมาะสม ซึ่งผู้สอบแต่ละคนจะได้ทำข้อสอบที่ยากขึ้นเรื่อยๆ จนถึงข้อที่ยาก จนทำไม่ถูกจำนวนหนึ่งการทดสอบก็จะยุติลง

โดยในการให้คะแนนนั้นตั้งอยู่บนข้อตกลงเบื้องต้นว่าถ้าผู้สอบทำกลุ่มข้อสอบใดได้ถูกต้องย่อมจะทำข้อสอบข้อที่ง่ายกว่านั้นได้ถูกต้องด้วย แต่อาจจะทำข้อที่ยากกว่านั้นไม่ได้ (ศิริชัย กาญจนวาสี, 2555)

Wright และ Douglas (1975) อ้างถึงใน ศิริชัย กาญจนวาสี (2555) ได้นำเสนอวิธีการทดสอบปรับเหมาะที่สร้างสเกลความยากของข้อสอบจากการวิเคราะห์ด้วยโมเดลแบบ -1 พารามิเตอร์ของราซส์ โดยผู้สอบได้รับแบบสอบที่เรียงข้อสอบตามระดับความยากและแนะนำให้ผู้สอบเริ่มทำข้อสอบข้อที่มีความยากปานกลาง ในการทำข้อสอบให้เลือกทำข้อที่ง่ายลง และยากขึ้นได้จนผู้สอบพบว่าข้อใดบ้างที่ยากเกินไปหรือง่ายเกินไปสำหรับตนเอง คะแนนของผู้สอบได้จากการรวมจำนวนข้อที่ตอบถูกกับจำนวนข้อที่ตนเองตัดสินว่าง่ายเกินไป ซึ่งวิธีการให้คะแนนลักษณะนี้ยังเป็นเรื่องที่ถูกเถียงกันถึงความเหมาะสม โดยลอร์ด (Lord, 1969 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) ได้ใช้วิธีการทดสอบปรับเหมาะแบบ 2 ขั้นตอน โดยขั้นตอนแรกเป็นการใช้แบบสอบฉบับสั้นสำหรับจัดระดับความสามารถของผู้สอบ ซึ่งเรียกว่า “Routing Test” เพื่อดูว่าผู้สอบควรใช้แบบสอบในขั้นที่สอง (Second – stage test) ระดับความยากเท่าใด โดยลอร์ดได้เสนอวิธีการประมาณค่าความสามารถของผู้สอบจากการทำแบบสอบทั้งสองฉบับร่วมกัน ผลการวิจัยจากการสร้างสถานการณ์จำลองพบว่าการทดสอบปรับเหมาะแบบ 2 ขั้นตอนให้ผลการประมาณค่าความสามารถของผู้สอบกลุ่มเก่งและกลุ่มอ่อนได้ดีกว่า วิธีการทดสอบขั้นตอนเดียวแบบประเพณีนิยม แต่ก็ให้ผลที่ไม่แตกต่างกันสำหรับกลุ่มผู้สอบที่มีความสามารถในระดับปานกลาง ต่อมาในปี 1980 ลอร์ด (Lord, 1980 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) ได้ทำการวิจัยเกี่ยวกับการทดสอบที่สถาบันบริการทางการทดสอบทางการศึกษา (ETS) ได้ให้ความสนใจการทดสอบปรับเหมาะ โดยแบบสอบที่มีความยาวคงที่ไม่น่าจะมีประสิทธิภาพสำหรับผู้สอบส่วนใหญ่ โดยเฉพาะอย่างยิ่งผู้สอบที่มีความสามารถสูง และความสามารถต่ำ เนื่องจากผู้สอบที่มีความสามารถสูงไม่จำเป็นต้องทำข้อสอบที่ง่าย แต่ผู้สอบที่มีความสามารถต่ำก็ไม่จำเป็นต้องทำข้อสอบข้อที่ยากมากๆ เพราะข้อสอบจะให้สารสนเทศสำหรับผู้สอบน้อยมาก โดยสามารถทำให้แบบสอบสั้นลงได้ และไม่สูญเสียความแม่นยำในการประมาณค่าความสามารถของผู้สอบ

## 1.2 หลักการของการทดสอบปรับเหมาะ

ในการทดสอบปรับเหมาะกับความสามารถของผู้สอบ หลักการคัดเลือกข้อสอบสำหรับแต่ละบุคคลอยู่บนพื้นฐานของผลการตอบข้อสอบข้อที่ผ่านมาของผู้สอบนั้น เมื่อเหมาะสมกับความสามารถของผู้สอบมีประสิทธิภาพเหนือกว่าการทดสอบ แบบประเพณีนิยม (Lord, 1971; Hulin, Drasgow and Parsons, 1982; Weiss, 1988 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) ซึ่งเป็นลักษณะที่ตรงกันข้ามกับแบบทดสอบแบบประเพณีนิยม (conventional tests) ที่ผู้สอบทุกคนจะถูกสอบด้วยแบบทดสอบชุดเดียวกัน Weiss (2004) ทั้งนี้ Jian-quan, Dan-min, Xia, and Jing-jing (2007) ได้

สรุปความ แตกต่างระหว่างการทดสอบแบบประเพณีนิยมและการทดสอบที่เหมาะสมไว้ ดังตาราง 2.1

**ตาราง 2.1 ความแตกต่างระหว่างการทดสอบแบบประเพณีนิยมและการทดสอบที่เหมาะสม**

รายการ	แบบทดสอบแบบประเพณีนิยม	แบบทดสอบแบบที่เหมาะสม
องค์ประกอบของแบบทดสอบ	ผู้สอบแต่ละคนใช้แบบทดสอบชุดเดียวกัน	ผู้สอบแต่ละคนใช้ข้อสอบที่ต่างกัน
ผู้สอบแต่ละคนใช้ข้อสอบที่แตกต่างกัน		
ความยากของแบบทดสอบ	ใช้ความยากเฉลี่ยที่เหมาะสมกับผู้สอบ	ความยากจะขึ้นกับผู้สอบแต่ละคน
ความยาวของแบบทดสอบ	ผู้สอบใช้แบบทดสอบที่มีความยาวเท่ากันและมีความยาวมากกว่าแบบทดสอบแบบที่เหมาะสม	ผู้สอบใช้แบบทดสอบที่มีความยาวแตกต่างกันและแบบทดสอบจะสั้นกว่าแบบทดสอบแบบประเพณีนิยม
โอกาสในการสอบ	เจาะจงเวลาในการทดสอบให้กับผู้สอบทุกคน	ผู้สอบเลือกเวลาได้ตามความเหมาะสม
การบริการการสอบ	ใช้เวลาในการทดสอบมาก	ใช้เวลาในการทดสอบน้อย
การให้ผลคะแนน	ไม่สามารถให้ผลคะแนนได้ทันที	สามารถให้ผลคะแนนได้ทันที

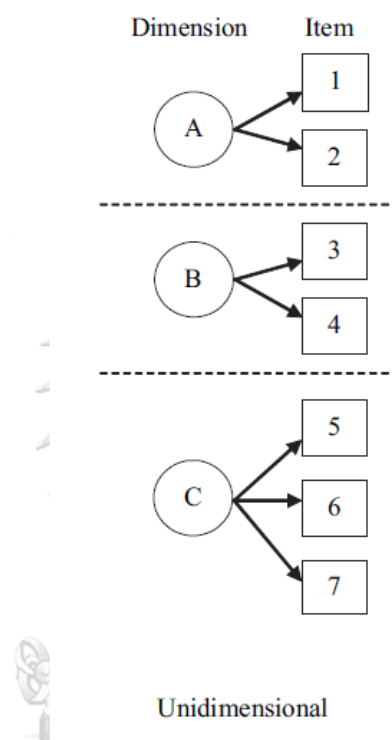
ที่มา : Jian-quan et al. (2007)

### 1.3 การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์

การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์ เป็นการทดสอบที่มีการคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ ซึ่งเป็นการใช้ระบบคอมพิวเตอร์ในการสอบเพื่อประเมินความสามารถของผู้สอบแต่ละคน มีการคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ เริ่มต้นจากข้อสอบข้อที่ยากง่ายปานกลาง และเมื่อทำถูกจะได้ข้อสอบข้อถัดไปที่ยากขึ้นเรื่อยๆ โดยกระบวนการสอบจะดำเนินการไปจนกระทั่งประเมินความสามารถของผู้สอบได้

การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบเอกมิติ (Unidimensional computerized adaptive testing) เป็นการทดสอบที่ผสมผสานกันระหว่างแนวคิดของการ ทดสอบแบบที่เหมาะสมด้วยคอมพิวเตอร์ (Computerized Adaptive Testing) และทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional Item response theory) ซึ่งเป็นการทดสอบที่มีประสิทธิภาพในการวัดสูง มีความแม่นยำ และสามารถลดจำนวนข้อสอบที่ผู้สอบต้องทำการทดสอบได้ ในช่วงเวลา 10 ปีที่ผ่านมาการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์ (CAT) ได้รับความนิยมและกล่าวถึงเป็นอย่างมากถึงประสิทธิภาพและความแม่นยำในการประมาณค่าความสามารถของผู้สอบ อีกทั้งยังช่วยลดเวลาและจำนวนข้อสอบที่ใช้ในการทดสอบ วิธีการทดสอบเป็นมาตรฐานและยืดหยุ่นในด้านการทดสอบด้วย ซึ่งในการประมาณค่าความสามารถของการทดสอบด้วยการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์

(CAT) นั้นจะใช้ทฤษฎีการตอบสนองข้อสอบเป็นพื้นฐานในการคำนวณคะแนนความสามารถของผู้สอบ ภายใต้ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional) (Segall, 1996) ซึ่งเป็นโมเดลการตอบสนองข้อสอบที่ตรวจให้คะแนน 2 ค่าแบบเอกมิติ (Unidimensional IRT Models) ตัวแปรคุณลักษณะแฝงหรือระดับความสามารถของผู้สอบมีเพียงองค์ประกอบเดียว หรือมิติเดียว ดังภาพ 2.1



ภาพ 2.1 โมเดลการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional IRT Model)

(Y.-Y. Cheng, Wang, & Ho, 2008)

การทดสอบปรับเหมาะแบบเหมาะพหุมิติด้วยคอมพิวเตอร์ (Multidimensional computerized adaptive testing) เป็นการทดสอบที่ผสมกันระหว่างแนวคิดของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing) และทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional item response theory) ซึ่งเป็นการทดสอบที่มีประสิทธิภาพในการวัดสูง มีความแม่นยำและสามารถลดจำนวนข้อสอบที่ผู้สอบต้องทำการทดสอบได้ ในเวลาที่ผ่านมาการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) ได้รับความนิยมและกล่าวถึงเป็นอย่างมาก ถึงประสิทธิภาพและความแม่นยำในการประมาณค่าความสามารถของผู้สอบ อีกทั้งยังช่วยลดเวลาและจำนวนข้อสอบที่ใช้ในการทดสอบ วิธีการทดสอบเป็นมาตรฐานและยืดหยุ่นในด้านการทดสอบด้วย ซึ่งในการประมาณค่าความสามารถของการทดสอบด้วย CAT นั้นจะใช้ทฤษฎีการตอบสนองข้อสอบเป็นพื้นฐานในการคำนวณคะแนนความสามารถของผู้สอบ ภายใต้ข้อตกลงเบื้องต้นของทฤษฎี





## 1.4 ประเภทของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์

ประเภทของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์กับความสามารถของผู้สอบ (ทศน์ศิริรินทร์ สว่างบุญ, 2554; ศิริชัย กาญจนวาสี, 2555; Weiss, 1973 อ้างถึงใน สุภารัตน์ หวลมุกดา, 2550; อนุสรณ์ เกิดศรี, 2557) ได้แบ่งวิธีการคัดเลือกข้อสอบที่มีระดับความยากเหมาะสมกับความสามารถของผู้สอบ เป็น 2 วิธี ได้แก่ ยุทธวิธีสองขั้นตอน และยุทธวิธีหลายขั้นตอน รายละเอียดดังนี้

### 1.4.1 ยุทธวิธีสองขั้นตอน (Two-Stage Strategies)

ยุทธวิธีสองขั้นตอนเป็นรูปแบบง่ายที่สุดในยุทธวิธีการทดสอบปรับเหมาะ เป็นยุทธวิธีที่เน้นที่ขั้นตอนการคัดเลือกข้อสอบเพื่อให้ได้จำนวนข้อสอบที่น้อยที่สุดในการประมาณค่าความสามารถของผู้สอบได้อย่างถูกต้อง โดยใช้แบบทดสอบ 2 ฉบับในการทดสอบสองขั้นตอน แบบทดสอบแต่ละฉบับในแต่ละขั้นตอนจะประกอบไปด้วยกลุ่มของข้อสอบที่มุ่งวัดเนื้อหาใดเนื้อหาหนึ่ง ซึ่งกำหนดทิศทางการตอบสำหรับผู้สอบ ประกอบด้วย 2 ขั้นตอน มีรายละเอียดดังนี้

ขั้นตอนแรก หรือ แบบทดสอบเพื่อกำหนดทิศทาง (First-stage test or Routing Test) ประกอบด้วยข้อสอบจำนวนไม่มากนัก ข้อสอบอาจมีระดับความยากปานกลางหรือมีความยากแตกต่างกันจากง่ายไปยาก ผู้สอบต้องทำข้อสอบครบทุกข้อ ผลคะแนนที่ได้นำไปประมาณค่าความสามารถของผู้สอบเริ่มแรกก่อนนำไปใช้ในขั้นตอนที่สองเพื่อเลือกแบบทดสอบให้เหมาะสมกับความสามารถของผู้สอบเป็นรายบุคคล

ขั้นตอนที่สอง หรือแบบทดสอบวัดผล (Second-stage test หรือ Main หรือ Measurement Test) ขั้นตอนนี้เป็นแบบทดสอบที่ถูกคัดเลือกให้เหมาะสมกับความสามารถของผู้สอบ ซึ่งประมาณได้จากการทดสอบในขั้นแรก ในขั้นตอนนี้ประกอบด้วยแบบทดสอบย่อย ประมาณ 3 – 5 ฉบับ แยกตามระดับความยาก ในขั้นตอนนี้แบบทดสอบให้สารสนเทศได้มากกว่าแบบทดสอบแยกทิศทาง ถ้าผู้สอบตอบข้อสอบในขั้นตอนแรกได้ถูกต้องมากข้อจะได้รับแบบทดสอบวัดฉบับความยากสูง แต่ถ้าผู้สอบตอบข้อสอบในขั้นตอนแรกได้ถูกต้องน้อยข้อ จะได้รับแบบทดสอบวัดฉบับความยากต่ำลงมา

ศิริชัย กาญจนวาสี (2555) ได้กล่าวว่า ยุทธวิธีสองขั้นตอน มีข้อดีที่ช่วยลดจำนวนข้อสอบที่แต่ละคนจะต้องทำ โดยไม่ทำให้ความแม่นยำในการวัดลดลง แต่มีข้อจำกัดที่ในการทดสอบกับคนจำนวนมากในเวลาเดียวกัน การตรวจข้อสอบกำหนดทิศทางจะต้องรีบเร่งตรวจและใช้คนจำนวนมาก รวมทั้งการปรับเหมาะของกระบวนการทดสอบเกิดขึ้นเพียงครั้งเดียวหลังจากการทดสอบในขั้นแรก ซึ่งถ้าเกิดความคลาดเคลื่อนในขั้นตอนนี้อาจทำให้เกิดความคลาดเคลื่อนในการทดสอบในขั้นที่สองได้ ความคลาดเคลื่อนในการจำแนกกลุ่มความสามารถของผู้สอบเกิดขึ้นได้ประมาณร้อยละ 20 ตลอดจนแบบทดสอบที่มีช่วงค่าความยากที่แคบจะทำให้ได้สารสนเทศค่อนข้างน้อย แต่ถ้าสร้าง

แบบทดสอบวัดผลในขั้นที่สองให้มีช่วงความยากที่ค่อนข้างกว้างก็จะขัดแย้งกับหลักการของการทดสอบที่เหมาะสม สามารถการแก้ปัญหาดังกล่าวได้โดยใช้คอมพิวเตอร์ในการทดสอบที่เหมาะสม

**1.4.2 ยุทธวิธีหลายขั้นตอน (Multi-Stage Strategies)** ยุทธวิธีนี้แตกต่างจากยุทธวิธีสองขั้นตอน เนื่องจาก ยุทธวิธีสองขั้นตอนตัดสินทางแยกครั้งเดียว จากแบบทดสอบแยกทิศทางไปยังแบบทดสอบวัด แต่ยุทธวิธีหลายขั้นตอนตัดสินทางแยกหลายลำดับขั้นหลังจากผู้สอบตอบคำถามแต่ละข้อ ประกอบด้วย 2 รูปแบบ คือ

1) รูปแบบการทดสอบแบบทางแยกคงที่ (Fixed Branching Model) รูปแบบนี้ใช้ในยุทธวิธีหลายขั้นตอนเหมือนยุทธวิธีสองขั้นตอนในส่วนของการกำหนดทางแยกคงที่ แตกต่างกันในส่วนของ การตัดสินใจแยกทาง ซึ่งในการใช้ยุทธวิธีสองขั้นตอนกำหนดการตัดสินใจเพียงครั้งเดียวจากการทำแบบทดสอบกำหนดทิศทางในขั้นแรกไปการทำแบบทดสอบวัดผลในขั้นที่สอง ส่วนรูปแบบทางแยกคงที่ในการใช้ยุทธวิธีหลายขั้นตอนจะกำหนดการตัดสินใจหลังจากที่ผู้สอบทำข้อสอบในแต่ละข้อตามโครงสร้างของแบบทดสอบที่กำหนดไว้ล่วงหน้า โดยคำนวณค่าพารามิเตอร์ของข้อสอบไว้แล้ว ผู้สอบแต่ละคนจะได้ทำแบบสอบที่มีโครงสร้างของแบบทดสอบเดียวกันแต่จะได้ทำข้อสอบในข้อที่แตกต่างกันตามวิธีการคัดเลือกข้อสอบที่กำหนดไว้ รูปแบบในการกำหนดโครงสร้างของแบบทดสอบสรุปได้ดังนี้

1.1) รูปแบบปิรามิดหรือรูปแบบโครงสร้างต้นไม้ (Pyramidal model) รูปแบบปิรามิดมีหลายลักษณะ ได้แก่ ปิรามิดขนาดขั้นคงที่ ปิรามิดขนาดขั้นแปรผัน ปิรามิดชนิดรอบบิน-มินโร ปิรามิดชนิดข้างตัว ปิรามิดชนิดข้างตัดสัปดาห์การสะท้อนกลับ ปิรามิดชนิดมีหลายข้อในแต่ละขั้น และปิรามิดชนิดให้น้ำหนักตัวเลือก

1.2) รูปแบบปรับระดับแบบยืดหยุ่น (Flexilevel model)

1.3) รูปแบบปรับระดับขั้น (Stradaptive model)

2) รูปแบบทางแยกแปรผัน (Variable Branching Model) รูปแบบนี้เป็นรูปแบบที่ไม่ได้กำหนดโครงสร้างและไม่ได้กำหนดข้อสอบไว้ล่วงหน้าว่าถ้าผู้สอบตอบถูกต้องจะไปทำข้อสอบข้อใดหรือถ้าตอบผิดจะต้องไปทำข้อสอบข้อใด แต่จะดำเนินการโดยใช้ข้อสอบที่กำหนดค่าฟังก์ชันสารสนเทศของข้อสอบไว้แล้ว การคัดเลือกข้อสอบเพื่อให้เหมาะสมกับความสามารถของผู้สอบจะดำเนินการโดยใช้วิธีการทางสถิติในการประมาณค่าความสามารถ ซึ่งวิธีการประมาณค่าความสามารถที่นิยมใช้มี 2 วิธี ได้แก่ การประมาณค่าแบบเบส์ (Bayesian estimation) และการประมาณค่าแบบความเป็นไปได้สูงสุด (maximum likelihood estimation)

3) รูปแบบเฟล็กซิเลเวล รูปแบบนี้ประกอบด้วยชุดของข้อสอบจำนวนหนึ่ง ข้อสอบแต่ละข้อมีช่วงห่างของค่าความยากเท่ากัน โดยทำการเรียงจากข้อที่ง่ายที่สุดไปยังข้อที่ยากที่สุด

การทดสอบเริ่มต้นด้วยการให้ผู้ตอบทำข้อสอบข้อที่มีความยากปานกลาง ถ้าตอบถูกก็จะไปตอบข้อที่ยากขึ้น แต่ถ้าตอบผิดข้อต่อไปก็จะง่ายลง

4) รูปแบบปรับระดับขั้น (Stradaptive Model) การทดสอบรูปแบบนี้เริ่มจากการเตรียมคลัง ข้อสอบโดย แบ่งแบบทดสอบเป็นกลุ่มข้อสอบย่อยตามค่าความยากและจัดเข้าแต่ละชั้น (Stratum) ของคลังข้อสอบ กลุ่มข้อสอบย่อยในแต่ละชั้นกระจายอยู่รอบๆ ระดับค่าความยากแสดงเป็นยอดโค้ง (Peak) ส่วนสูงของยอดโค้งแสดงถึงความหนาแน่นของจำนวนข้อสอบในระดับความยากตามที่กำหนด ค่าความยากเฉลี่ยของข้อสอบชั้นถัดไปจะสูงกว่าค่าความยากเฉลี่ยของข้อสอบชั้นก่อน นั่นคือ ชั้นที่ 1 บรรจุด้วยกลุ่มข้อสอบมีค่าความยากน้อยที่สุด ชั้นที่ 2 บรรจุด้วยกลุ่มข้อสอบมีค่าความยาก เฉลี่ยสูงกว่าชั้นที่ 1 ชั้นอื่นๆ บรรจุด้วยกลุ่มข้อสอบมีค่าความยากเฉลี่ยสูงขึ้นไปเรื่อยๆ จนถึงขั้นสูงสุด การทดสอบปรับเหมาะรูปแบบปรับระดับขั้นเริ่มด้วย การประมาณค่าระดับความสามารถของผู้สอบ พิจารณาจากสารสนเทศเบื้องต้นหรือข้อมูลส่วนตัวนำไปกำหนดจุดนำเข้า (Entry Point) ผู้สอบความสามารถต่ำจะเริ่มต้นการทดสอบด้วยการทำข้อสอบง่าย และผู้สอบความสามารถสูงจะเริ่มต้นการทดสอบด้วยข้อสอบยาก ผลการตอบข้อสอบข้อแรกนำไปประมาณค่าความสามารถและกำหนดทางแยกตามกฎที่เลือกใช้ กรณีกฎทางแยกขึ้นหนึ่ง-ลงหนึ่งเมื่อตอบข้อสอบถูกต้อง จะนำไปสู่ข้อสอบในขั้นต่อไปที่มีระดับ ความยากสูงสุดในขั้นนั้น ถ้าตอบข้อสอบผิด ข้อสอบข้อต่อไปจะอยู่ในขั้นต่ำลงมาที่ให้ค่าความยากโดยเฉลี่ย ดังนั้น การออกแบบทางแยกในรูปแบบปรับระดับขั้นเป็นไปเพื่อหาค่าความยากของข้อสอบให้เหมาะสมกับผู้สอบ ชั้นคลังข้อสอบที่ผู้สอบตอบข้อสอบทั้งหมดถูกต้องอาจเป็นขั้นพื้นฐาน (Basal Stratum) และชั้นคลังข้อสอบที่ผู้สอบตอบข้อสอบผิดทั้งหมดอาจเป็นขั้นเพดาน (Ceiling Stratum) ความสามารถจริงของผู้สอบตกอยู่ระหว่าง 2 ขั้นนี้ โดยถือว่า ผู้สอบสามารถตอบข้อสอบได้ถูกต้องประมาณ 50 เปอร์เซนต์ของข้อสอบที่ทำ

ไวส์ (Weiss, 1973 อ้างถึงใน สุภารัตน์ หวลมุกดา, 2550) อธิบายข้อดีของการทดสอบแบบปรับระดับขั้น (Stradaptive Testing) ว่าเหนือกว่าการทดสอบในกลยุทธ์สองขั้นตอนและกลยุทธ์หลายขั้นตอนในรูปแบบอื่นๆ ดังนี้

การทดสอบปรับเหมาะรูปแบบปรับระดับขั้นสามารถย้อนกลับตำแหน่งการทดสอบเดิมได้ (Complete Recovery) การย้อนกลับนี้เป็นไปไม่ได้ในรูปแบบปิรามิดและเฟล็กซิเลเวล สำหรับกลยุทธ์สองขั้นตอนพบว่ามักเกิดความคลาดเคลื่อนบ่อยๆ คลังข้อสอบรูปแบบนี้สามารถให้สารสนเทศสูงสุดแก่ผู้สอบ ผู้สอบส่วนมากได้รับสัดส่วนการตอบถูกต้องทั้งหมดประมาณ .50 ของการทดสอบ และมีกฎการยุติถูกออกแบบไว้ชัดเจนสำหรับอธิบายการเดา ผู้สอบทั้งหมดไม่จำเป็นต้องเริ่มทดสอบด้วยข้อสอบเดียวกัน และการทดสอบโดยใช้สารสนเทศเดิมของผู้สอบสามารถลดจำนวนการใช้ข้อสอบในผู้สอบแต่ละคนได้ วิธีนี้ยอมให้จำนวนข้อสอบเปลี่ยนแปลงตามผู้สอบ การทดสอบดำเนินอย่างต่อเนื่องจนได้ระดับความแม่นยำตามต้องการ ถ้าคลังข้อสอบมีจำนวนข้อสอบเพียงพอ

การทดสอบจะให้ความแม่นยำเท่าเทียมกันทุกระดับความสามารถของผู้สอบ ผลการทดสอบปรับเหมาะสมรูปแบบปรับระดับขั้นให้คะแนนคงเส้นคงวา เมื่อคะแนนมีความแน่นอนสามารถใช้ในการทำนายครั้งต่อไปได้ถูกต้อง การทดสอบปรับเหมาะสมรูปแบบปรับระดับขั้นออกแบบมาเพื่อใช้กับเครื่องคอมพิวเตอร์ แต่ยังสามารถใช้เป็นเครื่องมือการทดสอบเฉพาะตามจุดประสงค์ได้

การพัฒนาคลังข้อสอบสำหรับการทดสอบรูปแบบปรับระดับขั้น (Weiss, 1973 อ้างถึงใน สุภารัตน์ หวลมุกดา, 2550) มีแนวทางว่าข้อสอบต้องมีจำนวนมากเพียงพอและครอบคลุมระดับความสามารถของผู้สอบ ขนาดคลังข้อสอบขึ้นกับคุณภาพข้อสอบ และขนาดคลังข้อสอบเมื่อแบ่งชั้นแล้วจะได้ผลดีเมื่อมีข้อสอบมากกว่า 200 ข้อขึ้นไป ข้อสอบจะนำไปหาค่าความยากและค่าอำนาจจำแนกของข้อสอบจากกลุ่มผู้สอบขนาดใหญ่ข้อสอบอำนาจจำแนกต่ำจะขจัดออกไป ต่อจากนั้นจัดคลังข้อสอบโดยแบ่งชั้นตามระดับค่าความยากในแต่ละชั้นจะได้ยอดโค้งของข้อสอบโดยไม่มีการทับซ้อนค่าความยากของข้อสอบระหว่างชั้น จำนวนชั้นคลังข้อสอบหรือจำนวนยอดโค้งข้อสอบขึ้นอยู่กับขนาดคลังข้อสอบครั้งแรก จำนวนชั้นยิ่งมากจะให้ความแม่นยำสำหรับกลุ่มผู้สอบระดับความสามารถต่างๆ อย่างเท่าเทียมกัน จำนวนชั้นควรมีประมาณ 9-10 ชั้น ครอบคลุมระดับความสามารถโดยไม่จำเป็นต้องใช้คลังข้อสอบขนาดใหญ่ทุกชั้นขั้นบรรจุข้อสอบ 10-15 ข้อ และจำนวนข้อสอบไม่จำเป็นต้องเท่ากันทุกชั้น ชั้นความยาก ต่ำ ปานกลาง ต้องการข้อสอบมากกว่าชั้นความยากสูง ข้อสอบในแต่ละชั้นถูกจัดลำดับค่าอำนาจจำแนกข้อสอบจากมากไปน้อย ระยะแรกของการทดสอบเมื่อใช้ข้อสอบจำนวนน้อย การจัดค่าอำนาจจำแนก ข้อสอบเป็นช่วงกว้างจะเหมาะสมกว่า ในทางตรงข้าม ระยะท้ายของการทดสอบ เมื่อค่าประมาณความสามารถของผู้สอบใกล้เคียงความจริง ควรจัดค่าอำนาจจำแนกข้อสอบเป็นช่วงแคบ การทดสอบจึงจะได้ประสิทธิภาพดี

จุฬาลงกรณ์มหาวิทยาลัย

## 1.5 องค์ประกอบของระบบการทดสอบปรับเหมาะสมด้วยคอมพิวเตอร์

การทดสอบปรับเหมาะสมด้วยคอมพิวเตอร์มีองค์ประกอบของการดำเนินงานที่สำคัญ โดยมีรายละเอียดดังนี้

### 1.5.1 คลังข้อสอบ (Item Pool)

คลังข้อสอบ (Item Pool หรือ Question Banks, Item Collection, Item Reservoirs, Test Item Libraries) (Flaugh, 2000) ความตรงของแบบทดสอบขึ้นอยู่กับความเกี่ยวข้องและความเหมาะสมของข้อสอบที่ใช้แต่สำหรับการทดสอบปรับเหมาะสมด้วยคอมพิวเตอร์ความตรงของแบบทดสอบจะขึ้นอยู่กับทั้งจำนวนและเนื้อหาของข้อสอบในคลังข้อสอบ นอกจากข้อสอบจะต้องมีจำนวนมากเพียงพอ ในการประมาณค่าระดับความสามารถของผู้สอบซึ่งมีความแตกต่างหลากหลายกันได้อย่างถูกต้อง เหมาะสมแล้ว ข้อควรคำนึงที่สำคัญประการหนึ่งก็คือเนื้อหาของข้อสอบ เนื่องจากการมีข้อสอบในคลังจำนวนมากทำให้ มีโอกาสที่จะมีข้อสอบที่ไม่ดีได้มาก ซึ่งข้อสอบเหล่านั้นจะมี

ผลกระทบอย่างมากต่อการให้คะแนนสุดท้ายในการทดสอบปรับเหมาะ เพราะในการทดสอบแบบปรับเหมาะมีการตรวจให้คะแนนในทันที ข้อสอบที่ถูกเลือกให้ผู้สอบทำจึงมีผลต่อความถูกต้องในการประมาณค่าความสามารถของผู้สอบและการให้คะแนนสุดท้ายโดยไม่มีการแก้ไขหรือย้อนกลับได้ (Hambleton, Zaal and Pieters 2000 อ้างถึงใน Latu & Chapman, 2002) การทดสอบปรับเหมาะจะมีประสิทธิภาพสูงขึ้น ถ้ามีคลังข้อสอบขนาดใหญ่ ข้อสอบดังกล่าวได้รับการประมาณค่าพารามิเตอร์โดยใช้ ทฤษฎีการตอบสนองข้อสอบ ข้อสอบมีค่าอำนาจจำแนกสูง มีค่าความยากครอบคลุมช่วงความสามารถของผู้สอบ และมีค่าการเดาต่ำ Weiss (1988) อ้างถึงใน ศิริชัย กาญจนวาสี (2555) ได้เสนอไว้ว่าจำนวนข้อสอบในคลังข้อสอบควรมีน้อย 100-200 ข้อ Allen and Yen (1979) อ้างถึงใน ศิริชัย กาญจนวาสี (2555) กล่าวว่า ถ้าจะให้ดีต้องมีจำนวนข้อสอบจำนวน  $2n$  ข้อ เมื่อ  $n$  = จำนวนข้อสอบที่ใช้ในการทดสอบ แต่ถ้าการทดสอบมีการกำหนดจำนวนข้อที่จะใช้คงที่เหมือนกันทุกคน จำนวนข้อสอบควรมีน้อย  $n(n+1)/2$  โดยข้อสอบแต่ละข้อควรมีอำนาจจำแนกสูง ( $a > 0.80$ ) มีค่าความยากที่ครอบคลุม ระดับช่วงต่างๆ อย่างเหมาะสม  $[-2.0 < \theta < +2.0]$  ค่าสัมประสิทธิ์การเดา ( $c < 0.30$ ) และควรมีข้อสอบเท่าๆกันในแต่ละระดับความยาก (ศิริชัย กาญจนวาสี, 2555)

### 1.5.2 จุดเริ่มต้นการทดสอบ (starting point or entry level)

การทดสอบเริ่มต้นด้วยการคัดเลือกข้อสอบข้อแรก (initial item) การคัดเลือกข้อสอบข้อแรกนิยมใช้ข้อสอบที่มีความยากปานกลางซึ่ง (ทัศนศิริินทร์ สว่างบุญ, 2554; ศิริชัย กาญจนวาสี, 2555) สามารถแบ่งวิธีการคัดเลือกเป็น 2 กรณี

1) กรณีเมื่อประชากรผู้สอบมีความสามารถค่อนข้างใกล้เคียงกัน (homogeneous) หรือไม่มีข้อมูลสำหรับผลสัมฤทธิ์ที่ผ่านมาควรเริ่มต้นด้วยข้อสอบที่มีค่าความยากปานกลางเท่ากันสำหรับทุกคน

2) กรณีเมื่อประชากรผู้สอบมีความสามารถค่อนข้างแตกต่างกัน (heterogeneous) โดยพิจารณาจากข้อมูล ที่มีอยู่ เช่นกลุ่มอายุ ระดับชั้นที่ศึกษา ซึ่งควรเริ่มต้นด้วยข้อสอบที่มีค่าความยากปานกลางสำหรับแต่ละระดับการศึกษาของผู้สอบ

ระดับในการเริ่มต้นในการทดสอบปรับเหมาะ ผู้สอบแต่ละคนจะได้ข้อสอบที่มีระดับความยากต่างกันการเริ่มทำข้อสอบที่คลาดเคลื่อนไปจากความสามารถของผู้สอบไม่มีผลกระทบต่อการประมาณค่าความสามารถเท่าใด แต่ถ้าให้ผู้สอบเริ่มทำข้อสอบที่ตรงกับความสามารถของตนก็จะทำให้ลดจำนวนข้อที่จะใช้ในการทดสอบลงได้ (พิมพ์สิริ เจริญนเรศรัษฎ์, 2549) สอดคล้องกับ ศิริชัย กาญจนวาสี (2555) การทดสอบควรเริ่มต้นด้วยข้อสอบที่ตรงกับความสามารถของผู้สอบหรือใกล้เคียงกับความสามารถของผู้สอบมากที่สุด ซึ่งจะช่วยให้จำนวนข้อสอบที่ใช้ในการทดสอบสำหรับแต่ละบุคคลมีจำนวนลดลง และโดยทั่วไปนิยมเลือกข้อสอบที่มีความยากปานกลาง

การกำหนดจุดเริ่มต้นการทดสอบควรใช้ข้อสอบค่าความยากปานกลางจัดให้กับผู้สอบแต่ละคนเมื่อไม่ทราบค่าสารสนเทศเบื้องต้นของผู้สอบ การใช้สารสนเทศเบื้องต้นเกี่ยวกับระดับความสามารถของผู้สอบหรือข้อมูลการศึกษาที่ผ่านมาใช้กำหนดจุดเริ่มต้นการทดสอบมีความเหมาะสมแตกต่างกัน การนำไปใช้ขึ้นอยู่กับการออกแบบและเงื่อนไขการทดสอบแบบปรับเหมาะตามที่กำหนด

### 1.5.3 การคัดเลือกข้อสอบ (item selection)

การคัดเลือกข้อสอบและการประมาณค่าความสามารถของผู้สอบมีประสิทธิภาพมากขึ้นเมื่อนำทฤษฎีการตอบสนองข้อสอบ (item response theory) มาประยุกต์ใช้ โดยทั่วไปนิยมใช้วิธีที่สอดคล้องกัน เพื่อความสะดวกรวดเร็วในการคำนวณ (Hulin, Drasgow and Parson, 1983 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555)

1) กรณีที่ใช้การประมาณค่าความสามารถด้วยวิธีของ maximum likelihood

การประมาณค่าความสามารถด้วยวิธีของ maximum likelihood estimation นิยมใช้กับวิธีการคัดเลือกข้อสอบวิธีใดวิธีหนึ่งดังต่อไปนี้

1.1) คัดเลือกข้อสอบที่มีความยาก ( $b_i$ ) สอดคล้องกับระดับความสามารถที่ประมาณค่าได้ ( $\theta$ ) [Match  $b_i$  to  $\theta$ ] โดยการตรวจสอบข้อสอบทุกข้อเพื่อเลือกข้อสอบข้อที่มีค่า  $b_i$  ใกล้เคียงกับ  $\theta$  ที่เพิ่งประมาณได้ และใช้ข้อสอบข้อนั้นเป็นข้อถัดไป วิธีนี้ใช้กันมาตั้งแต่ตอนต้นยุคของ Tailored Testing เนื่องจากคำนวณง่ายและประหยัด

1.2) คัดเลือกข้อสอบที่มีตำแหน่งสารสนเทศสูงสุด ( $m_i$ ) สอดคล้องกับระดับความสามารถที่ประมาณได้ ( $\theta$ ) [Match  $m_i$  to  $\theta$ ] ข้อสอบมักมีปัจจัย การเดาเข้ามาเกี่ยวข้อง การคัดเลือกข้อสอบที่มีค่า  $b_i$  ใกล้เคียงที่สุดกับ  $\theta$  จึงมักมีแนวโน้มที่จะได้ข้อที่ยากเกินไปสำหรับผู้สอบ เนื่องจากข้อสอบแต่ละข้อให้สารสนเทศสูงสุดที่ตำแหน่ง  $\theta = m_i$  เมื่อ  $m_i$  สำหรับโมเดล 3 พารามิเตอร์ คำนวณได้ดังนี้

$$m_i = b_i + \frac{1}{Da_1} \ln \left[ \frac{1 + \sqrt{1 + 8c_i}}{2} \right]$$

เมื่อ	D	=	ค่าคงที่
	ln	=	natural logarithm
	$a_i$	=	ค่าอำนาจจำแนกของข้อสอบ
	$b_i$	=	ค่าความยากของข้อสอบ
	$c_i$	=	ค่าโอกาสการเดาข้อสอบถูก

$m_i$  เป็นค่า  $\theta$  ณ ตำแหน่งที่ข้อสอบข้อนั้นให้สารสนเทศสูงสุด ซึ่ง  $m_i > b_i$  ถ้า  $c_i > 0$  แสดงว่าข้อสอบที่มีค่าความยาก  $b_i$  เหมาะที่จะใช้ประมาณค่า  $\theta$  ของผู้สอบที่มีความสามารถสูงกว่า  $b_i$  เล็กน้อย ดังนั้นการคัดเลือกข้อสอบจากค่า  $b_i$  จึง สามารถปรับปรุงให้ดีขึ้น เมื่อพิจารณาจากค่า  $m_i$

1.3) คัดเลือกข้อสอบที่ให้สารสนเทศสูงสุดตรงตำแหน่ง  $\hat{\theta}$  (maximum Information Item information) โดยการตรวจสอบข้อสอบทุกข้อที่ยังไม่ได้นำมาใช้สอบ ที่สามารถให้สารสนเทศสูงสุด ณ ตำแหน่ง  $\hat{\theta}$  การคัดเลือกข้อสอบตามวิธีนี้จึง ต้องคำนวณสารสนเทศของข้อสอบ (item information) ตรงตำแหน่ง  $\hat{\theta}$  [ $I(\hat{\theta}, u_i)$ ]

1.4) วิธีสารสนเทศสูงสุดของฟิชเชอร์ (Maximum Fisher Information: MFI) เป็นวิธีที่ใช้กันอย่างแพร่หลายในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ ที่มีประสิทธิภาพในการประมาณค่าความสามารถ การทดสอบข้อสอบก่อนหน้านี้ในการบริหารจัดการทดสอบของผู้เข้าสอบ โดยทั่วไปเป็นการค้นหาข้อสอบ  $x$  ทำให้สูงสุด  $I[\hat{\theta}_{m-1}]$  สำหรับผู้เข้าสอบ ในการประมาณค่าความสามารถ  $\theta$  และ  $m - 1$  ของจำนวนข้อสอบในการบริหารการทดสอบต่อไป ในคลังข้อสอบแบบตัวเลือกมีการกำหนดรูปแบบโมเดลโลจิสติกส์ 3 พารามิเตอร์ (3PL) วิธีการคัดเลือกข้อสอบตามเกณฑ์ MFI จะคัดเลือกข้อสอบสำหรับ ข้อสอบ  $i$  ที่มีสารสนเทศมากที่สุด

$$I_i[\theta_{m-1}] = \frac{(Da_i)^2 (1-c_i)}{[c_i + e^{Da_i(\theta_{m-1} - b_i)}][1 + e^{-Da_i(\theta_{m-1} - b_i)}]^2}$$

เมื่อ  $a_i$  เป็น พารามิเตอร์ค่าอำนาจจำแนกข้อสอบ

$b_i$  เป็น พารามิเตอร์ค่าความยากของข้อสอบ

$c_i$  เป็น พารามิเตอร์การเดาข้อสอบได้ถูก

D เป็น ค่าคงที่ เท่ากับ 1.702

เกณฑ์สารสนเทศสูงสุดของฟิชเชอร์ เป็นที่นิยมเพราะเป็นวิธีที่มีประสิทธิภาพในการบริหารการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) ส่งผลให้สารสนเทศการทดสอบสูงสุดสำหรับผู้เข้าสอบแต่ละคน

ปัจจุบันการเลือกข้อสอบวิธีสารสนเทศสูงสุดประยุกต์ใช้กับการเลือกข้อสอบวิธีอื่นๆ เพื่อเพิ่มประสิทธิภาพในการเลือกข้อสอบ ได้แก่ การเลือกข้อสอบโดยใช้ค่าสารสนเทศโดยรวมเฉลี่ย (Kullback -Leibler Information : KLI) การเลือกข้อสอบด้วยวิธีเกณฑ์ข้างเคียงใกล้ที่สุด (Nearest-Neighbors Criterion ) วิธีเกณฑ์ข้างเคียงใกล้ที่สุดเป็นการรวมวิธีสารสนเทศสูงสุดของฟิชเชอร์และวิธีค่าความยากข้อสอบเหมาะสม (Optimal Item Difficulty) เข้าไว้ด้วยกัน นอกจากนี้ การศึกษาวิธีเลือกข้อสอบจากสารสนเทศเฉลี่ยสูงสุด นำไปใช้ผ่อนคลายค่าประมาณความสามารถระหว่างการเลือกข้อสอบ เช่น เกณฑ์สารสนเทศตามน้ำหนักทั่วไป (General Weight Information

Criterion) เกณฑ์สารสนเทศช่วงชั้น (Interval Information Criterion) และ เกณฑ์สารสนเทศตามน้ำหนักเป็นไปได้ (Likelihood Weight Information Criterion)

วิธีการคัดเลือกข้อสอบที่ใช้กันมากที่สุดควบคู่ไปกับการประมาณค่า  $\theta$  ด้วยวิธี Bayesian คือ การเลือกข้อสอบข้อที่ยังไม่ได้นำมาใช้ในการทดสอบนั้น ที่จะให้ค่าความแปรปรวนของค่าความสามารถที่คาดหวังมีค่าต่ำสุด (smallest posterior variance) นั่น คือ เป็นการเลือกข้อที่คาดว่า จะลดความไม่แน่นอนลงได้มากที่สุด ในการประมาณค่า  $\theta$

#### 1.5) วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI)

Y. Cheng and Chang (2009) เสนอวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index) สำหรับการคัดเลือกข้อสอบที่มีข้อบังคับอย่างเข้มงวด การวิจัยครั้งนี้การศึกษาผ่านการจำลอง ข้อมูลเพื่อเปรียบเทียบวิธีการจัดการข้อบังคับ 2 วิธี ได้แก่ 1) วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) และ 2) วิธี Weighted Deviation Modeling (WMD) โดยใช้วิธีการคัดเลือกข้อสอบแบบ Maximum Information และวิธี Randomized เป็นฐานในการเปรียบเทียบ โดยใช้ข้อมูลจากคลังข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ของการสอบเพื่อจัดตำแหน่งในมหาวิทยาลัย (College Placement CAT) ในวิชาพื้นฐานพีชคณิต คลังข้อสอบประกอบด้วยข้อสอบจาก 3 กลุ่มเนื้อหาวิชา แต่ละกลุ่มเนื้อหาวิชาแบ่งออกเป็น 4-10 ขอบเขตเนื้อหา ผลการศึกษาพบว่า 1) วิธี MPI และวิธี Weighted Deviation Modeling (WMD) ลดการฝ่าฝืนข้อบังคับในการเปรียบเทียบกับวิธี MPI และ Randomized และผลการทดสอบแสดงให้เห็นอย่างชัดเจนว่า MPI มีประสิทธิภาพดีกว่า WMD ทุกด้าน โดยมีอัตราการใช้ข้อสอบซ้ำสูงสุด การใช้ข้อสอบซ้ำเกิน และอัตราการทับซ้อนของแบบสอบต่ำกว่าวิธี WMD นอกจากนี้ยังพบว่าทุกวิธียกเว้น วิธี Randomized มีสัดส่วนของ Never Exposed สูงมากกว่า 50% แสดงว่าข้อสอบในคลังข้อสอบ ถูกใช้อย่างไม่มีประสิทธิภาพ (Chang & Ying, 1999) จากข้อค้นพบของการศึกษาครั้งนี้ วิธี MPI นี้สามารถปรับให้เข้ากับข้อบังคับที่ไม่ใช่ทางสถิติที่หลากหลายและสามารถควบคุมได้หลายอย่างพร้อมๆ กัน เช่น ความสมดุลของเนื้อหา การควบคุมการใช้ข้อสอบซ้ำ ความสมดุลของคำตอบ โดยที่วิธีการนี้สามารถนำไปปรับใช้กับโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ที่มีอยู่ในปัจจุบันได้ สะดวกโดยไม่จำเป็นต้องปรับค่าน้ำหนักความสัมพันธ์ระหว่างข้อบังคับและสารสนเทศ

วิธีดัชนีลำดับความสำคัญสูงสุด จะพิจารณาความแปรปรวนของค่าสารสนเทศโดยการนำดัชนีไปเป็นตัวคูณให้กับค่าสารสนเทศของข้อสอบ ดังนั้นในการคัดเลือกข้อสอบแทนที่จะพิจารณาจากค่าสารสนเทศเพียงอย่างเดียววิธีการนี้จะนำลำดับความสำคัญซึ่งกำหนดโดยผู้เชี่ยวชาญที่พิจารณาร่วมด้วย การคัดเลือกข้อสอบจะพิจารณาจากผลคูณของค่าสารสนเทศของข้อสอบกับดัชนีความสำคัญ ข้อใดให้ผลคูณมากที่สุด ข้อนั้นจะได้รับการคัดเลือกให้นำไปใช้กับผู้สอบ ดังนั้น วิธี MPI



จึง เกี่ยวข้องกับเมตริกซ์เงื่อนไขบังคับ ซึ่งจะบอกว่าข้อสอบแต่ละข้อประกอบด้วยเงื่อนไขบังคับใดบ้าง โดยกำหนดให้เมตริกซ์ดังกล่าว คือ เมตริกซ์ C.C เมื่อ  $C_{jk} = 1$  แสดงว่า ข้อสอบข้อที่ j เกี่ยวข้องกับเงื่อนไขบังคับที่ K

เมื่อ	C.C	คือ	เมตริกซ์ที่เกี่ยวข้องกับข้อบังคับ และมีขนาด $J \times K$
	j	คือ	จำนวนข้อสอบในคลังข้อสอบ
	K	คือ	จำนวนเงื่อนไขบังคับทั้งหมด
	k	คือ	ข้อบังคับเนื้อหาจะเกี่ยวข้องกับค่าน้ำหนัก (weight) $w_k$
	$C_j$	คือ	1 บ่งชี้ว่า ข้อบังคับ k เกี่ยวข้องกับข้อสอบที่ j
	$C_j$	คือ	0 บ่งชี้ว่า ข้อบังคับ k ไม่เกี่ยวข้องกับข้อสอบที่ j

โดยทั่วไป Matrix C.C จะถูกระบุก่อนการเลือกข้อสอบโดยผู้เชี่ยวชาญด้านเนื้อหาหรือผู้เชี่ยวชาญทางด้านจิตมิติ โดยที่แต่ละข้อบังคับ k จะเกี่ยวข้องกับค่าน้ำหนัก (weight)  $w_k$  ซึ่งการทดสอบส่วนใหญ่จะใส่ค่าน้ำหนักขนาดใหญ่ให้กับข้อบังคับที่สำคัญและค่าน้ำหนักที่น้อยกว่ากับข้อบังคับอื่นๆ ดังนั้น Priority Index ของข้อสอบข้อที่ j สามารถคำนวณได้ดังนี้

$$PI_j = I_j \prod_{k=1}^K (w_k f_k)^{C_{jk}} \dots \dots \dots (1)$$

เมื่อ PI แทน Fisher information ของข้อสอบข้อที่ J ที่ถูกประมาณค่าที่ระดับความสามารถปัจจุบัน

$X_j$  แทน ข้อสอบจากขอบเขตเนื้อหาที่แน่นอน

$X_k$  แทน ข้อสอบที่จะถูกเลือก

$f_{jk}$  แทน โควตาที่มีอยู่ของข้อบังคับ k

และผลของโควตาที่เหลืออยู่จะคำนวณได้จาก

$$f_k = \frac{(X_k - x_k)}{X_k} \dots \dots \dots (2)$$

เมื่อ  $C_{jk} = 0$  หมายความว่า ข้อสอบข้อที่ j ไม่ถูกควบคุมด้วยข้อบังคับ k

สมมติว่า ต้องการข้อบังคับ k' ดังนั้นอัตราการแสดงของแต่ละข้อเป็นค่าที่ต่ำกว่าหรือเท่ากับ r และระหว่างผู้สอบทั้งหมด N คน ผู้ซึ่งทำการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ มีผู้สอบ n คน ที่ได้เห็นข้อสอบ j แล้ว  $f_{jk'}$  สามารถคำนวณได้จาก

$$f_{jk'} = \frac{(r - (n/N))}{r} \dots \dots \dots (3)$$

จากสมการที่ 3 สามารถคำนวณดัชนีลำดับความสำคัญ (Maximum Priority Index: MPI) สำหรับข้อสอบทุกข้อที่อยู่ในคลังข้อสอบ วิธี MPI ในรูปแบบปัจจุบันถูกจำกัดข้อบังคับในรูปของขอบเขตบน อย่างไรก็ตาม โปรแกรม การทดสอบมักจะมีข้อบังคับที่เกี่ยวข้องกับขอบเขตล่าง ดังตัวอย่าง ผังข้อสอบสำหรับแบบสอบ คณิตศาสตร์อาจจะระบุไว้เพียงไม่เกิน 15 ข้อ (upper bound) แต่ก็ไม่น้อยกว่า 10 ข้อ (lower bound) ที่เป็นเนื้อหาทางด้านพีชคณิต ดังนั้นข้อบังคับจะเรียกว่า ข้อบังคับความสมดุลทางเนื้อหาที่ยืดหยุ่น (Flexible content balancing constraint) (Y. Cheng, Chang, & Yi, 2007; อนุสรณ์ เกิดศรี, 2557)

#### 1.5.4 การประมาณค่าความสามารถ (ability estimate)

Weiss (1974) อ้างถึงใน สุดารัตน์ หวลมุกดา (2550) และ ทศน์ศิริพันธ์ สว่างบุญ (2554) ได้กล่าวถึงวิธีการให้คะแนนการทดสอบที่เหมาะสมหรือการประมาณค่าความสามารถไว้หลายวิธี สรุปได้ดังนี้

1) ให้คะแนนตามค่าความยากของข้อสอบที่ยากที่สุดที่ตอบถูก (the most difficult of items answered correctly) เป็นการให้คะแนนโดยดูจากข้อสอบทั้งหมดที่ผู้สอบสามารถตอบได้ถูกต้อง หากข้อใดที่มีค่าความยากสูงสุด ที่ผู้สอบตอบถูกถือว่าเป็นค่าความสามารถของผู้สอบคนนั้น ค่าความยากดังกล่าวควรเป็นค่าความยากที่ได้จากการวิเคราะห์ข้อทดสอบตามทฤษฎีการตอบสนองข้อทดสอบ

2) ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก (average difficulty of all items answered correctly) เป็นการให้คะแนนโดยการนำค่าความยากของข้อสอบทุกข้อที่ผู้สอบตอบถูกมาหาค่าเฉลี่ย ค่าเฉลี่ยความยากนี้ถือเป็นค่าความสามารถของผู้สอบคนนั้น

3) ให้คะแนนตามค่าเฉลี่ยความยากของทุกข้อที่ทำ (average difficulty) วิธีการให้คะแนนแบบนี้ Lord เป็นผู้เสนอ ในปี ค.ศ.1970 โดยการนำค่าความยากของข้อสอบทุกข้อที่ผู้สอบได้ตอบโดยไม่คำนึงว่าจะตอบถูก หรือไม่มาหาค่าเฉลี่ย ค่าเฉลี่ยนี้ถือเป็นค่าความสามารถของผู้สอบ

4) ให้คะแนนตามค่าความยากของข้อสุดท้ายที่ตอบ (difficulty of the final item) นั้นคือไม่คำนึงว่าขั้นสุดท้ายจะตอบถูก หรือไม่ก็ตาม จะถือว่าค่าความยากของข้อสอบ ข้อสุดท้ายที่ผู้สอบตอบคือค่าความสามารถของผู้สอบผู้นั้น

5) ให้คะแนนตามค่าความยากของข้อสอบที่ต่อจากขั้นสุดท้ายที่ตอบ (difficulty of  $(N+1)^{th}$  item) ซึ่งมีลักษณะคล้ายกับวิธีที่ 4 แต่เป็นการตรวจสอบความสามารถต่อไปอีกขั้นหนึ่ง กล่าวคือ ถ้าขั้นสุดท้ายที่ผู้สอบสามารถตอบถูก คะแนนในขั้นต่อไปก็จะเพิ่มขึ้น แต่ถ้าขั้นสุดท้ายที่สอบผิดคะแนนขั้นต่อไปก็จะลดลง

6) การให้คะแนนตามวิธีของเบส์ มีหลายวิธี แต่วิธีที่นิยมใช้กันมาก คือ วิธีของเบส์ที่ปรับปรุงให้ทันสมัย (Bayesian updating) ซึ่งสามารถประมาณค่าความสามารถของผู้สอบได้ค่อนข้างคงที่ ซึ่งมีสูตรดังนี้

กรณีที่ตอบถูก

$$\theta_{m+1} = \theta_m + (1 - c) \left( \frac{\sigma_m^2}{\sqrt{\frac{1}{a^2} + \sigma_m^2}} \right) \left( \frac{O(D)}{C + (1-c)A(-D)} \right)$$

$$\sigma_{m+1}^2 = \sigma_m^2 \left( 1 - \left( \frac{1-c}{1 + \frac{1}{a^2 \sigma_m^2}} \right) \left( \frac{O(D)}{B} \right) \left( \frac{(1-c)O(D)}{B} - D \right) \right)$$

เมื่อ  $D = \frac{b - \theta_m}{\sqrt{\frac{1}{a^2} + \sigma_m^2}}$

และ  $B = c + (1+c) \times A(-D)$

กรณีที่ตอบผิด

$$\theta_{m+1} = \theta_m - \left( \frac{\sigma_m^2}{\sqrt{\frac{1}{a^2} + \sigma_m^2}} \right) \left( \frac{O(D)}{A(D)} \right)$$

$$\sigma_{m+1}^2 = \sigma_m^2 \left( 1 - \left( \frac{O(D)}{1 + \frac{1}{a^2 \sigma_m^2}} \right) \left( \frac{O(D)}{A(D)} + D \right) \div A(D) \right)$$

$\theta_m$  แทน ความสามารถของผู้สอบที่ประมาณค่าได้ก่อนตอบข้อสอบข้อที่ m+1 ซึ่งตามปกติแล้ว ถ้าไม่ทราบค่าความสามารถเบื้องต้นของผู้สอบมักกำหนดให้เท่ากับ 0

$\sigma_m^2$  แทน ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบก่อนตอบข้อที่ m+1 ถ้าไม่ทราบค่าความแปรปรวนดังกล่าวมาก่อนมักกำหนดให้เท่ากับ 1

$\theta_{m+1}$  แทน ค่าความสามารถของผู้สอบโดยประมาณหลังจากที่ตอบข้อที่ m+1

$\sigma_{m+1}^2$  แทน ค่าความแปรปรวนในการประมาณค่าความสามารถของผู้สอบเมื่อตอบข้อสอบข้อที่ m+1

a แทน พารามิเตอร์ค่าอำนาจจำแนกข้อสอบข้อที่ m+1

b แทน พารามิเตอร์ค่า ความยากของข้อสอบข้อที่ m+1

- c แทน พารามิเตอร์ระดับโอกาสการเดาข้อสอบได้ถูก ข้อที่  $m+1$
- D แทน จุดบนแกน X
- $O(D)$  แทน ค่าออร์ดิเนต (ordinate) ของโค้งปกติที่จุด D
- A (D) แทน พื้นที่ใต้โค้งปกติจากค่า D จนถึงจุด D

7) การให้คะแนนโดยวิธีการประมาณค่าความเป็นไปได้สูงสุด (maximum likelihood) มีหลายวิธี แต่วิธีที่นิยมใช้มาก คือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (conditional maximum likelihood) การประมาณค่า โดยวิธีนี้มีข้อจำกัด คือ ถ้าผู้สอบตอบข้อสอบถูกหมดหรือผิดหมด จะไม่สามารถประมาณค่าได้ ขั้นตอนในการให้คะแนนมี ดังนี้ (Hambleton and Swanminathan, 1985 อ้างถึงใน ทศน์ศิริรินทร์ สว่างบุญ, 2554; ศิริชัย กาญจนวาสี, 2555)

ขั้นที่ 1 ประมาณค่าความสามารถเริ่มต้น ( $\theta_m = 0; m = 0$ ) โดยใช้สูตรดังนี้

$$\theta_0 = \ln \left( \frac{r_a}{K - r_a} \right)$$

- เมื่อ  $r_a = \sum a_i U_i$
- $U_i = 1$  เมื่อตอบข้อสอบถูก
- $U_i = 0$  เมื่อตอบข้อสอบผิด
- $a_i$  คือ อำนาจจำแนกของข้อสอบข้อที่ i
- K คือ จำนวนข้อสอบทั้งหมดที่ผู้สอบตอบ

ขั้นที่ 2 หาค่า  $P_1(\theta_m)$  และ  $Q(\theta_m)$  โดยใช้สูตร ดังนี้

$$P_1(\theta_m) = c_i + (1 - c_i) \frac{e^{Da_1(\theta_m - b_1)}}{Da_i(\theta_m - b_i)}$$

$$Q(\theta_m) = 1 - P_1(\theta_m)$$

$$= \frac{1 - c_i}{1 + e^{Da_1(\theta_m - b_1)}}$$

- เมื่อ  $\theta_m$  คือ ความสามารถของผู้สอบที่ประมาณค่าได้ในครั้งที่ m
- $a_1$  คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i
- $b_1$  คือ ค่าความยากของข้อสอบข้อที่ i
- $c_1$  คือ ค่าการเดาของข้อสอบข้อที่ t
- D คือ ค่าคงที่ ซึ่งมีค่าเท่ากับ 1.7

e คือ ค่าคงที่ ซึ่งมีค่าเท่ากับ 2.7182

ขั้นที่ 3 หาค่าปรับแก้ ( $h_m$ ) โดยใช้สูตร

$$h_m = \frac{D[r_m - \sum P_i(\theta_m)]}{-D^2 \sum P_i(\theta_m) Q_i(\theta_m)}$$

ขั้นที่ 4 ประเมินค่าความสามารถของผู้สอบใหม่ ( $\theta_m + 1$ ) โดยใช้สูตร

$$\theta_m + 1 = \theta_m - h_m$$

ขั้นที่ 5 ทำการคำนวณซ้ำในขั้นที่ 2,3 และ 4 จนกระทั่ง  $h_m$  เข้าใกล้ศูนย์

$$(h_m < 0.001)$$

จากวิธีการให้คะแนนหรือการประมาณค่าความสามารถ วิธีการประมาณค่าความเป็นไปได้สูงสุด (maximum likelihood estimate) และการประมาณค่าแบบเบย์ (Bayesian estimate) วิธีการคัดเลือกข้อสอบที่ใช้สำหรับการประมาณค่าความเป็นไปได้สูงสุด ประกอบด้วยการคัดเลือกข้อสอบที่มีค่าความยากสอดคล้องกับค่าความสามารถที่ประมาณได้ การคัดเลือกข้อสอบที่มีตำแหน่งสารสนเทศ (ค่าความยาก ค่าอำนาจจำแนกและค่าการเดา) สูงสุดสอดคล้องกับความสามารถที่ประมาณได้และการคัดเลือกข้อสอบที่ให้สารสนเทศสูงสุดที่ตำแหน่งความสามารถนั้น ส่วนวิธีการคัดเลือกข้อสอบที่ใช้สำหรับการประมาณค่าแบบเบย์ใช้การคัดเลือกข้อสอบข้อที่ยังไม่ได้นำมาใช้ในการทดสอบนั้นที่จะให้ค่าความแปรปรวนของค่าความสามารถที่คาดหวังมีค่าต่ำสุด

### 1.5.5 เกณฑ์การยุติการสอบ (termination criteria or Stopping criteria)

ลักษณะสำคัญของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ คือ การทดสอบจะดำเนินไปเรื่อยจนกระทั่งถึงเกณฑ์ที่ระบุไว้ให้ยุติการทดสอบ การดำเนินการสอบก็จะยุติลง เกณฑ์ที่ใช้กันอยู่ทั่วไปมีอยู่ 2 แบบ (ศิริชัย กาญจนวาสี, 2555) คือ

1) ความยาวคงที่ (Fixed-Length) เป็นการกำหนดจำนวนข้อสอบที่ใช้ในการทดสอบให้คงที่ สำหรับทุกคน เช่น กำหนดให้การทดสอบปรับเหมาะใช้ข้อสอบจำนวน 25 ข้อ ( $n = 25$ ) เมื่อผู้สอบทำข้อสอบได้ครบ 25 ข้อ การทดสอบก็จะยุติลง เกณฑ์นี้ค่อนข้างเป็นประโยชน์ในการศึกษาเนื่องจากจำนวนข้อสอบ เท่ากันทำให้สามารถเปรียบเทียบสารสนเทศของแบบสอบได้โดยตรง ในทางปฏิบัติการกำหนดให้ทุกคนทำข้อสอบจำนวนเท่ากัน อาจมีคุณภาพของการวัดผลได้แตกต่างกัน เกณฑ์ความยาวคงที่นี้ใช้งานง่าย สามารถทำนายอัตราการใช้ข้อสอบซ้ำได้ แต่ต้องระวังการใช้กฎความยาวคงที่อาจให้ฟังก์ชันสารสนเทศเป็นโค้งลักษณะแบน ดังนั้นกฎความยาวคงที่อาจไม่ให้ความแม่นยำเท่าเทียมกันที่ระดับความสามารถต่างๆ

2) ความยาวแปรผัน (Variable-Length) เป็นการกำหนดความยาวของข้อสอบแบบไม่คงที่ สำหรับผู้สอบแต่ละคน โดยจะผันแปรไปตามความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ (Standard Error:  $SEE(\theta)$ ) การทดสอบจะดำเนินไปจนกว่าการประมาณค่าความสามารถ ( $\theta$ ) ของผู้สอบมีความคลาดเคลื่อนมาตรฐานลดต่ำลงจนถึงระดับที่ยอมรับได้ การทดสอบจึงยุติลง ดังนั้นการใช้เกณฑ์ยุติการทดสอบจากค่าความคลาดเคลื่อนมาตรฐานจะให้ผลดี คือ มีความแม่นยำในการวัดสูงสำหรับทุกระดับความสามารถของผู้สอบ

$$SEE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

เมื่อ  $SEE(\theta)$  คือ ความคลาดเคลื่อนมาตรฐานของการประมาณค่า  $\theta$

$I(\theta)$  คือ ค่าฟังก์ชันสารสนเทศ หรือ ค่าสารสนเทศที่ได้รับจากแบบสอบสำหรับผู้มีความสามารถ  $\theta$

การกำหนดเกณฑ์การยุติการสอบขึ้นอยู่กับ วัตถุประสงค์ในการทดสอบ ลักษณะคลังข้อสอบ และการควบคุมการทดสอบ การใช้กฎความยาวคงที่โดยกำหนดจำนวนข้อสอบเท่าๆ กัน ขนาดความคลาดเคลื่อนมาตรฐานเมื่อสิ้นสุดการทดสอบ การใช้กฎความยาวแปรผันโดยจัดข้อสอบให้อย่างต่อเนื่องจนกระทั่งระดับความแม่นยำของแต่ละผู้สอบมาถึงเกณฑ์กำหนด

### 1.5.6 การควบคุมการใช้ข้อสอบซ้ำ (Exposure Control)

การใช้ข้อสอบซ้ำ (Item Exposure) คือ การนำข้อสอบจากคลังข้อสอบไปใช้กับผู้สอบในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ ผู้สอบแต่ละคนจะได้รับข้อสอบที่ถูกคัดเลือกให้เหมาะสมกับความสามารถของตนเอง ข้อสอบที่นำไปใช้นั้นจะให้สารสนเทศสูงสุด ณ ระดับความสามารถของผู้สอบที่ถูกประมาณค่า เมื่อพิจารณาจากโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์

$$P_1(\theta_m) = c_i + (1 - c_i) \frac{e^{Da_1(\theta_m - b_1)}}{Da_i(\theta_m - b_i)} \text{ และฟังก์ชันสารสนเทศของข้อสอบ}$$

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

เมื่อ  $I_i(\theta)$  คือ สารสนเทศที่ได้จากข้อสอบข้อที่  $i$  สำหรับผู้สอบความสามารถ  $\theta$

$P_i(\theta)$  คือ ความน่าจะเป็นที่ผู้สอบความสามารถ  $\theta$  จะตอบข้อสอบที่  $i$  ได้ถูกต้อง

$P'_i(\theta)$  คือ อนุพันธ์ของ  $P_i(\theta)$  หรือ ความชันของฟังก์ชันการตอบสนองข้อสอบข้อที่  $i$  เมื่อพิจารณาจากตำแหน่งความสามารถ  $\theta$

$Q_i(\theta)$  คือ ความน่าจะเป็นที่ผู้สอบความสามารถ  $\theta$  จะตอบข้อที่  $i$  ผิด โดยที่

$$Q_i(\theta) = 1 - P_i(\theta)$$

ค่าสารสนเทศของข้อสอบจะมีค่าสูงขึ้นได้นั้น มี 2 กรณี คือ 1) เมื่อค่าพารามิเตอร์อำนาจจำแนกของข้อสอบมีค่ามากขึ้นแล้วค่าสารสนเทศของข้อสอบจะมีค่าเพิ่มขึ้น และ 2) เมื่อค่าความสามารถของผู้สอบ  $\theta$  เข้าใกล้ค่าพารามิเตอร์ความยากของข้อสอบแล้วค่าสารสนเทศของข้อสอบจะมีค่าสูงขึ้นจากข้อค้นพบดังกล่าวทำให้อนุมานได้ว่า วิธีการคัดเลือกข้อสอบที่มีค่าสารสนเทศสูงสุด (ML) ซึ่งเป็นวิธีที่ได้รับความนิยมมากในการทดสอบปรับเหมาะเนื่องจากให้ค่าสารสนเทศสูงสุด มักคัดเลือกข้อสอบที่มีค่าพารามิเตอร์อำนาจจำแนกสูงๆ จึงทำให้ข้อสอบบางส่วนในคลังข้อสอบถูก นำมาใช้บ่อย หรือถูกเปิดเผยกับผู้สอบถี่เกินไป (Overexposure) ขณะที่ข้อสอบส่วนที่เหลือแทบไม่ถูกนำออกมาใช้เลย (Underexposure or Neverexposure) และการใช้ข้อสอบอย่างไม่สมดุลคือเลือกใช้เฉพาะข้อสอบที่มีค่าพารามิเตอร์อำนาจจำแนกสูง แบบสอบที่ผู้สอบแต่ละคนได้รับจะถูก ละเลยในเรื่องของความสมดุลของเนื้อหา (Balancing) ซึ่งได้รับกำหนดไว้ในผังแบบสอบ (test blueprint) หรือ ตารางกำหนดคุณลักษณะของข้อสอบ (Table of Content Specification)

ผลของการใช้ข้อสอบซ้ำถี่เกินไปก่อให้เกิดปัญหาในการวัดและประเมิน เช่น กลุ่มของข้อสอบที่มีค่าพารามิเตอร์อำนาจจำแนกสูงๆ เมื่อนำออกมาใช้บ่อยจนรู้จักแพร่หลายในกลุ่มผู้สอบ จะนำไปสู่ปัญหาในเรื่องความปลอดภัยของแบบสอบ (Test Security) ซึ่งเกิดจากการร่วมใช้ข้อสอบ (Item Sharing) ระหว่างผู้สอบ ผู้สอบจะได้รับข้อมูลของแบบสอบจากผู้ที่เคยสอบก่อนหน้านี้และสามารถสอบตอบคำถามข้อนั้นได้อย่างง่ายดายโดยการล่วงรู้ข้อสอบก่อน (Item Pre-Knowledge) โดยไม่ต้องใช้ความสามารถของตนเอง คะแนนที่สังเกตได้จะขาดความถูกต้อง ผู้สอบสามารถทำคะแนนได้มากเกินความสามารถจริงของตนเอง และจะสูญเสียคุณสมบัติทางจิตมิติขาดความตรงตามสภาพเนื่องจากแบบสอบไม่สามารถวัดได้ตรงตามสภาพของความสามารถจริงของผู้สอบ ถ้าปัญหาดังกล่าวไม่ได้รับการแก้ไขจะส่งผลกระทบต่อความปลอดภัยของข้อสอบและประสิทธิภาพการใช้ข้อสอบในคลังข้อสอบ (Pool Utilization) ต่อมานักทดสอบทั้งทางด้านจิตวิทยาและทางด้านการศึกษาหลายท่านจึงให้ข้อเสนอว่า เพื่อให้เกิดความเหมาะสมในทางปฏิบัติ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์ควรให้ความสำคัญในเรื่องการควบคุมการใช้ข้อสอบซ้ำ (Item Exposure Control) และการสร้างความสมดุลของเนื้อหา (Content Balancing) (สุदारัตน์ หวลมุกดา, 2550)

วิธีการควบคุมการใช้ข้อสอบซ้ำมีหลายวิธี ดังนี้

ตาราง 2.2 จำแนกวิธีการควบคุมการใช้ข้อสอบซ้ำและจุดเด่นและข้อจำกัดของแต่ละวิธี

ประเภทกลยุทธ์	จุดเด่น	ข้อจำกัด
<b>Randomization</b> ประกอบด้วย <ul style="list-style-type: none"> <li>- 5-4-3-2-1 Procedure</li> <li>- Randomesque strategy</li> <li>- Choose One of Three</li> <li>- Within .10 Logits Procedure</li> <li>- Progressive strategy</li> </ul>	เป็นวิธีที่ง่ายในการนำไปใช้	เนื่องจากการใช้กระบวนการสุ่มจึงไม่รับประกันในเรื่องของอัตราการใช้ข้อสอบซ้ำที่เกิดขึ้น และสารสนเทศได้ต่ำเพราะเป็นเลือกข้อสอบอย่างสุ่ม
<b>Conditional Selection</b> ประกอบด้วย <ul style="list-style-type: none"> <li>- Simpson-Hetter (SH) strategy</li> <li>- Extended SH strategy</li> <li>- Stocking and Lewis Multinomial strategy</li> <li>- Restricted Maximum Information strategy</li> <li>- Stocking and Lewis Conditioning on Estimated Ability</li> <li>- Chen and Lei strategy</li> <li>- Shadow Test approach</li> </ul>	- ควบคุมความน่าจะเป็นของข้อสอบที่ได้รับการเลือกแล้วให้มีการใช้ข้อสอบเป็นไปตามกำหนด - วิธีนี้รับรองว่าอัตราการใช้ข้อสอบสูงสุดเป็นไปตามกำหนดล่วงหน้า	ขั้นตอนมีความซับซ้อนและต้องศึกษาในสถานการณ์จำลองโดย การทำซ้ำจำนวนมากและในการใช้งานจริง เมื่อองค์ประกอบต่างๆ ในการทดสอบเปลี่ยนแปลงไป จำเป็นต้องศึกษาในสถานการณ์จำลองใหม่ทุกครั้งเพื่อกำหนด พารามิเตอร์ควบคุมอัตราการใช้ข้อสอบซ้ำ
<b>Stratified Strategies</b> ประกอบด้วย <ul style="list-style-type: none"> <li>- a-Stratified strategy (a-STR)</li> <li>- a-STR with b-Blocking</li> <li>- 0-1 Stratification strategy</li> <li>- a-STR with Content</li> </ul>	- เพิ่มอัตราการใช้ข้อสอบซ้ำในข้อสอบที่มีค่าอำนาจจำแนกต่ำ และทำให้อัตราการใช้ข้อสอบซ้ำของข้อสอบทั้งคลังข้อสอบมีความเท่าเทียมกัน	ข้อสอบที่บรรจุภายในแต่ละชั้นควรมีค่าอำนาจจำแนกใกล้เคียงกัน และมีช่วงของค่าความยากของข้อสอบครอบคลุมช่วงความสามารถของผู้สอบ
<b>Balancing</b> <ul style="list-style-type: none"> <li>- Multi-dimensional</li> </ul>		
<b>Stratification</b> <ul style="list-style-type: none"> <li>- 0-1 Stratification strategy</li> </ul>		



ตาราง 2.2 จำแนกวิธีการควบคุมการใช้ข้อสอบซ้ำและจุดเด่นและข้อจำกัดของแต่ละวิธี (ต่อ)

ประเภทกลยุทธ์	จุดเด่น	ข้อจำกัด
<b>Combined Strategies</b>	เป็นการนำเอาลักษณะเด่นของ	ข้อจำกัดที่มีจะได้รับสืบทอด
ประกอบด้วย	แต่ละวิธีมารวมกัน ทำให้มี	มาจาก วิธีการต้นแบบที่
- Progressive Restricted strategy	ความสามารถมากกว่ากลยุทธ์	นำมาใช้
- Progressive-restricted	เดียว จากงานวิจัยในอดีตพบว่า มี	
standard error (PR-SE)	การนำเทคนิคมาบูรณาการเข้า	
- a-Stratified Design with the	ด้วยกันเป็นจำนวนมาก	
Sympton-Hetter		
- Algorithm		
- Incorporation of the SH into		
a-STR with Content Blocking		
- Constraints CAT using Shadow		
Test		
- Computerized Adaptive		
Sequential Testing		
- Adaptive Multi-stage Item		
Bundles		
- Multiple Forms Structures		
- Testlet-Based Adaptive		
Mastery Testing		
- Two-Phase ISP for Flexible		
Content Balancing		

ที่มา Elissavet, Evangelos, and Economides (2007) และ อนุสรณ์ เกิดศรี (2557)

วิธีการควบคุมการใช้ข้อสอบซ้ำ วิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PR-SE)

วิธีจำกัดความก้าวหน้า (Progressive-restricted Procedure) พัฒนาโดย Revuelta and Ponsoda (1998) ซึ่งเป็นวิธีที่พัฒนามาจากวิธีสุ่มเลือกแบบก้าวหน้า (Progressive-restricted Procedure) และ วิธีจำกัดสารสนเทศสูงสุด (restricted maximum information) ซึ่งพัฒนาเพื่อการป้องกันไม่ให้มีการเปิดเผยข้อสอบมากเกินไปและเพิ่มอัตราการคัดเลือกข้อสอบที่ไม่ค่อยได้รับการถูกเลือก อีกทั้งยังเป็นวิธีที่มีการรักษาความแม่นยำในการประมาณค่าความสามารถ

วิธีจำกัดความก้าวหน้า (Progressive-restricted Procedure) ได้พัฒนามาจากวิธีจำกัดสารสนเทศสูงสุด (maximum information method) ที่ข้อสอบที่มีสารสนเทศสูงสุดจะได้รับ

การคัดเลือกเข้าสู่ระบบการทดสอบ และวิธีจำกัดสารสนเทศสูงสุด (restricted maximum information) ซึ่งคัดเลือกข้อสอบที่มีสารสนเทศสูงสุด แต่ก็ข้อสอบก็ไม่ได้การคัดเลือก หลังจากปรากฏในการทดสอบ (100K%) ซึ่งการคำนวณ  $k$  เริ่มต้นจากการบริหารจัดการการทดสอบ ข้อสอบข้อที่  $i$  ได้รับการบริหารจัดการ  $a_i$  ครั้ง ในการทดสอบ  $t$  อัตราการเปิดเผยข้อสอบ  $k_i$  คำนวณจาก  $a_i/t$  เฉพาะข้อสอบที่มีต่ำกว่าอัตราการเปิดเผย  $k$  เหมาะสมที่จะได้รับการคัดเลือก

วิธีจำกัดความก้าวหน้า (Progressive-restricted Procedure) นำวิธีการสุ่มสารสนเทศสูงสุด โดยคำนวณค่าน้ำหนักข้อสอบแต่ละข้อโดยใช้สูตรคำนวณ

$$W_i = (1 - s) R_i + S I_i \text{ -----สมการที่ 1}$$

เมื่อ  $W_i$  = การกำหนดน้ำหนักของข้อสอบ

$S$  = ตำแหน่งความสัมพันธ์ของข้อสอบ

$R_i$  = สุ่มเลือกจำนวนข้อสอบที่มีการกระจายเหมือนกัน

$I_i$  = สารสนเทศของข้อสอบ

$s$  คำนวณได้จาก  $h/m$  เมื่อ  $h$  คือจำนวนข้อสอบที่ได้รับการบริหารจัดการในการทดสอบครั้งก่อน ในการทดสอบที่มีความยาวสูงสุด  $m$  ข้อ ซึ่งองค์ประกอบจะสูญเสียความสำคัญของการทดสอบเพราะจะถูกคูณด้วย  $(12s)$  ตรงกันข้ามบางส่วนของสารสนเทศจะเป็นส่วนที่สำคัญจะถูกคูณด้วย  $s$  ทางเลือกที่ทำให้น้ำหนักข้อสอบมีค่าเพิ่มขึ้นของข้อสอบในคลังข้อสอบถูกใช้ (เนื่องจากองค์ประกอบสุ่ม) ในขณะที่ยังคงรักษาความแม่นยำของการประมาณค่า (เนื่องจากค่าสารสนเทศข้อสอบ)

วิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error) พัฒนาโดย McClarty et al. (2006) ซึ่งใช้สูตรคำนวณเดียวกับวิธีจำกัดความก้าวหน้า (Progressive-restricted Procedure) ในสมการที่ 1 แต่มีการนิยาม  $s$  ใหม่เป็นอัตราส่วนของความคลาดเคลื่อนมาตรฐาน (SE) กฎการหยุด มากกว่าความคลาดเคลื่อนมาตรฐาน (SE) ปัจจุบัน การใช้ความคลาดเคลื่อนมาตรฐาน (SE) ไม่ได้ขึ้นอยู่กับตำแหน่งตามลำดับ แต่ใช้ความคลาดเคลื่อนมาตรฐาน (SE) เพื่อตรวจสอบองค์ประกอบการสุ่ม หรือ สารสนเทศที่มีอิทธิพลต่อการถ่วงน้ำหนัก เช่นเดียวกับขั้นตอนการสุ่ม ขั้นตอนการทำงานร่วมกันง่ายกว่าวิธีการแบบเงื่อนไข

McClarty et al. (2006) ได้เปรียบเทียบวิธีจำกัดความก้าวหน้าความคลาดเคลื่อนมาตรฐาน (Progressive-restricted standard error) กับวิธีจำกัดความก้าวหน้า (Progressive-restricted Procedure) และไม่มีการควบคุมการเปิดเผยข้อสอบ โดยกำหนดการให้คะแนนแบบ

2 ค่า และ 3 พารามิเตอร์โลจิสติก (3PL) และการให้คะแนนแบบบางส่วน (partial credit model) และกำหนดขนาดคลังข้อสอบ ลักษณะการกระจายของข้อมูลและกฎการหยุด พบว่าภายใต้โมเดล 3 พารามิเตอร์โลจิสติก (3PL) วิธีจำกัดความก้าวหน้าความคลาดเคลื่อนมาตรฐานใช้ข้อสอบน้อยกว่าวิธีจำกัดความก้าวหน้า ที่แตกต่างในจำนวนของการบริหารจัดการข้อสอบได้มากขึ้นภายใต้การกระจายข้อมูลปกติ ผลการทดสอบคลังข้อสอบขนาดเล็กพบว่า ค่าความสามารถของผู้สอบมีความคลาดเคลื่อนมาตรฐาน (SE) เพิ่มขึ้น ซึ่งสอดคล้องกับการจำลองข้อมูลที่มีความคลาดเคลื่อน (SE) เพิ่มขึ้น ความสัมพันธ์ระหว่างการประมาณค่าและการรู้ค่าความสามารถของผู้สอบคล้ายคลึงกันระหว่างวิธี Progressive-restricted standard error กับวิธีจำกัดความก้าวหน้าในทุกเงื่อนไข และเปอร์เซ็นต์ข้อสอบที่เข้าซ้อนกันมีค่าต่ำสำหรับทั้งสองในทุกเงื่อนไข สำหรับการให้คะแนนแบบบางส่วน (partial credit model) ไม่พบความแตกต่างระหว่างวิธี Progressive-restricted standard error กับวิธีจำกัดความก้าวหน้า วิธี Progressive-restricted standard error พบการใช้คลังข้อสอบขนาดใหญ่ในการบริหารจัดการข้อสอบ โดยรวม Progressive-restricted standard error ดำเนินการบริหารการทดสอบเหมือนกับวิธีจำกัดความก้าวหน้า และสนับสนุนการใช้ วิธี Progressive-restricted standard error กับ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบความยาวแปรผัน

วิธี Progressive-restricted standard error มีจุดแข็งในการดำเนินการบริหารจัดการเมื่อเปรียบเทียบกับขั้นตอนการควบคุมการเปิดเผยข้อสอบวิธีอื่นๆ ในการทดสอบ ในขณะที่ วิธีจำกัดความก้าวหน้า เหมือนกับวิธีการอื่นๆ ในการทดสอบแบบกำหนดความยาวคงที่ เนื่องจากมีการกำหนดตำแหน่งตามลำดับ ในขณะที่วิธี Progressive-restricted standard error เหมาะสมทั้งการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบกำหนดความยาวคงที่ และ แบบความยาวแปรผัน เนื่องจาก s จะกำหนดค่าความคลาดเคลื่อนมาตรฐาน (SE) ของการประมาณค่า

### 1.5.7 การสร้างความสมดุลของเนื้อหา (Content Balancing)

ความสมดุลของเนื้อหา (Content Balancing) ตามการกำหนดคุณลักษณะของข้อสอบ (Content Specification) เป็นสิ่งสำคัญโดยเฉพาะอย่างยิ่งในการวัดผลสัมฤทธิ์ทางการศึกษา ในการทดสอบแบบดั้งเดิมที่เป็นมาตรฐาน แบบสอบที่มีรูปแบบคู่ขนานส่วนมากถูกสร้างตามรายละเอียดของ ตารางการกำหนดคุณลักษณะของข้อสอบ (Table of Content Specification) และในการออกแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ควรคำนึงถึงข้อจำกัดในทางปฏิบัติ (Practical Constraints) เช่น การกำหนดคุณลักษณะของข้อสอบ และความสมดุลของเนื้อหา (Content Balancing) เพื่อรับประกันว่าข้อสอบที่ผู้สอบแต่ละคนได้รับเป็นตัวแทนตามสัดส่วนของแต่ละขอบเขตเนื้อหา (Content Area) และเพื่อให้แน่ใจว่าผู้ที่เข้าสอบได้รับการทดสอบในเนื้อหาที่เท่ากัน และเหมาะสมในระหว่างการใช้ CAT วิธีการคัดเลือกข้อสอบควรจะต้องเลือกข้อสอบที่ให้

สารสนเทศมากที่สุดในการประมาณค่าความสามารถของผู้ที่เข้าสอบในช่วงเวลานั้น แต่ไม่บรรลุถึงการกำหนดแบบ แผนของเนื้อหา (Content Specifications) ของข้อสอบทั้งหมด ตัวอย่างเช่น การทดสอบทางด้านการคำนวณเพื่อที่จะให้ทราบถึงความสามารถของผู้เข้าสอบในเรื่อง การบวก การลบ การคูณ และ การหาร ทั้งหมด แต่ใน CAT อาจจะต้องคัดเลือกให้ทำข้อสอบเฉพาะเรื่องบางเรื่องเท่านั้น เช่น เรื่องการ ลบ และการคูณ ซึ่งผู้เข้าสอบจะไม่ได้รับการสอบในเรื่องการบวก และการหาร ผู้สอบที่ไม่มีความรู้ เกี่ยวกับการบวกและการหาร ก็จะถูกประมาณค่าความสามารถสูงเกินจริงสำหรับผู้เข้าสอบที่มี ความรู้เกี่ยวกับเรื่องบวกและหาร จะได้รับการประมาณค่าความสามารถที่แท้จริงของต่ำเกินไป ดังนั้นการสร้างคุณสมบัติของเนื้อหาจึงเป็นสิ่งจำเป็นในการออกแบบระบบ CAT (Boyd, 2003 อ้างถึงใน Elissavet et al., 2007)

## 1.6 งานวิจัยที่เกี่ยวข้องกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์

### 1.6.1 งานวิจัยที่เกี่ยวข้องกับการคัดเลือกข้อสอบ

จากการศึกษางานวิจัย ของ Y. Cheng and Chang (2009) ที่ทำการศึกษาวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index) สำหรับการคัดเลือกข้อสอบที่มีข้อบังคับอย่างเข้มงวด พบว่า การคัดเลือกข้อสอบวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) เป็นวิธีการที่มีการลดการฝ่าฝืนข้อบังคับ และมีประสิทธิภาพสามารถปรับให้เข้ากับข้อบังคับที่ไม่ใช่ทางสถิติที่หลากหลายและสามารถควบคุมได้หลายอย่างพร้อมๆ กัน เช่น ความสมดุลของเนื้อหา การควบคุมการใช้ข้อสอบซ้ำ ความสมดุลของคำตอบ สามารถนำไปปรับใช้กับโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ที่มีอยู่ในปัจจุบันได้ สะดวก โดยไม่จำเป็นต้องปรับค่าน้ำหนักความสัมพันธ์ระหว่างข้อบังคับและสารสนเทศ ซึ่งสอดคล้องกับงานวิจัยของ He, Diao, and Hauser (2014) ที่ได้ศึกษาเปรียบเทียบวิธีการคัดเลือกข้อสอบ ที่พัฒนาขึ้นเพื่อใช้กับข้อจำกัดอย่างรุนแรงในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CATs) ข้อจำกัดที่รุนแรงของ CATs อ้างถึงการทดสอบปรับเหมาะที่พยายามจะตอบสนองต่อชุดที่ซับซ้อนของข้อจำกัดที่มักจะไม่ได้ข้อสรุปกัน เช่น ข้อสอบอาจนำไปสู่ความพึงพอใจของข้อจำกัดหลายอย่างในเวลาเดียวกัน พบว่า วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) ทำงานได้ดีที่สุดในวิธีการทั้งหมดในแง่ของความถูกต้องในการวัดและการจัดการข้อจำกัดอื่น ๆ

สำหรับการศึกษางานวิจัยของ Deng, Ansley, and Chang (2010) ได้ทำการศึกษาระบบและวิธีการคัดเลือกข้อสอบที่มีสารสนเทศสูงสุดในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ พบว่า การคัดเลือกข้อสอบวิธีสารสนเทศสูงสุดของฟิชเชอร์ (maximum Fisher's information: MFI) มีความแม่นยำในการประมาณค่าความสามารถ ใช้งานที่ดีเมื่อข้อสอบได้รับการคัดเลือกภายใต้การควบคุมและการเปิดเผยกับการขอบเขตพื้นที่ในการคัดเลือกที่จำกัด ผลการทดสอบให้ความหมาย

สำหรับการเลือกวิธีการคัดเลือกข้อสอบที่เหมาะสมที่ประยุกต์ได้ ซึ่งสอดคล้องกับงานวิจัยของ Daniel, Barbara, and Brandon (2010) และ Ying Cheng, Patton, and Shao (2014) ที่พบว่า วิธีการคัดเลือกข้อสอบของสารสนเทศสูงสุดของฟิชเชอร์ (MFI) มีประสิทธิภาพสามารถรักษาความสมดุลของการใช้คลังข้อสอบให้ประมาณค่าคุณลักษณะแฝงที่ถูกต้องในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) และเมื่อใช้รวมกับการควบคุมการเปิดเผยข้อสอบไวต่อโอกาสที่เกิดขึ้นจริง โดยเฉพาะอย่างยิ่งเมื่อการทดสอบความถูกต้องของกลุ่มตัวอย่างที่มีขนาดเล็ก

### 1.6.2. งานวิจัยที่เกี่ยวข้องกับการควบคุมการใช้ข้อสอบซ้ำ

Lee and Dodd (2011) ทำการศึกษาการเปรียบเทียบการควบคุมการใช้ข้อสอบซ้ำ ลักษณะของคลังข้อสอบและการกระจายของประชากรสำหรับ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์ ที่ใช้ Partial Credit Model การศึกษานี้การตรวจสอบวิธีการควบคุมการใช้ข้อสอบซ้ำภายใต้ชุดต่างๆ ของลักษณะคลังข้อสอบและการกระจายความสามารถในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ที่ขึ้นอยู่กับ Partial Credit Model มีการศึกษา ตัวแปร 3 ตัว ได้แก่ ลักษณะของคลังข้อสอบ (120 ข้อ สำหรับง่าย, ปานกลางและยากในคลังข้อสอบ) การกระจายความสามารถมี 2 วิธี (การกระจายปกติและการกระจายข้อมูลเบ้ในเชิงลบ) และมีการควบคุมการใช้ข้อสอบซ้ำ 3 วิธี (Randomsque, วิธี progressive-restricted และวิธี maximum information) จำนวนของดัชนีการวัดความแม่นยำ เช่น สถิติเชิงพรรณนา, ความสัมพันธ์ระหว่างที่รู้จักและระดับการประมาณค่าความสามารถ, bias, RMSES และAAD อัตราการทำซ้ำ การใช้ข้อสอบ และการทับซ้อนของข้อสอบที่ได้รับการคำนวณในการประเมินผลกระทบของการจับคู่หรือไม่ได้จับคู่ ในคลังข้อสอบและการกระจายความสามารถในความถูกต้องของการประมาณค่าความสามารถและประสิทธิภาพของการควบคุมการใช้ข้อสอบซ้ำโดยเป็นที่คาดหวังว่าคลังข้อสอบขนาดกลางให้ข้อสอบที่มีความแม่นยำในการวัดที่ดีขึ้นกว่าคลังข้อสอบที่ง่ายและยาก วิธีการ progressive-restricted มีประสิทธิภาพในอัตราการทำซ้ำสูงสุดการทับซ้อนเฉลี่ยของข้อคำถามและการใช้คลังข้อสอบดีกว่าทั้ง วิธีการสุ่มและวิธีการให้สารสนเทศสูงสุด คลังข้อสอบที่ง่ายมีข้อมูลที่บิดเบือนในทางลบซึ่งเป็นเงื่อนไขที่ไม่ตรงกันผลซึ่งเป็นประสิทธิภาพที่ไม่ดี

Leroux, Lopez, Hembry, and Dodd (2013) ได้ศึกษาเรื่องการเปรียบเทียบวิธีการควบคุมการใช้ข้อสอบซ้ำใน CATs โดยใช้โมเดล 3 พารามิเตอร์ ซึ่งในการศึกษานี้เป็นการเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (progressive-restricted standard error ; PR-SE) ของวิธีการควบคุมการใช้ข้อสอบซ้ำที่ใช้โดยทั่วไปในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ 3 วิธี ได้แก่ วิธีการสุ่ม (randomesque), วิธี Sympton-Hetter (SH) และ ไม่มีการควบคุมการใช้ข้อสอบ ประสิทธิภาพของวิธีการทั้ง 4 วิธี นี้ถูกประเมินจากการใช้โมเดลโลจิสติก 3 พารามิเตอร์ ภายใต้เงื่อนไขที่จัดกระทำตามขนาดของคลังข้อสอบเล็กกับใหญ่ และเกณฑ์

การยุติการสอบ (ความยาวจำกัดและความยาวแปรผัน) PR-SE ให้ประโยชน์ของการจำกัดที่คล้ายกันกับ SH โดยไม่ต้องการสำหรับการศึกษาการจำลองสถานการณ์ที่ดำเนินการมาก่อนหน้านี้ โดยภาพรวมสำหรับคลังข้อสอบขนาดใหญ่และเล็ก วิธี PR-SE สามารถบริหารจัดการได้เกือบทั้งหมดของคลังข้อสอบ ในขณะที่วิธีการอื่นๆสามารถบริหารจัดการได้เพียง 52% หรือน้อยกว่าสำหรับคลังข้อสอบขนาดใหญ่ และทำได้ 80% หรือน้อยกว่าสำหรับคลังข้อสอบขนาดเล็ก วิธี PR-SE ยังทำให้มีจำนวนการซ้อนทับกันน้อยที่สุดของข้อสอบระหว่างแบบสอบข้ามเงื่อนไขและการบริหารจัดการข้อสอบน้อยกว่าวิธี SH ซึ่งวิธี PR-SE ยังให้ผลที่คล้ายกันและยอมรับได้ความถูกต้องแม่นยำของการวัดถูกเปรียบเทียบกับวิธีการควบคุมการเปิดเผยข้อสอบวิธีอื่นๆ ในขณะที่มีการปรับปรุงการใช้คลังข้อสอบอย่างกว้างขวาง

### 1.6.3. งานวิจัยที่เกี่ยวข้องกับคลังข้อสอบ

Reckase (2010) ได้ทำการศึกษาเกี่ยวกับการออกแบบคลังข้อสอบ ให้เหมาะสมกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ด้วยเทคนิค p-optimal และ r-optimal พบว่าการออกแบบคลังข้อสอบนั้นส่งผลต่อประสิทธิภาพของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ แต่ผลจากการศึกษายังไม่สามารถให้ข้อสรุปที่ชัดเจนได้ว่าคลังข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ควรจะมีความใหญ่เท่าใด

He and Reckase (2014) ได้ศึกษาเรื่องการออกแบบคลังข้อสอบสำหรับการดำเนินการตามความยาวของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CATs) จะทำงานได้ดีจะต้องมีคลังข้อสอบที่มีจำนวนข้อสอบที่มีคุณภาพดีอย่างเพียงพอ ซึ่งนักวิจัยหลายคนได้ชี้ให้เห็นว่าการพัฒนาคลังข้อสอบสำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ไม่เพียงแต่เรื่องขนาดคลังข้อสอบที่สำคัญ แต่ยังมีเรื่องการแจกแจงของพารามิเตอร์ข้อสอบและการพิจารณาในทางปฏิบัติ เช่น การแจกแจงเนื้อหาและประเด็นการเปิดเผยข้อสอบ ยังมีงานวิจัยเล็ก ๆ น้อย ๆ ในการออกแบบคลังข้อสอบที่จะมีคุณสมบัติที่พึงประสงค์เหล่านั้น ในรายงานการวิจัยนี้ได้ให้ขั้นตอนที่ละเอียดขึ้นด้วยคู่มือคำแนะนำเกี่ยวกับขั้นตอนการออกแบบคลังข้อสอบโดยใช้วิธีการทั้งและการรวมกัน ในการออกแบบคลังข้อสอบสำหรับการทดสอบขนาดใหญ่เพื่อรับใบอนุญาต การทดสอบปรับเหมาะถูกใช้กับการทดสอบปรับเหมาะที่ซับซ้อนกับตัวแปรความยาวของแบบสอบ การตัดสินใจขึ้นอยู่กับกฎการยุติการทดสอบ ความสมดุลของเนื้อหา และควบคุมการใช้ข้อสอบซ้ำ กระบวนการออกแบบที่เกี่ยวข้องกับการจำลองที่ครอบคลุม เพื่อระบุการออกแบบคลังข้อสอบทางเลือกและประเมินผลการทำงานกับชุดของเกณฑ์ การออกแบบผลลัพธ์รวมถึงขนาดคลังข้อสอบที่ต้องการและการแจกแจงค่าพารามิเตอร์ข้อสอบ ผลการวิจัยพบว่ากลไกที่ใช้ในการระบุตัวตนของคลังข้อสอบที่ต้องการคุณสมบัติการทำงานที่ดี

#### 1.6.4 งานวิจัยที่เกี่ยวข้องกับกฎการหยุด

Choi, Grad, and Dodd (2011) และคณะได้ศึกษาเรื่อง กฎการหยุดใหม่ สำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ วัตถุประสงค์ของการศึกษาในปัจจุบัน คือ การแนะนำกฎการหยุดใหม่ สำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) ที่ทำนายการลดความคลาดเคลื่อนมาตรฐาน (PS-ER) โดยการใช้กฎการหยุดความแปรปรวนหลังการทำนายที่จะตรวจสอบความคลาดเคลื่อนมาตรฐานที่จะเป็นผลมาจากการดำเนินงานของข้อสอบเพิ่มเติม และประสิทธิภาพการทำงานของการทำงานการทำนายการลดความคลาดเคลื่อนมาตรฐาน (PS-ER) เมื่อเทียบกับกฎการหยุดความคลาดเคลื่อนมาตรฐานเพียงเล็กน้อย และการปรับปรุงรูปแบบข้อมูลของกฎการหยุดเพียงเล็กน้อยในการจำลองการทดสอบปรับเหมาะที่ดึงมาจากคลังข้อสอบ โดยพบว่า การทำนายความคลาดเคลื่อนมาตรฐาน (PS-ER) ทำให้คลังข้อสอบในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์มีประสิทธิภาพ มีการดำเนินงานข้อสอบน้อยกว่าการคาดการณ์เมื่อได้ประโยชน์จากข้อมูลจำนวนน้อย และเพิ่มความแม่นยำในการตรวจวัดเมื่อข้อมูลสมบูรณ์

#### 1.6.5 งานวิจัยที่เกี่ยวข้องกับความยาวของข้อสอบ

Patton, Cheng, Yuan, and Diao (2013) ได้ศึกษาเรื่อง อิทธิพลของขนาดความคลาดเคลื่อนของข้อสอบบนตัวแปรความยาวของข้อสอบในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ในการวิจัยนี้ได้กล่าวถึงตัวแปรความยาวของข้อสอบในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (VL-CAT) ที่รวมทั้งความยาวของข้อสอบและแบบสอบที่เหมาะสมสำหรับผู้สอบแต่ละคน โดยทั่วไปขึ้นอยู่กับค่าความคลาดเคลื่อนมาตรฐานของประมาณค่าความสามารถ ซึ่งขึ้นอยู่กับค่าพารามิเตอร์ของข้อสอบ อย่างไรก็ตามข้อสอบที่ถูกเลือกมาเพื่อประมาณค่าพารามิเตอร์พื้นฐานและการใช้ประโยชน์บนโอกาสที่อาจจะเกิดขึ้น ในการศึกษาครั้งนี้ผู้วิจัยได้ตรวจสอบผลกระทบของประโยชน์บนโอกาสของความยาวแบบสอบและความถูกต้องของการจัดกลุ่มในสถานการณ์ VL-CAT ที่หลากหลาย ผลการวิจัยยืนยันว่าการใช้ประโยชน์ของโอกาสเกิดขึ้นใน VL-CAT และมีผลกระทบที่ซับซ้อนบนความยาวของแบบสอบ การประมาณค่าความสามารถ และความถูกต้องของการจัดกลุ่ม ผลการวิจัยมีความสำคัญต่อการใช้สำหรับออกแบบและนำ VL-CATs ไปใช้

จากการศึกษาวิจัยที่เกี่ยวข้องกับการคัดเลือกข้อสอบและการควบคุมการใช้ข้อสอบซ้ำของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ พบว่า วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) เป็นวิธีเพื่อใช้ควบคุมเงื่อนไขบังคับอย่างเข้มงวดในการคัดเลือกข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ จะพิจารณาความแปรปรวนของค่าสารสนเทศโดยการนำดัชนีไปเป็นตัวคูณให้กับค่าสารสนเทศของข้อสอบ ดังนั้นในการคัดเลือกข้อสอบแทนที่จะพิจารณาจากค่าสารสนเทศเพียงอย่างเดียววิธีการนี้จะนำลำดับความสำคัญพิจารณาร่วมด้วย และสามารถปรับให้เข้ากับข้อบังคับที่ไม่ใช่ทางสถิติที่หลากหลายและสามารถควบคุมได้หลายอย่างพร้อมๆ กัน เช่น

ความสมดุลของเนื้อหา การควบคุมการใช้ข้อสอบซ้ำ ความสมดุลของคำตอบ โดยที่วิธีการนี้ สามารถนำไปปรับใช้กับโปรแกรมการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์ที่มีอยู่ในปัจจุบันได้ สะดวก โดยไม่จำเป็นต้องปรับค่าน้ำหนักความสัมพันธ์ระหว่างข้อบังคับและสารสนเทศ และวิธีการควบคุมการใช้ข้อสอบซ้ำด้วยวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PR-SE) โดยนำอัตราส่วนของความคลาดเคลื่อนมาตรฐาน (SE) มาใช้ในกฎการหยุด โดยการใช้ความคลาดเคลื่อนมาตรฐาน (SE) ไม่ได้ขึ้นอยู่กับตำแหน่งตามลำดับ แต่ใช้ความคลาดเคลื่อนมาตรฐานเพื่อตรวจสอบองค์ประกอบการสุ่ม หรือ สารสนเทศที่มีอิทธิพลต่อการถ่วงน้ำหนัก วิธีการ Progressive-restricted standard error มีการดำเนินการบริหารจัดการที่ดีเมื่อเปรียบเทียบกับขั้นตอน การควบคุมการเปิดเผยข้อสอบวิธีอื่นๆ และการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบกำหนดความยาวคงที่ และ แบบความยาวแปรผัน เนื่องจาก จะกำหนดค่าความคลาดเคลื่อนมาตรฐาน (SE) ของการประมาณค่าความสามารถผู้สอบ ทั้งสองวิธีเป็นวิธีการที่ยืดหยุ่นเหมาะสมกับการนำมาปรับใช้กับการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด

## ตอนที่ 2 การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด

การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์เป็นวิธีการประเมินตามลำดับที่เลือกข้อคำถามจากคลังข้อสอบเพื่อให้ตรงกับระดับความสามารถของผู้สอบในการบริหารการทดสอบ สองวิธีได้รับความนิยมมากที่สุดในการออกแบบการทดสอบตามลำดับได้แก่ การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายข้อ และการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) สองวิธีการนี้ได้ถูกนำไปใช้ในการวัดหลายงาน เช่น การทดสอบการศึกษา การรายงานการรักษาผู้ป่วย และการวัดคุณภาพชีวิต

การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์ทั้งสองวิธีมีจุดแข็งที่ต่างกัน การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายข้อ มีจุดแข็งในการทดสอบที่ผู้เข้ารับการทดสอบได้รับการจัดข้อสอบที่มีความยากตามระดับความสามารถของตน โดยที่ข้อสอบไม่ง่ายหรือยากจนเกินไปเหมือนกับ การทดสอบแบบดั้งเดิมที่ใช้กระดาษในการทดสอบ และจำนวนข้อสอบที่ใช้ในการทดสอบมีจำนวนลดลงโดยที่ไม่สูญเสียความแม่นยำในการประมาณค่าความสามารถของผู้เข้ารับการทดสอบ และส่วน การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มีจุดแข็งที่มีการจัดข้อสอบเป็นชุด กระบวนการในการเลือกข้อสอบเป็นชุดจะกระทำตามลำดับขั้น (stage) เท่านั้น จึงอนุญาตให้ผู้สอบสามารถย้อนหรือข้ามข้อสอบภายในขั้นที่กำลังดำเนินการทดสอบได้ เพื่อทำการทบทวนหรืออาจเปลี่ยนคำตอบ และการทดสอบแบบ MST ยังอนุญาตให้ผู้พัฒนาแบบทดสอบสามารถกำกับติดตาม และควบคุมคุณภาพของชุดแบบทดสอบ และการประมาณค่าความสามารถของผู้สอบเพื่อคัดเลือกชุดข้อสอบที่เหมาะสมกับความสามารถของผู้สอบจะทำการประมาณค่าจากชุดของข้อสอบในขั้นแรก



จึงทำให้ลดความผิดพลาดจากการประมาณค่าความสามารถผู้สอบจากข้อสอบข้อแรกเพียงข้อเดียว ซึ่งหากเกิดความผิดพลาดในการประมาณค่าความสามารถจากข้อสอบข้อแรกจะทำให้ผู้เข้ารับการทดสอบอาจให้เวลานานกว่าปกติในการทดสอบ (Zheng & Chang, 2015)

Wang et al. (2016) ได้ศึกษาจุดแข็งในการดำเนินการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ทั้งสองวิธี จนพัฒนาไปสู่การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เพื่อประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ โดยทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดประกอบไปด้วย แนวคิดการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน ซึ่งมีรายละเอียด ดังนี้

## 2.1 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อเป็นการเลือกข้อสอบให้เหมาะสมกับความสามารถของผู้สอบ โดยแนวทางในการคัดเลือกข้อสอบขึ้นอยู่กับผลการตอบข้อสอบข้อที่ผ่านมาของผู้แต่ละคน ในการทดสอบผู้สอบจะเริ่มทำข้อสอบข้อแรกโดยเป็นข้อที่มีค่าความยากระดับปานกลาง หลังจากนั้นจะทำการประมาณค่าความสามารถของผู้สอบจากผลการตอบข้อสอบ เพื่อทำการคัดเลือกข้อสอบข้อถัดไปที่มีความยากและอำนาจจำแนกที่เหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งมีหลักคือหากผู้สอบตอบข้อสอบข้อที่ผ่านมาถูกต้อง ข้อถัดไปจะมีค่าระดับความยากที่ยากขึ้น แต่หากผู้สอบตอบข้อสอบผิด ข้อถัดไปจะมีค่าระดับความยากที่ลดลง จนกระทั่งการประมาณค่าความสามารถผู้สอบอยู่ในระดับที่น่าเชื่อถือ คืออยู่ในเกณฑ์ของความคลาดเคลื่อนที่กำหนดจึงยุติการทดสอบ (Čisar, Radosav, Markoski, Pinter, & Čisar, 2010; ศิริชัย กาญจนวาสี, 2555)

### 2.1.1 องค์ประกอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ

#### 1) คลังข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ

การวิเคราะห์ข้อสอบเพื่อจัดทำคลังข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อจะวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory Model: IRT) ซึ่งถ้าจำนวนข้อสอบในคลังข้อสอบจะต้องมีจำนวนข้อสอบที่มากพอจะส่งผลต่อการทดสอบปรับเหมาะจะมีประสิทธิภาพสูงขึ้น ถ้ามีคลังข้อสอบขนาดใหญ่ ข้อสอบดังกล่าวจะได้รับการประมาณค่าพารามิเตอร์โดยใช้ทฤษฎีการตอบสนองข้อสอบ ข้อสอบมีค่าอำนาจจำแนกสูง มีค่าความยากครอบคลุมช่วงความสามารถของผู้สอบ และมีค่าการเดาต่ำ Weiss (1988) อ้างถึงใน ศิริชัย กาญจนวาสี (2555) ได้เสนอไว้ว่าจำนวนข้อสอบในคลังข้อสอบควรมีอย่างน้อย 100-200 ข้อ Allen and Yen (1979) อ้างถึงใน ศิริชัย กาญจนวาสี (2555) กล่าวว่า ถ้าจะให้ดีต้องมีจำนวนข้อสอบจำนวน  $2n$  ข้อ เมื่อ  $n$  = จำนวนข้อสอบที่ใช้ในการทดสอบแต่ถ้าการทดสอบมีการกำหนดจำนวนข้อที่จะใช้คงที่เหมือนกันทุกคน จำนวนข้อสอบควรมีอย่างน้อย  $n(n+1)/2$  โดยข้อสอบแต่ละข้อควรมี

อำนาจจำแนกสูง ( $a > 0.80$ ) มีค่าความยากที่ครอบคลุม ระดับช่วงต่างๆ อย่างเหมาะสม  $[-2.0 < \theta < +2.0]$  ค่าสัมประสิทธิ์การเดา ( $c < 0.30$ ) และควรมีข้อสอบเท่าๆกันในแต่ละระดับความยาก

2) การประมาณค่าความสามารถในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ การประมาณค่าความสามารถในการทดสอบแบบรายข้อจะใช้การประมาณค่าด้วยวิธี Bayesian และ วิธี Maximum likelihood (ML) มีรายละเอียดดังนี้

2.1) การให้คะแนนตามวิธี Bayesian มีหลายวิธี แต่วิธีที่นิยมใช้กันมาก คือ วิธีของ เบส์ที่ปรับปรุงให้ทันสมัย (Bayesian updating) ซึ่งสามารถประมาณค่าความสามารถของผู้สอบได้ ค่อนข้างคงที่ ซึ่งมีสูตรดังนี้

กรณีที่ตอบถูก

$$\theta_{m+1} = \theta_m + (1-c) \left( \frac{\sigma_m^2}{\sqrt{\frac{1}{a^2} + \sigma_m^2}} \right) \left( \frac{O(D)}{C + (1-c)A(-D)} \right)$$

$$\sigma_{m+1}^2 = \sigma_m^2 \left( 1 - \left( \frac{1-c}{1 + \frac{1}{a^2 \sigma_m^2}} \right) \left( \frac{O(D)}{B} \right) \left( \frac{(1-c)O(D)}{B} - D \right) \right)$$

เมื่อ  $D = \frac{b - \theta_m}{\sqrt{\frac{1}{a^2} + \sigma_m^2}}$

และ  $B = c + (1+c) \times A(-D)$

กรณีที่ตอบผิด

$$\theta_{m+1} = \theta_m - \left( \frac{\sigma_m^2}{\sqrt{\frac{1}{a^2} + \sigma_m^2}} \right) \left( \frac{O(D)}{A(D)} \right)$$

$$\sigma_{m+1}^2 = \sigma_m^2 \left( 1 - \left( \frac{O(D)}{1 + \frac{1}{a^2 \sigma_m^2}} \right) \left( \frac{O(D)}{A(D)} + D \right) \div A(D) \right)$$

$\theta_m$  แทน ความสามารถของผู้สอบที่ประมาณค่าได้ก่อนตอบข้อสอบข้อที่ m+1 ซึ่งตามปกติแล้วถ้าไม่ทราบค่าความสามารถเบื้องต้นของผู้สอบมักกำหนดให้เท่ากับ 0

$\sigma_m^2$  แทน ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบก่อนตอบข้อที่ m+1 ถ้าไม่ทราบค่าความแปรปรวนดังกล่าวมาก่อนมักกำหนดให้เท่ากับ 1

$\theta_{m+1}$  แทน ค่าความสามารถของผู้สอบโดยประมาณหลังจากที่ตอบข้อที่ m+1

$\sigma_{m+1}^2$  แทน ค่าความแปรปรวนในการประมาณค่าความสามารถของผู้สอบเมื่อ

ตอบข้อสอบข้อที่  $m+1$

- a แทน พารามิเตอร์ค่าอำนาจจำแนกข้อสอบข้อที่  $m+1$
- b แทน พารามิเตอร์ค่า ความยากของข้อสอบข้อที่  $m+1$
- c แทน พารามิเตอร์ระดับโอกาสการเดาข้อสอบได้ถูก ข้อที่  $m+1$
- D แทน จุดบนแกน X
- $O(D)$  แทน ค่าออร์ดิเนต (ordinate) ของโค้งปกติที่จุด D
- $A(D)$  แทน พื้นที่ใต้โค้งปกติจากค่า D จนถึงจุด D

2.2) การให้คะแนนโดยวิธีการประมาณค่าความเป็นไปได้สูงสุด (maximum likelihood) มีหลายวิธี แต่วิธีที่นิยมใช้มาก คือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (conditional maximum likelihood) การประมาณค่า โดยวิธีนี้มีข้อจำกัด คือ ถ้าผู้สอบตอบข้อสอบถูกหมดหรือ ผิดหมด จะไม่สามารถประมาณค่าได้ ขั้นตอนในการให้คะแนนมีดังนี้ (Hambleton and Swanminathan, 1985 อ้างถึงใน ทศน์ศิริรินทร์ สว่างบุญ, 2554; ศิริชัย กาญจนวาสี, 2555)

ขั้นที่ 1 ประมาณค่าความสามารถเริ่มต้น ( $\theta_m = 0; m = 0$ ) โดยใช้สูตรดังนี้

$$\theta_0 = \ln \left( \frac{r_a}{k - r_a} \right)$$

เมื่อ  $r_a = \sum a_i U_i$

$U_i = 1$  เมื่อตอบข้อสอบถูก

$U_i = 0$  เมื่อตอบข้อสอบผิด

$a_i$  คือ อำนาจจำแนกของข้อสอบข้อที่ i

K คือ จำนวนข้อสอบทั้งหมดที่ผู้สอบตอบ

ขั้นที่ 2 หาค่า  $P_1(\theta_m)$  และ  $Q(\theta_m)$  โดยใช้สูตร ดังนี้

$$P_1(\theta_m) = c_i + (1 - c_i) \frac{e^{Da_1(\theta_m - b_1)}}{Da_i(\theta_m - b_i)}$$

$$Q(\theta_m) = 1 - P_1(\theta_m)$$

$$= \frac{1 - c_i}{1 + e^{Da_1(\theta_m - b_1)}}$$

เมื่อ  $\theta_m$  คือ ความสามารถของผู้สอบที่ประมาณค่าได้ในครั้งที่ m

$a_1$  คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

$b_i$	คือ	ค่าความยากของข้อสอบข้อที่ i
$c_i$	คือ	ค่าการเดาของข้อสอบข้อที่ i
D	คือ	ค่าคงที่ ซึ่งมีค่าเท่ากับ 1.7
e	คือ	ค่าคงที่ ซึ่งมีค่าเท่ากับ 2.7182

ขั้นที่ 3 หาค่าปรับแก้ ( $h_m$ ) โดยใช้สูตร

$$h_m = \frac{D[r_m - \sum P_i(\theta_m)]}{-D^2 \sum P_i(\theta_m) Q_i(\theta_m)}$$

ขั้นที่ 4 ประเมินค่าความสามารถของผู้สอบใหม่ ( $\theta_m + 1$ ) โดยใช้สูตร

$$\theta_m + 1 = \theta_m - h_m$$

ขั้นที่ 5 ทำการคำนวณซ้ำในขั้นที่ 2, 3 และ 4 จนกระทั่ง  $h_m$  เข้าใกล้ศูนย์ ( $h_m < 0.001$ )

### 3) วิธีการคัดเลือกข้อสอบในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ

ประสิทธิภาพของแบบทดสอบแบบปรับเหมาะแบบรายข้อพิจารณาจากแบบทดสอบแบบปรับเหมาะที่ให้สารสนเทศเกี่ยวกับความสามารถของผู้สอบมากที่สุดโดยใช้ข้อสอบจำนวนน้อย ดังนั้น การเลือกใช้ข้อสอบแต่ละข้อจึงมีความสำคัญต่อการทดสอบ ข้อสอบจะให้สารสนเทศสูงสุดเมื่อมีค่าความยาก (b) ใกล้เคียงค่าความสามารถของผู้สอบมากที่สุด ค่าอำนาจจำแนก (a) สูง ค่าการเดา (c) เข้าใกล้ศูนย์ และเลือกข้อสอบคุณสมบัติดังกล่าวจัดให้แก่ผู้สอบ ในขณะเดียวกันการเลือกข้อสอบต้องคำนึงถึงการป้องกันการใช้ข้อสอบข้อใดข้อหนึ่งมากเกินไปและเพิ่มการใช้ข้อสอบที่ใช้บ่อยหรือไม่ได้ใช้เลย รักษาสมดุลเนื้อหา ร่วมกับการให้ประสิทธิภาพ การประมาณค่าความสามารถของผู้สอบไปพร้อมๆ กัน

เดวิส และดอร์ต (Davis and Dodd, 2003 อ้างถึงใน สุภารัตน์ ทวลมุกดา, 2550) สรุปจุดมุ่งหมายการเลือกข้อสอบว่า เพื่อให้ความแม่นยำในการประมาณค่าความสามารถของผู้สอบมากที่สุด การรักษาความปลอดภัยของข้อสอบ และรักษาสมดุลเนื้อหาแบบทดสอบพร้อมกันด้วย

วิธีการเลือกข้อสอบแต่ละวิธีให้คุณลักษณะแตกต่างกัน ได้แก่

การเลือกข้อสอบโดยใช้วิธีการสุ่มเลือกข้อสอบ (Random Item Selection) เนื่องจากวิธีการสุ่มเลือกเปิดโอกาสให้ข้อสอบทุกข้อมีโอกาสถูกเลือกเท่าเทียมกัน วิธีนี้จึงช่วยกระจายการใช้ข้อสอบในคลังข้อสอบ แต่ให้ผลลดประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและไม่ให้สมดุลเนื้อหา

การเลือกข้อสอบโดยพิจารณาจากฟังก์ชันสารสนเทศ (Information Function) วิธีนี้สามารถให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูง แต่ข้อสอบที่ให้สารสนเทศสูงมักนำมาใช้บ่อยเกินไป ในขณะที่ข้อสอบบางข้อเลือกใช้น้อยหรือไม่ได้เลือกเลย ข้อสอบใช้มากเกินไปทำให้ผู้สอบเกิดการเรียนรู้หรือข้อสอบเป็นที่รู้จักกันทั่วไปในกลุ่มผู้สอบ เสี่ยงต่อความปลอดภัยของข้อสอบและสูญเสียคุณสมบัติทางจิตวิทยาการทดสอบจากความเที่ยงตรงของข้อสอบลดลง

วิธีการเลือกข้อสอบที่นิยมแพร่หลาย คือการเลือกข้อสอบวิธีสารสนเทศสูงสุด (Maximum Information Item Selection) และ การเลือกข้อสอบวิธีเบย์เซียน (Bayesian Item Selection) มีรายละเอียดดังนี้

1. การเลือกข้อสอบวิธีสารสนเทศสูงสุด (Maximum Information Item Selection) วิธีสารสนเทศสูงสุดจะเลือกข้อสอบที่ให้สารสนเทศสูงสุด ตามค่าประมาณความสามารถของผู้สอบครั้งหลังสุด เมื่อสารสนเทศแบบทดสอบสูงขึ้น ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถจะลดลง ทำให้ค่าประมาณความสามารถเข้าใกล้ค่าความสามารถจริงของผู้สอบมากที่สุด การเลือกข้อสอบสารสนเทศสูงสุดในการทดสอบปรับเหมาะมีความแม่นยำเทียบได้กับการทดสอบ ที่ใช้กระดาษและดินสอแต่ใช้จำนวนข้อสอบน้อยกว่า 60 เปอร์เซนต์ เปรียบเทียบระหว่างวิธีสารสนเทศ สูงสุดกับวิธีการสุ่มเลือกข้อสอบ พบว่า วิธีการสุ่มเลือกข้อสอบใช้ข้อสอบมากกว่า 20 เปอร์เซนต์ และ ยังให้ความถูกต้องในการประมาณค่าความสามารถของผู้สอบน้อยกว่าวิธีสารสนเทศสูงสุด (Eggen & Straetmans, 2000)

ปัจจุบันการเลือกข้อสอบวิธีสารสนเทศสูงสุดประยุกต์ใช้กับการเลือกข้อสอบวิธีอื่นๆ เพื่อเพิ่มประสิทธิภาพในการเลือกข้อสอบ ได้แก่ การเลือกข้อสอบโดยใช้ค่าสารสนเทศโดยรวมเฉลี่ย (Kullback-Leibler Information : KLI) การเลือกข้อสอบด้วยวิธีเกณฑ์ข้างเคียงใกล้ที่สุด (Nearest-Neighbors Criterion) (P. E. Cheng & Liou, 2003) วิธีเกณฑ์ข้างเคียงใกล้ที่สุดเป็นการรวมวิธีสารสนเทศสูงสุดของพิชเชอร์และวิธีค่าความยากข้อสอบเหมาะสม (Optimal Item Difficulty) เข้าไว้ด้วยกัน นอกจากนี้ การศึกษาวิธีเลือกข้อสอบจากสารสนเทศเฉลี่ยสูงสุด นำไปใช้ผ่อนคลายค่าประมาณความสามารถระหว่างการเลือกข้อสอบ เช่น เกณฑ์สารสนเทศตามน้ำหนักทั่วไป (General Weight Information Criterion) เกณฑ์สารสนเทศช่วงชั้น (Interval Information Criterion) และ เกณฑ์สารสนเทศตามน้ำหนักเป็นไปได้ (Likelihood Weight Information Criterion)

2. การเลือกข้อสอบวิธีเบย์เซียน (Bayesian Item Selection) วิธีนี้เลือกข้อสอบที่ให้ความแปรปรวนของการประมาณค่าความสามารถของผู้สอบน้อยที่สุด การประยุกต์วิธีเบย์เซียนสำหรับใช้ในการเลือกข้อสอบ พบได้ในงานวิจัยของ Van der Linden (1999) [ได้แก่ Maximum Expected (Posterior Weighted) Information] การเลือกข้อสอบทั้งวิธีสารสนเทศสูงสุดและ

วิธีการเลือกข้อสอบวิธีเบเซียน สามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยได้ ทั้ง 2 วิธีให้ผลการเลือกข้อสอบใกล้เคียงกันเมื่อความยาวแบบทดสอบมากเพียงพอ การทดสอบปรับเหมาะสามารถใช้วิธีสารสนเทศสูงสุดและวิธีเบเซียนในการทดสอบเดียวกัน พบในผลงานวิจัยของ Van der Linden (1999)

#### 4) การควบคุมการใช้ข้อสอบซ้ำในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ

การควบคุมการใช้ข้อสอบซ้ำเป็นปัญหาที่สำคัญของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ เนื่องมาจากการใช้ข้อสอบซ้ำเกินไปก่อให้เกิดปัญหาในการวัดและประเมิน เช่น กลุ่มของข้อสอบที่มีค่าพารามิเตอร์อำนาจจำแนกสูงๆ เมื่อนำออกมาใช้บ่อยจนรู้จักแพร่หลายในกลุ่มผู้สอบจะนำไปสู่ปัญหาในเรื่องความปลอดภัยของแบบสอบ (Test Security) ซึ่งเกิดจากการร่วมใช้ข้อสอบ (Item Sharing) ระหว่างผู้สอบ ผู้สอบจะได้รับข้อมูลของแบบสอบจากผู้ที่เคยสอบก่อนหน้าและผู้สามารถ สอบตอบคำถามข้อนั้นได้อย่างง่ายดายโดยการล่วงรู้ข้อสอบก่อน (Item Pre-Knowledge) โดยไม่ต้องใช้ความสามารถของตนเอง ถ้าปัญหาดังกล่าวไม่ได้รับการแก้ไข จะส่งผลกระทบต่อความปลอดภัยของข้อสอบและประสิทธิภาพการใช้ข้อสอบในคลังข้อสอบ (Pool Utilization) ตามมา นักทดสอบทั้งทางด้านจิตวิทยาและทางด้านการศึกษาหลาย ท่านจึงให้ข้อเสนอว่า เพื่อให้เกิดความเหมาะสมในทางปฏิบัติการทดสอบปรับเหมาะด้วย คอมพิวเตอร์ควรให้ความสำคัญในเรื่อง การควบคุมการใช้ข้อสอบซ้ำ (Item Exposure Control) และ การสร้างความสมดุลของเนื้อหา (Content Balancing) (Chang & Zhang, 2002; S.-Y. Chen et al., 2003; S.-Y. Chen & Doong, 2008; S. Y. Chen & Macredie, 2005; P. E. Cheng & Liou, 2003)

โดยสามารถแบ่งวิธีการควบคุมการใช้ข้อสอบซ้ำ ดังนี้

1. กลุ่มที่ใช้กระบวนการสุ่มเข้ามาช่วยควบคุมการใช้ข้อสอบซ้ำในขั้นตอนวิธีการคัดเลือกข้อสอบ(Randomization) วิธีนี้การเลือกกลุ่มข้อสอบที่เหมาะสมที่สุดจากชุดข้อสอบ แทนการเลือกข้อสอบสารสนเทศสูงสุดเพียงข้อเดียว และสุ่มเลือกข้อสอบข้อหนึ่งจากกลุ่มข้อสอบจัดให้กับผู้สอบ วิธีการเลือกอย่างสุ่มไม่รับรองว่าการใช้ข้อสอบจะควบคุมได้ตามเกณฑ์ที่ต้องการ ตัวอย่างวิธีนี้ เช่น

- วิธี 5-4-3-2-1 (5-4-3-2-1 Procedure) เสนอโดย McBride & Martin อ้างถึงใน Elissavet et al. (2007) วิธีนี้เน้นการเลือกข้อสอบในระยะเริ่มต้นของการทดสอบ เมื่อผู้สอบมีแนวโน้มว่าเริ่มต้นการทดสอบที่ระดับความสามารถเท่าเทียมกัน (Stocking & Lewis, 2002) เริ่มจากการเลือกข้อสอบสารสนเทศสูงสุดมา 5 ข้อ ข้อสอบข้อแรกที่จัดให้สุ่มเลือกจากข้อสอบสารสนเทศสูงสุด 5 อันดับ ข้อสอบข้อที่ 2 สุ่มเลือกจากข้อสอบสารสนเทศสูงสุดที่เหลืออีก 4 อันดับ และดำเนินการไปเรื่อยๆ จนถึงข้อที่ 5 ข้อดีของวิธีนี้คือทำได้ง่าย แต่มีข้อจำกัดบางประการ คือ อัตราการใช้ข้อสอบซ้ำสูงโดยเฉพาะกลุ่มข้อสอบสารสนเทศสูงและไม่รับรองว่าการใช้ข้อสอบจะควบคุมได้ตามเกณฑ์ที่ต้องการ

- วิธีสุ่มเลือกจากกลุ่มข้อสอบ (Randomesque Procedure) เสนอโดย Kingsbury & Zara อ้างถึงใน Elissavet et al. (2007) วิธีนี้จะสุ่มเลือกข้อสอบเพียงข้อเดียวจากกลุ่มข้อสอบสารสนเทศสูงสุดจำนวนเท่าๆ กัน

- วิธีสุ่มเลือกแบบก้าวหน้า (Progressive Procedure) เสนอโดย Revuelta & Ponsoda อ้างถึงใน Elissavet et al. (2007) เป็นวิธีรวมการสุ่มเลือกและการเลือกจากสารสนเทศสูงสุดเข้าไว้ด้วยกันในระยะแรกใช้วิธีการสุ่มเลือกมากกว่าการเลือกจากสารสนเทศสูงสุด แต่เมื่อการทดสอบดำเนินต่อไป สารสนเทศจะมีผลต่อการเลือกข้อสอบมากกว่าวิธีการสุ่มเลือกเพื่อควบคุมโอกาสที่ข้อสอบจะถูกนำ ออกมาแสดงมากเกินไป (Conditional Selection)

2. กลุ่มที่ใช้ขั้นตอนวิธีการคัดเลือกข้อสอบอย่างมีเงื่อนไข วิธีการเลือกอย่างมีเงื่อนไขเป็นวิธี ควบคุมความน่าจะเป็นของข้อสอบที่ได้รับการเลือกแล้วให้มีการใช้ข้อสอบเป็นไปตามกำหนด วิธีนี้รับรองว่าอัตราการใช้ข้อสอบสูงสุดเป็นไปตามกำหนด แต่ขั้นตอนมีความซับซ้อน และต้องศึกษาในสถานการณ์จำลองในด้านจำนวนครั้งสำหรับการใช้ปฏิบัติ เมื่อองค์ประกอบต่างๆในการทดสอบเปลี่ยนแปลงไปจำเป็นต้องศึกษาในสถานการณ์จำลองก่อนใช้ในสถานการณ์จริง มีวิธีต่างๆ ได้แก่

- วิธี Simpson-Hetter: SH เสนอโดย Simpson & Hetter อ้างถึงใน Elissavet et al. (2007) เป็นที่รู้จักกันแพร่หลาย ดำเนินการโดยใช้ความถี่ของข้อสอบที่จัดให้กับกลุ่มผู้สอบจำนวนมากในสถานการณ์จำลองเปรียบเทียบกับอัตราการใช้ข้อสอบเป้าหมายเพื่อหาพารามิเตอร์การใช้ข้อสอบ ขบวนการนี้ทำซ้ำ ๆ กันจนพารามิเตอร์การใช้ข้อสอบแต่ละข้อมีค่าระหว่าง 0 ถึง 1 พารามิเตอร์นี้ใช้สำหรับการเลือกข้อสอบในการทดสอบสถานการณ์จริงโดยสร้างตัวเลขสุ่มจากการแจกแจงเหมือนกัน (0 ถึง 1)เปรียบเทียบกับพารามิเตอร์การใช้ข้อสอบซ้ำ ถ้าพารามิเตอร์การใช้ข้อสอบซ้ำมากกว่าตัวเลขสุ่มจะจัด ข้อสอบให้ แต่ถ้าพารามิเตอร์การใช้ข้อสอบซ้ำน้อยกว่าตัวเลขสุ่ม จะทำการเลือกข้อสอบข้อใหม่แทน

- วิธี Simpson-Hetter แบบมีเงื่อนไข (Conditional Simpson-Hetter) เป็นวิธีที่พัฒนา จากวิธี Simpson-Hetter ที่เสนอโดย Stocking & Lewis อ้างถึงใน Elissavet et al. (2007) พารามิเตอร์การใช้ข้อสอบซ้ำได้จากการประมาณระดับค่าความสามารถแทนการแจกแจงค่าความสามารถของผู้สอบระหว่างขั้นตอนสถานการณ์จำลอง จะมีการสร้าง  $n \times m$  เมทริกซ์ เมื่อแถว  $n$  คือจำนวนข้อสอบในคลังข้อสอบ และแถวตั้ง  $m$  คือ จำนวนจุดค่าความสามารถที่กระจายบนช่วงการแจกแจงค่าความสามารถ ดังนั้น ข้อสอบแต่ละข้อจะมีพารามิเตอร์การใช้ข้อสอบซ้ำที่แต่ละจุดค่าความสามารถของผู้สอบ

3. กลุ่มที่ใช้กลวิธีในการแบ่งคลังข้อสอบออกเป็นชั้นๆ วิธีการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก (a-Stratified Method: a-STR) วิธีนี้เสนอโดย (Chang & Ying, 1996) ประยุกต์มาจากรูปแบบปรับระดับขั้นของ Weiss เป็นวิธีการเลือกข้อสอบ

และมีประสิทธิภาพในการควบคุมการใช้ข้อสอบซ้ำโดยอัตโนมัติ เริ่มจากการแบ่งชั้นคลัง ข้อสอบตาม ค่าอำนาจจำแนกของข้อสอบ และแบ่งแบบทดสอบเข้าในแต่ละชั้นคลังข้อสอบ ชั้นแรก บรรจุข้อสอบ อำนาจจำแนกต่ำสุดชั้นต่อไปบรรจุข้อสอบอำนาจจำแนกสูงกว่าชั้นแรก ทำเช่นนี้เรื่อยๆ จนกระทั่งชั้น สุดท้ายบรรจุข้อสอบค่าอำนาจจำแนกสูงสุด การทดสอบจะดำเนินจากชั้นแรกจนถึงชั้น สุดท้าย วิธีนี้ กระจายการใช้ข้อสอบซ้ำในคลังข้อสอบให้เท่าเทียมกัน

#### 5) เกณฑ์การยุติการทดสอบ

ในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ จะใช้เกณฑ์การยุติ 2 วิธี คือ ความยาวคงที่ (Fixed-Length) เป็นการกำหนดจำนวนข้อสอบที่ใช้ในการทดสอบให้คงที่ สำหรับ ทุกคน และ ความยาวแปรผัน เป็นการกำหนดความยาวของข้อสอบแบบไม่คงที่ สำหรับผู้สอบแต่ละ คน โดยจะผันแปรไปตามความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ (Standard Error: SEE( $\theta$ )) การทดสอบจะดำเนินไปจนกว่าการประมาณค่า ความสามารถ ( $\theta$ ) ของผู้สอบมีความคลาดเคลื่อนมาตรฐานลดต่ำลงจนถึงระดับที่ยอมรับได้ การทดสอบจึงยุติลง

#### 2.1.2 จุดแข็งและจุดอ่อนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อมีจุดแข็งและจุดอ่อน ดังรายละเอียด ตาราง 2.3

**ตาราง 2.3 จุดแข็งและจุดอ่อนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ**

จุดแข็ง	จุดอ่อน
- การทดสอบที่ผู้เข้ารับการทดสอบได้รับการจัดข้อสอบที่มี ความยากตามระดับความสามารถของตน	- อาจเกิดการประมาณค่าความสามารถที่เกินจริง (over estimate) และต่ำกว่าจริง (under estimate)
- จำนวนข้อสอบที่ใช้ในการทดสอบมีจำนวนลดลงโดยที่ไม่ สูญเสียความแม่นยำในการประมาณค่าความสามารถของผู้เข้า รับการทดสอบ	- ผู้เข้ารับการทดสอบไม่สามารถกลับไปแก้ไข ข้อสอบก่อนหน้าได้
- ขั้นตอนการเลือกข้อสอบจะเลือกข้อสอบแต่ละข้อภายหลัง การประมาณค่าความสามารถของผู้สอบจากคลังข้อสอบด้วย กระบวนการทางคอมพิวเตอร์ โดยอยู่บนพื้นฐานของผลการ ตอบข้อสอบจากข้อก่อนหน้า	

ที่มา: Zheng and Chang (2015); ฅณภัทร ชัยมงคล (2558); ทศนศิริรินทร์ สว่างบุญ ศิริเดช สุชีวะ ศิริชัย กาญจนวาสิ และ Muraki (2555)



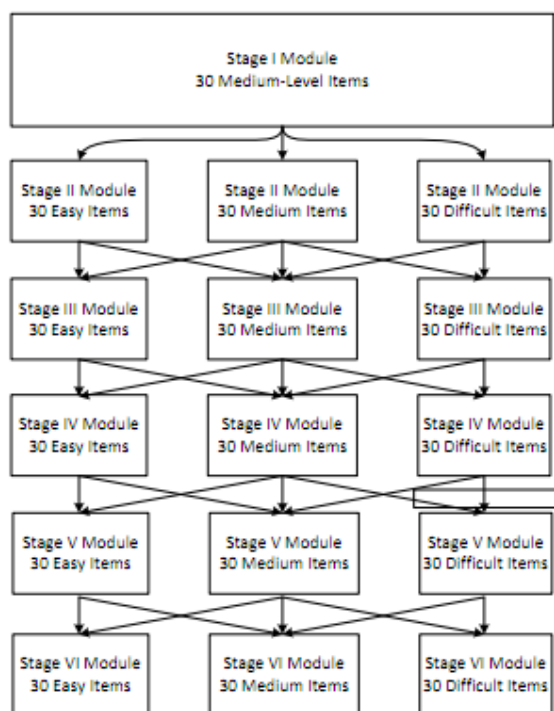
## 2.2 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (Multi-Stage Test Design)

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (Multi-Stage Test Design: MST) จะเป็นชุดของข้อสอบที่ถูกเลือกตามความสามารถของผู้สอบก่อนที่จะทำข้อสอบชุดนั้น ซึ่งกลุ่มของรูปแบบหลัง ๆ จะมีลักษณะคล้ายกับการทดสอบเส้นตรงแบบดั้งเดิมในเรื่องของรูปแบบของแบบสอบในท้ายที่สุดจะถูกบริหารจัดการจากการบริหารการสอบก่อนหน้า การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ถูกใช้ในทางปฏิบัติเพื่อควบคุมคุณภาพของความสามารถ ซึ่งเหมือนกันกับรูปแบบเส้นตรงในทุก ๆ รายการทดสอบจนในท้ายที่สุดจะถูกบริหารจัดการเป็นแบบเฉพาะในขั้นสูง วิธีการทดสอบแบบหลายขั้นตอน อนุญาตให้แต่ละรายการทดสอบสามารถพบทวนรายละเอียดด้านสถิติ เนื้อหา และอื่น ๆ เช่น ส่วนประกอบของกราฟ ภาพและเสียง ก่อนที่จะบริหารจัดการสอบ การอนุญาตนี้เพื่อควบคุมคุณภาพการติดตามที่สูงกว่าวิธีการที่ใช้ในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) ทั่วไป ดังนั้นจึงเป็นการเพิ่มประโยชน์ขององค์ประกอบด้านการปรับเหมาะเมื่อเปรียบเทียบกับวิธีเส้นตรงแบบดั้งเดิม (Zenisky, 2004 อ้างถึงใน Armstrong & Little, 2003; Brossman & Guille, 2014; Guille et al., 2011; Luecht & Sireci, 2011)

การดำเนินการของการทดสอบแบบหลายขั้นตอน นั้นจะใช้ข้อสอบชุดแรกเป็นกลุ่ม ๆ หรือเรียกว่า โมดูล (Modules) ในแต่ละโมดูลจะมีรายละเอียดของค่าสถิติและเนื้อหา ซึ่งโมดูลของการทดสอบแบบหลายขั้นตอน จะถูกสร้างเป็นรูปแบบบล็อก โดยทั่วไปในแต่ละโมดูลจะสร้างตามระดับของเนื้อหาจับคู่กับรายการเนื้อหาในแบบสอบทั้งฉบับ ในส่วนของรายละเอียดด้านสถิตินั้นตั้งใจที่จะเปลี่ยนข้ามกันระหว่างโมดูล เช่น ความแตกต่างของแต่ละโมดูลจะเกี่ยวข้องกับระดับความยากง่ายของข้อสอบ ยกตัวอย่างเช่น การทดสอบใช้วิธีการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ซึ่งสร้างโมดูลยาก สำหรับผู้สอบที่มีความสามารถสูง โมดูลง่าย สำหรับผู้สอบที่ไม่สามารถทำข้อสอบได้ดีและโมดูลปานกลางสำหรับผู้สอบที่ทำข้อสอบได้ไม่มากหรือน้อยเกินไป หลังจากแต่ละโมดูลถูกสร้างขึ้น นักพัฒนาแบบสอบจะพบทวนว่าโมดูลเหล่านั้นมีเนื้อหาและค่าสถิติตามนั้นหรือไม่

โดยทั่วไประหว่างการบริหารการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ผู้สอบทั้งหมดจะใช้โมดูลเหมือนกันซึ่งจะใช้โมดูลที่มีความยากของข้อสอบในระดับปานกลางบนพื้นฐานของการตอบข้อสอบในโมดูลแรกนี้ ผู้สอบแต่ละคนจะถูกประมาณค่าคะแนนความสามารถ ดังนั้น จะอยู่บนฐานของกฎการตัดสินใจเฉพาะเจาะจงก่อนที่ผู้สอบจะถูกใช้ในโมดูลที่ 2 ต่อไปซึ่งประกอบด้วยข้อสอบที่ยาก ปานกลาง หรือง่าย ซึ่งได้จากการประมาณค่าความสามารถในขั้นตอนที่ 1 ถ้าแบบสอบประกอบด้วยขั้นตอนเพียง 2 ขั้นตอน การทดสอบก็จะยุติหลังจากขั้นตอนที่ 2 อย่างไรก็ตามการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) นั้น โครงสร้างมักจะประกอบด้วยมากกว่า 2 ขั้นตอน (Zenisky, 2004) ในสถานการณ์นี้ หลังจากขั้นตอนที่ 2 แล้ว

คะแนนความสามารถของผู้สอบแต่ละคนจะถูกประมาณค่าซ้ำบนพื้นฐานการตอบทั้งหมด ทั้งขั้นตอนที่ 1 และ 2 แล้วผู้สอบก็จะข้ามไปในขั้นตอนที่ 3 ซึ่งมีทั้งข้อสอบยาก ปานกลาง หรือง่าย ขึ้นอยู่กับการประมาณค่าความสามารถของผู้สอบก่อนหน้านี้ กระบวนการนี้จะทำต่อเนื่องไปจนกระทั่งครบหมดทุกขั้นตอน หลังจากนั้น การทดสอบก็จะสิ้นสุด คะแนนความสามารถของผู้สอบแต่ละคนจะถูกประมาณค่าขึ้นอยู่กับผลการตอบข้อสอบทั้งหมด

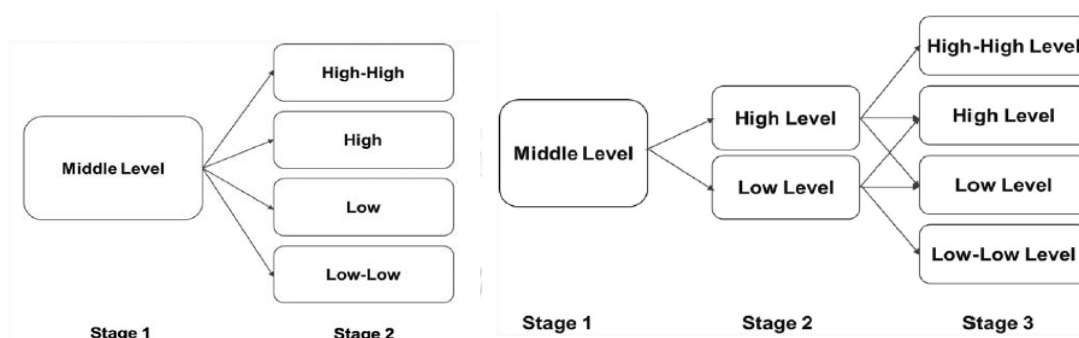


ภาพ 2.3 ตัวอย่างขั้นตอนและกฎเส้นทางของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) (Brossman & Guille, 2014)

นอกจากนี้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ยังสามารถกล่าวได้ว่ามีประสิทธิภาพมากสำหรับการทดสอบปรับเหมาะเพื่อการการทดสอบเชิงจำแนก เมื่อเปรียบเทียบกับ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) รูปแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มีองค์ประกอบที่เพิ่มความหลากหลาย เช่น กลุ่มโมดูล ขั้นตอน และเส้นทาง (Luecht, 2000 อ้างถึงใน Park, Kim, Chung, & Dodd, 2014) ซึ่งโมดูลเป็นการเก็บรวบรวมข้อสอบที่เป็นหน่วยที่เล็กที่สุดของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST)

โดยธรรมชาติของการปรับเหมาะของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จะเกิดขึ้นหลังจากผู้สอบตอบข้อสอบทั้งหมดในโมดูลนั้น ข้อสอบภายในโมดูลถูกควบคุมด้วยคุณลักษณะค่าสถิติที่เหมือนกัน แม้ว่าในส่วนหนึ่งของเนื้อหาและคุณลักษณะที่ไม่ใช่ค่าสถิติจะ

แปรเปลี่ยน ขั้นตอนต่อมาของหน่วยในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ แบบหลายขั้นตอน (MST) จะยังคงใช้โมดูลหนึ่งหรือมากกว่าได้ เนื่องจากผู้สอบจะทำข้อสอบเพียงหนึ่งโมดูลภายในแต่ละขั้นตอน คุณลักษณะของค่าสถิติของโมดูลแปรเปลี่ยนตามจุดมุ่งหมายของการทดสอบ กลุ่มของหน่วยที่ใหญ่ที่สุดของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และยังมีขั้นตอนที่หลากหลายอยู่ เช่น โครงสร้างกลุ่มแบบ 1-3-3 จะประกอบด้วย 3 ขั้นตอน ดังนี้ ในขั้นตอนที่ 1 จะมี 1 โมดูล ขั้นตอนที่ 2 จะมี 3 โมดูล และขั้นตอนที่ 3 จะมี 3 โมดูล



ภาพ 2.4 ตัวอย่างโมเดล การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST):

2-stage โมเดลการออกแบบ 1 – 4 และ 3-stage โมเดลการออกแบบ 1 – 4

(Wang et al., 2016)

## 2.2.1 จุดแข็งและจุดอ่อนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอนมีจุดแข็งและจุดอ่อน ดังนี้

ตาราง 2.4 จุดแข็งและจุดอ่อนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน

จุดแข็ง	จุดอ่อน
- กระบวนการในการเลือกข้อสอบจะกระทำเฉพาะระหว่างขั้น (stage) เท่านั้นไม่ได้ใช้กระบวนการเลือกภายในขั้น ดังนั้นจึงอนุญาตให้ผู้สอบสามารถย้อนหรือข้ามข้อสอบภายในขั้นที่กำลังดำเนินการทดสอบได้ เพื่อทำการทบทวนหรืออาจเปลี่ยนคำตอบ	ต้องมีการจัดเตรียมชุดข้อสอบ (panel) และโมดูล (module) ของข้อสอบไว้ล่วงหน้า ซึ่งใช้คนในการดำเนินการ ดังนั้นหากเป็นการทดสอบที่มีขนาดใหญ่อาจเกิดความผิดพลาดในการจัดชุดข้อสอบได้
- อนุญาตให้ผู้พัฒนาแบบทดสอบสามารถกำกับติดตามและควบคุมคุณภาพของชุดแบบทดสอบ	
- ในการประมาณค่าความสามารถของผู้สอบเพื่อคัดเลือกชุดข้อสอบที่เหมาะสมกับความสามารถของผู้สอบจะทำการประมาณค่าจากชุดของข้อสอบในขั้นแรก จึงทำให้ลดความผิดพลาดจากการประมาณค่าความสามารถผู้สอบจากข้อสอบข้อแรกเพียงข้อเดียว	

ที่มา: Zheng and Chang (2015); Zheng et al. (2012)

### 2.2.3 งานวิจัยที่เกี่ยวข้องกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (Multi-Stage Test Design)

Brossman and Guille (2014) ได้ศึกษาวิจัยเรื่องการเปรียบเทียบรูปแบบการทดสอบแบบหลายขั้นตอนและแบบเส้นตรงสำหรับการสอบขนาดกลางเพื่อรับใบอนุญาตและการรับรอง ส่วนใหญ่ การทดสอบในองค์กรขนาดใหญ่ ปัจจุบันใช้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) การออกแบบและการใช้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ในการทดสอบแต่ละขั้นนั้น รูปแบบหนึ่งเป็นการบริหารจัดการในแต่ละขั้นตอน ขั้นตอนที่ต้องเนื่องในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จะต้องมีความหลากหลายในเรื่องของความยากง่ายตามการประมาณค่าความสามารถของผู้สอบ แม้ว่า การศึกษาหลายครั้งมีการเปรียบเทียบความสามารถของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) กับรูปแบบการทดสอบแบบดั้งเดิม ผลการวิจัยชี้ว่าการศึกษาในสถานการณ์จำลองนั้นจำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่เพื่อลดจำนวนข้อมูลจำลอง ตามผลการวิจัย ถึงแม้ว่า สถิติจะชี้ให้เห็นว่าการประมาณค่ามีความน่าเชื่อถือ แต่การศึกษานี้ก็ให้ผลที่เหมาะสมกว่าในความสามารถของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) สำหรับการสอบขนาดใหญ่มื่อเทียบกับการสอบขนาดเล็กหรือขนาดกลาง จุดประสงค์ของการวิจัยนี้เป็นการศึกษาในข้อมูลจำลองสำหรับการสอบขนาดกลางเพื่อใบรับรองทางการแพทย์ ผลการวิจัยยืนยันว่า การประมาณค่าความสามารถที่ถูกต้องและสอดคล้องกับผลการตัดสินการสอบผ่านและตกที่ได้รับจริง สำหรับการทดสอบนี้ภายใต้รูปแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ดีกว่าเมื่อเปรียบเทียบกับรูปแบบดั้งเดิมแบบเส้นตรง

Park et al. (2014) ได้ศึกษาวิจัยเรื่องการเพิ่มประโยชน์ของคลังข้อสอบในการสร้าง การทดสอบหลายขั้นตอนโดยใช้การทดสอบรูปแบบผสม โดยการศึกษาเป็นการศึกษาวิธีการใช้คลัง ข้อสอบใหม่ในการสร้างการทดสอบหลายขั้นตอน (MST) โดยใช้การทดสอบรูปแบบผสมบนพื้นฐาน ของโมเดลการให้คะแนนบางส่วน (GPCM) ซึ่งรูปแบบของคลังข้อสอบแบบผสมนั้นประกอบด้วย ข้อสอบ 3 แบบ ได้แก่ 1) ข้อสอบที่ให้คะแนนแบบ 2 ค่า ถูกได้ 1 คะแนน และผิดได้ 0 คะแนน จำนวน 244 ข้อ 2) ข้อสอบที่ให้คะแนนแบบ 3 ค่า จำนวน 113 ข้อ และ 3) ข้อสอบที่ให้คะแนนแบบ 4 ค่า จำนวน 67 ข้อ ซึ่งวัดทั้งหมด 3 เนื้อหา จำนวน 424 ข้อ โดยแบ่งเป็นเนื้อหาที่ 1 จำนวน 126 ข้อ เนื้อหาที่ 2 จำนวน 148 ข้อ และเนื้อหาที่ 3 จำนวน 150 ข้อ การจำลองการทดสอบปรับเหมาะ ด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ของการทดสอบเชิงจำแนกได้ดำเนินการเพื่อประเมินผล การออกแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) โมเดลการเขียน โปรแกรมเชิงเส้น (linear programming : LP) ถูกนำมาใช้ในการดำเนินการในการทดสอบปรับ เหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ที่ประกอบขึ้นใหม่ ซึ่งขึ้นอยู่กับการสร้าง

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) เริ่มต้น โดยสามขั้นตอนต่อมา การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ที่ประกอบขึ้นใหม่ถูกดำเนินการสำหรับแต่ละหน่วยที่ประกอบใหม่ อัตราการทดแทนของการทดสอบสามหน่วย (TRRs; 0.22, 0.44, และ 0.66) ได้รับการตรวจสอบ เงื่อนไขของอัตราการส่งผ่านทั้งสามหน่วย (30%, 50% และ 70%) มีการพิจารณาในการทดสอบเชิงจำแนก

โดยการวิเคราะห์ข้อมูลจะพิจารณาจากเกณฑ์การตัดสินใจ ซึ่งแบ่งออกเป็นอัตราการจำแนกที่ถูกต้อง (correct classification rate : CCR) อัตราความคลาดเคลื่อนของการตัดสินใจผิดพลาดทางบวก (false positive error rate: FPER) อัตราความคลาดเคลื่อนของการตัดสินใจผิดพลาดทางลบ (false negative error rate: FNER) และอัตราความคลาดเคลื่อนรวมทั้งหมด (total error rate: TER) สำหรับในส่วนของการใช้ประโยชน์ของคลังข้อสอบ จะดูที่การเพิ่มขึ้นของการใช้ประโยชน์ของคลังข้อสอบและอัตราการใช้ประโยชน์ของคลังข้อสอบสะสมในแต่ละการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) รวมถึงการตรวจสอบการเปิดเผยของแบบสอบย่อยหลังจากทำขั้นตอนที่ 3 ของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) แล้ว ซึ่งจะชี้ให้เห็นถึงรูปแบบของการเปิดเผยของแบบสอบย่อยอย่างต่อเนื่องใน การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) โดยผลการวิจัยแสดงให้เห็นว่าเงื่อนไขต่าง ๆ ในองค์ประกอบใหม่ของ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) โดยรวมเพิ่มอัตราการใช้คลังข้อสอบ ในขณะที่การสร้างการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ที่ต้องการยังคงอยู่ เงื่อนไขการทดสอบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ทั้งหมดได้ดำเนินการอย่างเท่าเทียมกันทั้งในแง่ของความแม่นยำของการตัดสินใจเชิงจำแนก

## 2.3 การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

จากแนวคิดการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อและการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) พบว่าแต่ละแนวคิดมีจุดเด่นและจุดอ่อนที่แตกต่างกัน ดังนั้น Wang et al. (2016) จึงได้มีการนำแนวคิดการทดสอบทั้งสองวิธีมารวมกันทำให้ได้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดเพื่อเพิ่มประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ

Wang et al. (2016) ได้ทำการจำลองการทดสอบปรับเหมาะด้วยแบบไฮบริด ซึ่งเป็นการรวมรูปแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ สำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด จะประกอบด้วย 2 ส่วน คือการทดสอบในส่วนแรกใช้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และในส่วนต่อไปจะใช้การทดสอบปรับเหมาะด้วย

คอมพิวเตอร์แบบรายข้อ เนื่องจากการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ สามารถทำตามกระบวนการเป็นแบบหลายขั้นตอน และขนาดของโมดูลเท่ากับ 1 เพื่อแก้ไขปัญหาการออกแบบจากการเปลี่ยนขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ไปยังขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ

การทดสอบปรับเหมาะด้วยแบบไฮบริด คือ การรวมกรอบแนวคิดการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ โดยนำวิธีการประมาณค่าความสามารถและวิธีการคัดเลือกข้อสอบในการออกแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) มาใช้ในการทดสอบ เช่น สารสนเทศสูงสุดของฟิชเชอร์ (Maximum Fisher Information: MFI) และการประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood Estimate: MLE) สามารถนำมาใช้ในขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดได้ สำหรับขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จะมี 2 ขั้นตอนหรือ 3 ขั้นตอนนี้ขึ้นอยู่กับข้อเสนอนี้ และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มีการลดความยาวของข้อสอบ โดยมีการศึกษาวิจัยเรื่องความยาวของโมดูลในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ขึ้นอยู่กับขั้นตอนของเส้นทางที่ยาวสามารถให้ประมาณการที่แม่นยำมากขึ้น (Hambleton & Xing, 2006; Jodoin et al., 2006; Patsula, 1999) และในการทดสอบปรับเหมาะด้วยแบบไฮบริดมีการดำเนินการออกแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) โดยมีการกำหนดเส้นทางและลดความยาวโมดูลตอบสนองต่อวัตถุประสงค์ในการประมาณค่าความสามารถที่ถูกต้อง สำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อจะใช้วิธี automatic test assembly (ATA) ในการตรวจสอบ

ปัญหาการออกแบบที่สำคัญที่สุดคือการกำหนดจุดการเปลี่ยนแปลงจาก การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ไปยังการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ เช่น การออกแบบไฮบริด เสนอวิธีแก้ปัญหา ½ ความสัมพันธ์ของความยาวขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ควรจะเป็นครึ่งหนึ่งของความยาวของการทดสอบทั้งหมด หากการทดสอบส่งต่อไปยังขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อจะช้ามาก ขั้นตอนไฮบริดนี้จะไม่มีข้อสอบพอที่จะทำให้การใช้งานคุณสมบัติการปรับเหมาะหลายระดับข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อและอาจทำให้มันมีประสิทธิภาพน้อยกว่าขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ เพื่อการยอมรับว่าการประมาณค่าที่เกิดจากขั้นตอน การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ถูกต้องมากขึ้นเมื่อเทียบกับประมาณค่าเบื้องต้นขึ้นอยู่กับคำตอบต่อจำนวนเดียวกันของข้อสอบที่จัดการสอบโดยการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ และมีความสัมพันธ์ลดลง สำหรับความยาวของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ควรจะ

เลือกตามความต้องการจริง มิฉะนั้นจะส่งผลต่อไปยังส่วนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อเร็วเกินไป ถ้าความยาวภายในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) สั้นเกินไปที่จะส่งผลต่อข้อมูลของระดับความสามารถที่จะนำไปสู่ความไม่แน่นอนต่อการประมาณค่าจากการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด จะคาดคะเนว่าความยาวในการทดสอบได้รับเป็นค่าคงที่  $L_T$  ความยาวสำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และส่วนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ มี  $L_1$  และ  $L_2$  ตามลำดับ สมมติว่าในมาตรฐานการออกแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ ข้อสอบ  $L_0$  จะได้รับเลือกจากความยากระดับกลางในรูปแบบเริ่มต้น

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดซึ่งเป็นการผสมผสานรูปแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ มีรายละเอียดสรุปได้ดังนี้

1. การทดสอบในระบบเริ่มต้นโดยการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) แล้วเปลี่ยนไปเป็นส่วนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ
2. ความสัมพันธ์ระหว่างความยาวของขั้นตอน การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และความยาวของข้อความยาวคงที่เป็น

$$L_0 \leq L_1 \leq \frac{L_T}{2}$$

3. รูปแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) กับการออกแบบ 2 หรือ 3 ขั้นตอน ขึ้นอยู่กับข้อเสนอนี้ และขนาดของโมดูลจะลดลงตามการเพิ่มขึ้นของคำสั่งขั้นตอนในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST)
4. วิธีการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) สามารถใช้วิธีATA ในการประกอบชุด (panel) ขนานกัน

#### **ขั้นตอนการทำงานของ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด**

สำหรับขั้นตอนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด จะเริ่มต้นการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ด้วยชุดข้อสอบที่กำหนดขึ้น (initial item) (MST ขั้นที่ 1) ที่มีความยากระดับปานกลาง จากนั้นจะทำการประมาณค่าความสามารถของผู้สอบจากการตอบข้อสอบของผู้สอบในชุดแรก ระบบจะคัดเลือกข้อสอบชุดถัดไปที่มีความยากใกล้เคียงกับความสามารถของผู้สอบ (MST ขั้นที่ 2) ในขั้นตอนนี้จะมีการแบ่งโมดูลข้อสอบออกเป็น 3 ระดับ ได้แก่ ระดับง่าย ระดับปานกลาง และระดับยาก เมื่อผู้สอบทำชุดข้อสอบตามระดับความสามารถแล้ว ระบบจะประมาณค่าความสามารถของผู้สอบจากการตอบข้อสอบในชุดที่ 2 เพื่อคัดเลือกข้อสอบถัดไปโดยการสุ่มข้อสอบที่ละข้อที่ใกล้เคียงความสามารถของผู้สอบ ซึ่งเป็น

ส่วนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ จนถึงเกณฑ์การยุติการทดสอบ สำหรับกระบวนการเตรียมชุดข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จะมีการจัดเตรียมชุดข้อสอบไว้ล่วงหน้า และการเตรียมข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ จะใช้คอมพิวเตอร์ในการคัดเลือกข้อสอบแต่ละข้อ

Wang et al. (2016) ได้ทำการพัฒนาการทดสอบปรับเหมาะด้วยแบบไฮบริด และนำมาเปรียบเทียบคุณภาพการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ในการศึกษาเป็นการจำลองข้อมูลโดยภายใต้เงื่อนไขที่แตกต่างกัน มีกำหนดการศึกษา ดังนี้ คลังข้อสอบที่มีเนื้อหา 5 กลุ่ม ประกอบด้วยข้อสอบหลายตัวเลือกจำนวน 600 ข้อ โดยเป็นโมเดลโลจิสติก 3 พารามิเตอร์ มีสารสนเทศข้อสอบพารามิเตอร์  $a$ -,  $b$ -,  $c$ - ความสมดุลของเนื้อหามาใช้ในการทดสอบแต่ละครั้งประกอบกับการกระจายเนื้อหาเดียวกันหรือคล้ายกันกับที่คลังข้อสอบ การใช้ข้อสอบซ้ำด้วยอัตราการใช้ข้อสอบซ้ำสูงสุด 0.2

จากการศึกษามีประเด็นที่น่าสนใจ คือ เมื่อทำการเปรียบเทียบการทดสอบปรับเหมาะด้วยแบบไฮบริด กับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ และการทดสอบปรับเหมาะหลายขั้นตอน (MST) เรียกวิธีการนี้ว่า PMCAT โดยมีการทดสอบด้วยการจำลองข้อมูลเริ่มจากการทดสอบเริ่มต้นด้วยการคัดเลือกข้อสอบข้อแรก (initial item) ด้วยวิธีการสุ่มจำนวน 5 ข้อ ที่มีความยากปานกลาง มีการจัดคลังข้อสอบด้วยวิธีการแบ่งชั้น โดยวิธีการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ มีการกำหนดความยาวของข้อสอบในระดับที่สอง จำนวน 20 และ 40 ข้อ มีการกำหนดกลุ่มตัวอย่างผู้เข้าสอบจำนวน 5,000 คน กำหนดค่าความสามารถที่แท้จริงที่  $\theta_s$  โดยการจำลองจาก  $N(0,1)$  และกำหนดค่าความสามารถที่แท้จริง  $\theta_s$  ไว้ในการออกแบบการทดลอง มีการกำหนด การใช้ข้อสอบซ้ำด้วยอัตราการใช้ข้อสอบซ้ำสูงสุด 0.2 และมีการแบ่งชั้นในการควบคุมเนื้อหา ค่าดัชนีก่อนหน้า (PI) การคัดเลือกข้อสอบใช้วิธีการ Fisher information หลังจากจบขั้นตอนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ แล้วจะใช้วิธีการ PMCAT การประมาณค่าความสามารถใช้วิธีการประมาณค่าความเป็นไปได้สูงสุด (maximum likelihood estimate, MLE)

ในขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ของ วิธีการ PMCAT ใช้ยุทธวิธีสองขั้นตอน รูปแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) 1-4 ขั้นตอนแรกในแต่ละโมดูลมีค่าเฉลี่ยความยาก anchored ที่ 0 และระดับค่าความสามารถและค่าความยากในโมดูลสุดท้าย anchored ที่ 4 จุดตัดคะแนน สำหรับการแบ่งชั้นผู้สอบใช้วิธีการเลือกข้อสอบแบบฮิวริสติก (heuristic) หลักการ  $\frac{1}{2}$  สำหรับจุดเปลี่ยนผ่าน ในการทดลองปัจจัยแรกในขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ในวิธีการไฮบริด มีการกำหนดระดับความยาวเฉพาะในส่วนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์



แบบหลายขั้นตอน (MST) 3 ระดับ (8, 10 และ 15) และ 5 ระดับ (10, 15, 20, 25 และ 30) และกำหนดความยาวไว้ที่ 20 และ 40 ตามลำดับ ระดับเหล่านี้เป็นตัวแทนประมาณภายใต้เงื่อนไขที่ความยาวของขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ที่สั้นกว่าเท่ากับ หรือยาวกว่าครึ่งหนึ่งที่สอดคล้องกันความยาวในการทดสอบทั้งหมด ปัจจัยที่สองของขั้นตอนในเส้นทางที่ยาวในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ที่จะส่งผลในการประมาณค่าความสามารถสุดท้ายดีกว่า ในแต่ละความยาวในการทดสอบมีการตั้งเงื่อนไขการทดลองสองประเภท ในขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) เส้นทางขั้นตอนที่ยาวหรือขั้นตอนการกำหนดเส้นทางที่สั้นกว่าโดยเฉพาะเมื่อระยะเวลาในการทดสอบเท่ากับ 20 สองขั้นตอนในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มีการออกแบบเป็น PMCAT (5, 3), (6, 4), และ (10, 5) และเรียงลำดับในสองขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) คือ (3, 5), (4, 6) และ (5, 10) สองตัวเลขในวงเล็บสอดคล้องกับระยะเวลาของขั้นตอนแรกและสองในขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ในทำนองเดียวกันเมื่อความยาวของในการทดสอบเท่ากับ 40, สองขั้นตอนในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ในการออกแบบ PMCAT (6, 4), (10, 5), (12, 8), (15, 10), และ (20, 10) สองขั้นตอนจากการเรียงลำดับในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) (4, 6), (5, 10), (8, 12), (10, 15) และ (10, 20) สำหรับแต่ละเงื่อนไขในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มีการกำหนด 10 คู่ขนาน ขึ้นกับวิธีการแก้ปัญหาจากด้านล่างขึ้นไป

เกณฑ์การประเมินมี 3 เกณฑ์การประเมิน นำมาใช้ในการประเมินความถูกต้องของการประมาณค่าและประสิทธิภาพ: รากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSEs), bias และ อัตราการจำแนกที่ถูกต้อง (correct classification rate : CCR) ทั้งสองดัชนีแรกใช้สำหรับการวัดความถูกต้องและมีประสิทธิภาพ เพราะการทดสอบนี้มีความตั้งใจในการจำแนกความสามารถของผู้สอบเป็นกลุ่มที่แตกต่างกัน ดัชนีอัตราการจำแนกที่ถูกต้อง (CCR) ถูกนำมาใช้ในการวัดการจำลองเป็นการจำลองแบบ 10 ครั้ง และหาค่าเฉลี่ย

ผลลัพธ์และการอภิปราย RMSEs, bias และ CCR สำหรับวิธี PMCAT ภายใต้เงื่อนไขที่แตกต่างกัน เมื่อความยาวในขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) น้อยกว่าครึ่งหนึ่งของความยาวของการทดสอบทั้งหมดและส่วนที่เป็นพื้นฐานการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ ส่วนของผลลัพธ์ค่า RMSEs และ CCR ไม่ว่าจะเป็นขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จากการออกแบบน้อยไปมากหรือมากไปน้อยเกือบทั้งหมด ค่า RMSEs จากวิธี PMCAT มีขนาดเล็กกว่าผู้ที่อยู่ในการออกแบบที่การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ และ bias ความสามารถของวิธี PMCAT และ

การออกแบบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายข้อมีความใกล้เคียงกันมาก เมื่อความยาวของขั้นตอนการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ความยาวครึ่งหนึ่งของความยาวของการทดสอบทั้งหมดที่มีค่าดัชนีจากพื้นฐานการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์ (CAT) ในกรณีนี้วิธี PMCAT มี RMSEs ขนาดใหญ่และ CCR ต่ำกว่าผู้ที่มาจากการออกแบบที่การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายข้อภายใต้เงื่อนไขทั้งหมด ซึ่งสนับสนุนหลักการแก้ปัญหาที่ความยาวของขั้นตอนในการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ใน PMCAT ว่าไม่ควรจะยาวกว่าครึ่งหนึ่งของความยาวของการทดสอบทั้งหมด

จากการศึกษาพบว่า วิธี PMCAT มีประสิทธิภาพมากกว่าการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายข้อ เมื่อความยาวในการทดสอบสั้น ในการออกแบบการปรับเหมาะด้วยคอมพิวเตอร์สามารถปรับปรุง การประมาณค่าความสามารถในการเริ่มต้นซึ่งจะส่งผลกระทบต่อในคัดเลือกข้อสอบที่มีประสิทธิภาพมากขึ้นภายหลังขั้นตอนของการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายข้อ เพื่อดูว่าการออกแบบไฮบริดได้รับการออกแบบที่เหมาะสมสามารถปรับปรุงการประมาณค่าความสามารถในการเริ่มต้น การเปรียบเทียบ RMSEs ของวิธี PMCAT ที่จุดเปลี่ยนแปลงเมื่อความยาวในการทดสอบเท่ากับ 20 ข้อ แสดงให้เห็นว่าเมื่อความยาวของขั้นตอน การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ที่ใช้แก้ปัญหาหลักการอิวิริสติก  $\frac{1}{2}$  RMSEs ที่จุดเปลี่ยนจาก PMCATs มีขนาดเล็กกว่าสอดคล้องกับขั้นตอนการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายข้อ ที่มีจำนวนตัวเลขของการจัดการข้อสอบ แสดงให้เห็นว่าการประมาณค่าความสามารถที่จุดเปลี่ยนใน วิธี PMCAT มีความใกล้เคียงกับค่าที่แท้จริงของตนเมื่อเทียบกับผู้ที่มาจากการออกแบบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายข้อ

และจากการศึกษาเพื่อเปรียบเทียบประสิทธิภาพการทำงานของแบบไฮบริดที่เหมาะสมกับผู้ที่มาจากการทดสอบวิธีเดียวที่จะตรวจสอบคุณสมบัติของคุณสมบัติของการออกแบบของไฮบริด ความยาวการทดสอบคงที่ 40 ข้อ วิธี PMCAT ที่ดีที่สุดในการกำหนดความยาว ขั้นตอนการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ที่มี ข้อสอบ 6 และ 4 ข้อ วิธี PMCAT เป็นวิธีที่ได้รับการเลือกในแบบไฮบริดที่เหมาะสมในการศึกษานี้

สำหรับการเปรียบเทียบวิธีการออกแบบแรกใช้การออกแบบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์ แบบรายข้อ ใช้ในการศึกษาที่ 1 การออกแบบที่สองคือการออกแบบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์ (CAT) ใช้เกณฑ์สารสนเทศ Kullback-Leibler (KL) เป็นวิธีการคัดเลือกข้อสอบที่เรียกว่า KLCAT โดยดัชนี KL มีจุดมุ่งหมายร่วมกันของแบบไฮบริดและวิธี KLCAT คือการปรับปรุง การคัดเลือกข้อสอบในช่วงแรกของการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายข้อ การออกแบบที่สามคือวิธีการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มี two-stage 1-4 โมเดล ขนาดโมดูลสำหรับแต่ละขั้นตอนเป็น 20 การออกแบบที่สี่ วิธีการทดสอบปรับ

เหมาะสมด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มี three-stage 1-2-4 model ขนาดโมดูลของ stage 1 -3 คือ 10, 10, และ 20 ตามลำดับ การดำเนินการสองรูปแบบที่ดีที่สุดคือ two-stage และ three-stage ในแบบเดียวกันกับคลังข้อสอบ เมื่อความยาวของการทดสอบ 40 ข้อ 10 parallel panels ถูกกำหนดขึ้นจากโมดูลโดยผสมกลยุทธ์ล่างขึ้นบนและจากบนลงล่าง

เงื่อนไขการจำลองสำหรับผู้สอบ คือ วิธีการประมาณค่าความสามารถและวิธีการคัดเลือกข้อสอบมีความสอดคล้องกับผู้ที่อยู่ในการศึกษาที่ 1 และ 2 ตาม 4 จุดตัดคะแนน กลุ่มตัวอย่างผู้สอบแบ่งออกเป็น 5 กลุ่มประกอบด้วย กลุ่ม 1 ผู้สอบมีความสามารถต่ำสุดและ กลุ่ม 5 ประกอบด้วยผู้สอบมีความสามารถสูง ผลการศึกษาสรุปว่าคุณสมบัติของแบบไฮบริด ในแง่ของการ RMSEs แบบไฮบริดดำเนินการกับ KLCAT และการออกแบบที่การทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) มี RMSEs ใกล้เคียงที่กลุ่มความสามารถที่แตกต่างกัน สำหรับวิธีการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มี RMSEs ค่อนข้างใหญ่ที่กลุ่มความสามารถสูง/ต่ำ เมื่อเทียบกับกลุ่มผู้ที่มีความสามารถในระดับกลาง แสดงให้เห็นข้อเสียอย่างหนึ่ง

ในวิธีการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จำนวนที่จำกัดของระดับโมดูลที่ยาก นำไปสู่สถานการณ์ที่ผู้สอบมีความสามารถสูงได้รับข้อสอบที่มากกว่าในระดับความยากที่ต่ำกว่าความสามารถของตน ในขณะที่ผู้สอบที่มีความสามารถต่ำได้รับข้อสอบที่มีระดับความยากสูงกว่าความสามารถและได้ผลใน RMSEs ขนาดใหญ่ อย่างไรก็ตามแบบไฮบริดได้เปรียบเทียบผลที่เกิดจากการสร้างขึ้นในขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) ผลดังกล่าวแสดงให้เห็นว่าแบบไฮบริดสามารถแก้ไขความถูกต้องในการประมาณค่าที่ต่ำของวิธี MST สำหรับผู้สอบไล่ตั้งอยู่ช่วงความสามารถปลายทั้งสอง

จากการศึกษาหลักการการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เป็นระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ที่มีการทำงานร่วมกันของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอนและการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ ขั้นตอนการทดสอบด้วยคอมพิวเตอร์แบบไฮบริดในทดสอบจะเริ่มต้นการการทดสอบด้วยการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ด้วยชุดข้อสอบที่กำหนดขึ้น (MST ขั้นที่ 1) ที่มีความยากระดับปานกลาง เมื่อผู้ตอบทำข้อสอบครบ ระบบจะทำการวิเคราะห์ความสามารถของผู้สอบ ( $\theta$ ) โดยใช้วิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (maximum likelihood : MLE) จากนั้นระบบจะคัดเลือกชุดข้อสอบถัดไปที่มีความยากใกล้เคียงกับความสามารถของผู้สอบ (MST ขั้นที่ 2) ในขั้นตอนนี้จะมีการแบ่งโมดูลข้อสอบออกเป็น 3 ระดับ ได้แก่ ระดับง่าย ระดับปานกลาง และระดับยาก ซึ่งในการคัดเลือกข้อสอบด้วยวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI)โดยมีการควบคุมการใช้ข้อสอบซ้ำ วิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า

(Progressive-restricted standard error: PR-SE) และอัตราการใช้ข้อสอบซ้ำสูงสุดเท่ากับ 20 เปอร์เซนต์ เมื่อผู้สอบทำชุดข้อสอบตามระดับความสามารถแล้ว แล้วระบบจะประมาณค่าความสามารถของผู้สอบจากการตอบข้อสอบในชุดที่ 2 เพื่อคัดเลือกข้อสอบถัดไปโดยการสุ่มข้อสอบที่ละข้อที่ใกล้เคียงความสามารถของผู้สอบ ซึ่งเป็นส่วนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ ทำการทดสอบจนการทดสอบมีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) ต่ำกว่าที่กำหนด เช่น น้อยกว่าหรือเท่ากับ 0.3 แล้ว ยุติการทดสอบ

จากการศึกษาวิจัยที่เกี่ยวข้องกับการคัดเลือกข้อสอบและการควบคุมการใช้ข้อสอบซ้ำของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ พบว่า วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) เป็นวิธีเพื่อใช้ควบคุมเงื่อนไขบังคับอย่างเข้มงวดในการคัดเลือกข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ จะพิจารณาความแปรปรวนของค่าสารสนเทศโดยการนำดัชนีไปเป็นตัวคูณให้กับค่าสารสนเทศของข้อสอบ ดังนั้นในการคัดเลือกข้อสอบแทนที่จะพิจารณาจากค่าสารสนเทศเพียงอย่างเดียววิธีนี้จะนำลำดับความสำคัญพิจารณาร่วมด้วย และสามารถปรับให้เข้ากับข้อบังคับที่ไม่ใช่ทางสถิติที่หลากหลายและสามารถควบคุมได้หลายอย่างพร้อมๆ กัน เช่น ความสมดุลของเนื้อหา การควบคุมการใช้ข้อสอบซ้ำ ความสมดุลของคำตอบ โดยที่วิธีนี้ สามารถนำไปปรับใช้กับโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ที่มีอยู่ในปัจจุบันได้ สะดวก โดยไม่จำเป็นต้องปรับค่าน้ำหนักความสัมพันธ์ระหว่างข้อบังคับและสารสนเทศ และวิธีการควบคุมการใช้ข้อสอบซ้ำด้วยวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PR-SE) โดยนำอัตราส่วนของความคลาดเคลื่อนมาตรฐาน (SE) มาใช้ในกฎการหยุด โดยการใช้ความคลาดเคลื่อนมาตรฐาน (SE) ไม่ได้ขึ้นอยู่กับตำแหน่งตามลำดับ แต่ใช้ความคลาดเคลื่อนมาตรฐานเพื่อตรวจสอบองค์ประกอบการสุ่ม หรือ สารสนเทศที่มีอิทธิพลต่อการถ่วงน้ำหนัก วิธีการ Progressive-restricted standard error มีการดำเนินการบริหารจัดการที่ดีเมื่อเปรียบเทียบกับขั้นตอน การควบคุมการเปิดเผยข้อสอบวิธีอื่นๆ และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบกำหนดความยาวคงที่ และ แบบความยาวแปรผัน เนื่องจาก จะกำหนดค่าความคลาดเคลื่อนมาตรฐาน (SE) ของการประมาณค่าความสามารถผู้สอบ ทั้งสองวิธีเป็นวิธีการที่ยืดหยุ่นเหมาะสมกับการนำมาปรับใช้กับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ซึ่งการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดนี้จะทำการศึกษาภายใต้คลังข้อสอบตามหลักการทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) ใช้รูปแบบโลจิสติก 3 พารามิเตอร์ (3 Parameter Logistic Model : 3 PLM) คือ ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -4 ถึง 4 ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0 ถึง  $\infty$  ค่าการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.50 และการคำนวณผลการตอบข้อสอบตอบถูกต้องได้ 1 คะแนน และตอบผิดได้คะแนนเท่ากับ 0 (Wang et al., 2016) การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

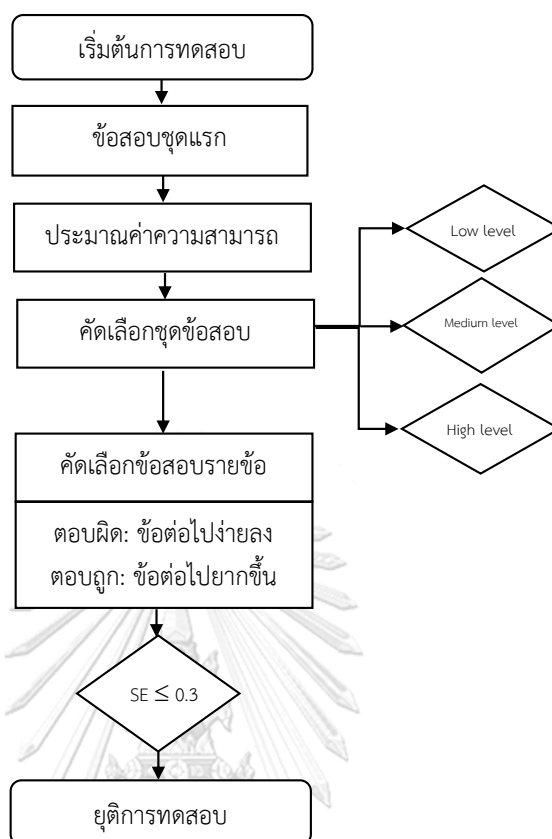
จากการเปรียบเทียบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ การทดสอบปรับเหมาะด้วย คอมพิวเตอร์แบบหลายขั้นตอน (MST) การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด (Hybrid design) จะเห็นได้ว่าการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดเป็นการนำหลักการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอนและการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อมารวมกันทำงานเพื่อช่วยการประมาณค่าความสามารถของผู้สอบให้ใกล้เคียงกับค่าที่แท้จริง และสามารถควบคุมการทดสอบให้มีประสิทธิภาพด้วยกระบวนการทางคอมพิวเตอร์ (Wang et al., 2016) ดังตาราง 2.5



ตาราง 2.5 การเปรียบเทียบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบรายชื่อ การทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด (Hybrid design)

เงื่อนไข	การทดสอบที่เหมาะสมด้วย คอมพิวเตอร์ แบบรายชื่อ	การทดสอบที่เหมาะสมด้วย คอมพิวเตอร์ แบบหลายขั้นตอน	การทดสอบที่เหมาะสมด้วย คอมพิวเตอร์ แบบไฮบริด
เริ่มต้นการทดสอบ	ใช้ข้อสอบเพียง ข้อ ใน 1 การประมาณค่า ความสามารถของผู้สอบ	ใช้ชุดข้อสอบ โดยให้ผู้สอบทำ ข้อสอบทั้งหมด ในการประมาณ ค่าความสามารถของผู้สอบ	ใช้ชุดข้อสอบ ที่กำหนดไว้โดยให้ ผู้สอบทำข้อสอบทั้งหมด ในการ ประมาณค่าความสามารถของผู้สอบ
การเลือกข้อสอบข้อ ถัดไป	เลือกข้อถัดไปจากค่า ความสามารถที่ได้จากการ ประมาณค่าในข้อแรก ถ้าทำ ถูก ข้อถัดไปจะมีความยาก ขึ้น ถ้าทำผิดข้อถัดไปจะมี ความยากลดลง	เลือกชุดข้อสอบข้อถัดไปโดยดู จากค่าความสามารถในการ ประมาณค่าในขั้นแรก โดย เลือกชุดที่มีความยากใกล้เคียง กับผู้สอบ ถ้าผู้สอบมี ความสามารถต่ำจะเลือกชุด ข้อสอบที่ง่ายลง ถ้าผู้สอบมี ความสามารถสูงจะเลือกชุด ข้อสอบที่ยากขึ้น	- เลือกข้อสอบชุดถัดไปโดยดูจาก ค่าความสามารถในการประมาณค่า จาก การตอบข้อสอบในชุดเริ่มต้น โดยเลือกชุดที่มีความยากใกล้เคียงกับ ผู้สอบ ถ้าผู้สอบมีความสามารถต่ำจะ เลือกชุดข้อสอบที่ง่ายลง ถ้าผู้สอบมี ความสามารถสูงจะเลือกชุดข้อสอบที่ ยาก - เลือกข้อสอบถัดไปเป็นรายชื่อ จากค่าความสามารถที่ได้จากการ ประมาณค่าใน การตอบข้อสอบ ชุดที่ โดยเลือกข้อสอบที่มี ความ 2 ยากใกล้เคียงกับผู้สอบ ถ้าผู้สอบมี ความสามารถต่ำจะเลือกข้อสอบที่ง่าย ลง ถ้าผู้สอบมีความสามารถสูงจะ เลือกข้อสอบที่ยากขึ้น
การเตรียมข้อสอบใน ข้อถัดไป	มีการจัดเตรียมข้อสอบใน คลังข้อสอบ	มีการจัดเตรียมชุดข้อสอบไว้ ล่วงหน้า	- ขั้นตอนแรก และขั้นตอนที่มี 2 การจัดเตรียมชุดข้อสอบไว้ล่วงหน้า - ขั้นตอนที่สาม มีการจัดเตรียม ข้อสอบรายชื่อในคลังข้อสอบ
การประมาณค่า ความสามารถ	ประมาณค่าความสามารถ จากข้อสอบเพียงข้อเดียวใช้ วิธี maximizing the fisher information (MFI)	ใช้การประมาณค่าจากชุดของ ข้อสอบในขั้นตอนเริ่มต้น	ประมาณค่าความสามารถจากข้อสอบ ใช้วิธี การประมาณค่าด้วย ความเป็นไปได้สูงสุด (maximum likelihood)

ที่มา: Wang et al. (2016) และ ฅณัทร ชัยมงคล (2558)



ภาพ 2.5 แผนภาพการทำงานของ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

กระบวนการในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ประกอบไปด้วย รายละเอียดดังนี้

- จุดเริ่มต้นการทดสอบ (starting point or entry level)

จุดเริ่มต้นในการทำข้อสอบจะทำให้การประมาณค่าความสามารถสูงหรือต่ำกว่าจริง ดังนั้นในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เริ่มต้นการทดสอบด้วยชุดข้อสอบที่กำหนดขึ้น (initial item) มีความยากปานกลาง (ค่า  $b$  อยู่ระหว่าง  $-0.5$  ถึง  $0.5$ ) เพื่อให้การประมาณค่าความสามารถมีความแม่นยำมากยิ่งขึ้นจัดให้กับผู้สอบทุกคนเมื่อไม่ทราบค่าสารสนเทศเบื้องต้นของผู้สอบ

- การประมาณค่าความสามารถ

สำหรับขั้นตอนการออกแบบระบบการประมาณค่า ใช้การประมาณค่าร่วมกัน วิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (maximum likelihood) คัดเลือกข้อสอบที่มีตำแหน่งสารสนเทศสูงสุด ( $m_i$ ) สอดคล้องกับระดับความสามารถที่ประมาณได้ ( $\theta$ ) [Match  $m_i$  to  $\theta$ ]

ข้อสอบมักมีปัจจัย การเดาเข้ามาเกี่ยวข้องของการคัดเลือกข้อสอบที่มีค่า  $b_i$  ใกล้เคียงที่สุดกับ  $\theta$  จึงมักมีแนวโน้มที่จะได้ข้อที่ยากเกินไปสำหรับผู้สอบ เนื่องจากข้อสอบแต่ละข้อให้สารสนเทศสูงสุดที่ตำแหน่ง  $\theta = m_i$  เมื่อ  $m_i$  สำหรับโมเดล 3 พารามิเตอร์ คำนวณได้ดังนี้

$$m_i = b_i + \frac{1}{Da_1} \ln \left[ \frac{1 + \sqrt{1 + 8c_i}}{2} \right]$$

เมื่อ	D	=	ค่าคงที่
	ln	=	natural logarithm
	$a_i$	=	ค่าอำนาจจำแนกของข้อสอบ
	$b_i$	=	ค่าความยากของข้อสอบ
	$c_i$	=	ค่าโอกาสการเดาข้อสอบถูก

$m_i$  เป็นค่า  $\theta$  ณ ตำแหน่งที่ข้อสอบข้อนั้นให้สารสนเทศสูงสุด ซึ่ง  $m_i > b_i$  ถ้า  $c_i > 0$  แสดงว่าข้อสอบที่มีค่าความยาก  $b_i$  เหมาะที่จะใช้ประมาณค่า  $\theta$  ของผู้สอบที่มีความสามารถสูงกว่า  $b_i$  เล็กน้อย ดังนั้น การคัดเลือกข้อสอบจากค่า  $b_i$  จึง สามารถปรับปรุงให้ดีขึ้น เมื่อพิจารณาจากค่า  $m_i$

- การคัดเลือกข้อสอบ (item selection)

กระบวนการคัดเลือกข้อสอบสามารถนำมาใช้กับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ซึ่งในแต่ละขั้น (stage) จะทำการเลือกข้อสอบที่มีสารสนเทศสูงสุดภายหลังการประมาณค่าความสามารถของผู้สอบ การคัดเลือกข้อสอบเป็นสิ่งสำคัญในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด วิธีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) เป็นวิธีการคัดเลือกข้อสอบที่เพิ่มประสิทธิภาพจากการใช้ค่าฟังก์ชันสารสนเทศสูงสุด (maximum item information function) แบบดั้งเดิม เป็นการกำหนดค่าถ่วงน้ำหนักตามเงื่อนไขที่กำหนด ทำให้การคัดเลือกข้อสอบมีความแม่นยำมากยิ่งขึ้น ดังนั้นในการวิจัยนี้จึงเลือกใช้วิธีลำดับความสำคัญสูงสุดในการคัดเลือกข้อสอบในแต่ละขั้น (Y. Cheng & Chang, 2009; ฅภัทร ชัยมงคล, 2558)

- การควบคุมการใช้ข้อสอบซ้ำ (exposure control)

ในการควบคุมการเปิดเผยของข้อสอบสามารถควบคุมได้ด้วยกระบวนการอัตโนมัติเหมือนกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ โดยเราสามารถป้องกันข้อสอบที่ถูกนำมาใช้บ่อย (over-exposed) จากกรณีที่มีการเปิดเผยและมีการแบ่งปันข้อสอบกับผู้สอบอื่น วิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PR-SE) เป็นวิธีที่นำอัตราส่วนของความคลาดเคลื่อนมาตรฐาน (SE) มาใช้ในกฎการหยุด โดยการใช้ความคลาดเคลื่อนมาตรฐาน (SE) เพื่อตรวจสอบองค์ประกอบการสุ่ม หรือ สารสนเทศที่มีอิทธิพลต่อการถ่วงน้ำหนัก



และเหมาะสมกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบกำหนดความยาวคงที่ และ แบบ ความยาวแปรผัน เนื่องจากจะกำหนดค่าความคลาดเคลื่อนมาตรฐาน (SE) ของการประมาณ ค่าความสามารถผู้สอบ ดังนั้นในการวิจัยนี้จึงเลือกใช้วิธีวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัด ความก้าวหน้าในการควบคุมการเปิดเผยข้อสอบ (McClarty et al., 2006)

#### - การยุติการสอบ (Stopping criteria)

การกำหนดค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ในทางปฏิบัติการทดสอบปรับเหมาะกับความสามารถของผู้สอบโดยใช้คอมพิวเตอร์ ควรให้ดำเนินการ ไปเรื่อยจนกว่า การประมาณค่า  $\theta$  มีความคลาดเคลื่อนมาตรฐาน (Standard Error: SE ( $\theta$ )) ลด ต่ำลงจนถึงระดับที่ยอมรับได้ การทดสอบจึงควรยุติลง ซึ่งการกำหนดค่าความคลาดเคลื่อนมาตรฐาน ในการประมาณค่าความสามารถของผู้สอบต่ำกว่าค่าที่กำหนด เช่น น้อยกว่าหรือเท่ากับ 0.30 จะให้ ค่าความตรงตามสภาพสูงสุด (พงษ์พิชิต พรหมสิทธิ์ โพศาล สุวรรณน้อย และ รังสรรค์ มณีเล็ก, 2555; รังสรรค์ มณีเล็ก, 2540; สิริลักษณ์ เกษรปทุมานันท์, 2549)

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ถูกพัฒนาขึ้นเพื่อช่วยการประมาณค่า ความสามารถของผู้สอบให้ใกล้เคียงกับค่าที่แท้จริงของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบ รายข้อที่อาจเกิดอาจเกิดการประมาณค่าความสามารถที่เกินจริง (over estimate) และต่ำกว่าจริง (under estimate) และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ซึ่งใช้ วิธีการการจัดเตรียมชุดของข้อสอบไว้โดยบุคคล ดังนั้นอาจเกิดความผิดพลาดในการจัดชุดข้อสอบได้ ด้วยเหตุนี้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดจะช่วยลดความผิดพลาดดังกล่าว และ สามารถควบคุมการทดสอบให้มีประสิทธิภาพด้วยกระบวนการทางคอมพิวเตอร์ (Wang et al., 2016)

### ตอนที่ 3 การประเมินผลระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์

การประเมินผลระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์มีการนำแนวคิดการประเมิน ตามมาตรฐานการประเมินระบบและการประเมินแบบฮิวริสติก (heuristic evaluation) มาใช้ร่วมกัน ซึ่งในการประเมินจะอาศัยผู้เชี่ยวชาญทางด้านคอมพิวเตอร์และการศึกษา สำหรับด้านเนื้อหาที่นำเข้ามา ใช้งานในระบบจะต้องอาศัยความเชี่ยวชาญจากผู้เชี่ยวชาญด้านการศึกษา ซึ่งรายละเอียดในการ ประเมินผลระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์มีรายละเอียดดังนี้

#### 3.1 การประเมินตามมาตรฐานการประเมินระบบ

การประเมินตามมาตรฐานการประเมินระบบของ Stufflebeam และคณะ ประกอบด้วย เนื้อหาหลักในการประเมินทั้งสิ้น 4 หมวด ได้แก่ การใช้ประโยชน์ (utility) ความเป็นไปได้

(feasibility) ความเหมาะสม (propriety) และความถูกต้อง (accuracy) ซึ่งมีรายละเอียดดังนี้ (ศิริชัย กาญจนวาสี, 2554)

- การใช้ประโยชน์ (utility) เป็นการประกันความเป็นประโยชน์ของระบบในการตอบสนองต่อความต้องการในการใช้สารสนเทศของผู้ที่เกี่ยวข้องให้ครอบคลุม ทันเวลา และมีผลต่อการนำไปใช้
- ความเป็นไปได้ (feasibility) เป็นการประกันความสอดคล้องของระบบกับสภาพความเป็นจริง เหมาะสมกับสถานการณ์ ปฏิบัติได้ ยอมรับได้ ประหยัดและคุ้มค่า
- ความเหมาะสม (propriety) เป็นการประกันว่าระบบมีความเหมาะสมซึ่งเป็นไปตามกฎระเบียบ จรรยาบรรณ คำนึงถึงสวัสดิภาพของผู้ที่เกี่ยวข้องและผู้ที่ได้รับผลกระทบ
- ความถูกต้อง (accuracy) เป็นการประกันว่าระบบมีการใช้เทคนิคที่เหมาะสม ให้สารสนเทศที่เพียงพอ

### 3.2 การประเมินแบบฮิวริสติก (heuristic evaluation)

การประเมินแบบฮิวริสติกเป็นวิธีการค้นหาปัญหาจากการใช้งานในการออกแบบหน้าจอระบบ (user interface) โดยมีชุดของนักประเมินจำนวนหนึ่งในการตรวจสอบหน้าจอระบบและตัดสินใจเพื่อยอมรับว่าหน้าจอสามารถใช้งานได้โดยทั่วไป พร้อมให้ความคิดเห็นว่าหน้าจอส่วนใดที่ดีและส่วนใดควรต้องปรับปรุง (Molich & Nielsen, 1990; Nielsen, 1992) การประเมินแบบฮิวริสติกเป็นวิธีการที่เร็วที่สุด ประหยัด และเป็นวิธีที่มีประสิทธิภาพในการระบุปัญหาของการใช้งานโดยการประเมินจะใช้กับระบบจริงหรือระบบต้นแบบ (prototype) ซึ่งควบคุมโดยผู้เชี่ยวชาญ (expert) ในการประเมินผู้เชี่ยวชาญจะใช้รายการตรวจสอบตามรายการประเมินแบบฮิวริสติก ซึ่งโดยธรรมชาติของการประเมินการใช้งานของหน้าจอระบบไม่สามารถทำได้โดยอัตโนมัติซึ่งคงยังต้องใช้มนุษย์ในการตรวจสอบ สำหรับรายการประเมินฮิวริสติกแบบดั้งเดิมตามแนวคิดของ Nielsen ประกอบด้วยรายการประเมินทั้งสิ้น 10 รายการโดยมีรายละเอียดดังนี้ (S. Y. Chen & Macredie, 2005; Gómez, Caballero, & Sevillano, 2014; ฅภัทร ชัยมงคล, 2558)

- ความชัดเจนของระบบ (Visibility of system status)  
ระบบควรจะมีการแจ้งเตือนผู้ใช้งานอยู่ตลอดเวลาว่ากำลังดำเนินการอะไรอยู่ โดยมีการให้ข้อมูลย้อนกลับที่เหมาะสมภายในระยะเวลาที่สมเหตุสมผล
- ความสอดคล้องของระบบและการใช้งานจริง (Match between system and the real world)  
ระบบควรพูดภาษาเดียวกับผู้ใช้งาน ด้วยคำ ประโยค และแนวคิดซึ่งผู้ใช้งานมีความคุ้นเคยมากกว่า การใช้คำศัพท์ที่เป็นของระบบ โดยให้เป็นไปตามแบบแผนของโลกความเป็นจริงซึ่งการจัดทำข้อมูลต้องมีลำดับที่เป็นธรรมชาติและมีตรรกะ

- การควบคุมและอิสระของผู้ใช้งาน (User control and freedom)

ผู้ใช้งานควรมีอิสระในการพัฒนาแนวทางในการใช้งานของตนเอง เลือกและเรียงลำดับเครื่องมือ และสามารถยกเลิกหรือทำกิจกรรมซ้ำที่ได้ทำไปแล้ว มากกว่าที่จะกำหนดระบบให้ผู้ใช้งาน

- ความสอดคล้องและมาตรฐาน (Consistency and standards)

ผู้ใช้งานไม่ควรสับสนกับคำพูด สถานการณ์ หรือการกระทำที่มีความแตกต่างกัน แต่มีความหมายเหมือนกัน และระบบควรมีการทำงานที่เป็นแบบแผน

- การป้องกันความผิดพลาด (Error prevention)

ข้อความแจ้งเตือนความผิดพลาดควรการออกแบบอย่างระมัดระวังเพื่อป้องกันปัญหาที่เกิดขึ้นในการใช้งานครั้งแรก

- การจำข้อมูลมากกว่าการเรียกคืนข้อมูล (Recognition rather than recall)

การสร้างวัตถุ การกระทำ และตัวเลือกในการใช้งาน ผู้ใช้งานไม่ควรที่จะต้องจำข้อมูลจากส่วนหนึ่งเพื่อไปใช้งานอีกส่วนหนึ่ง คำสั่งในการใช้งานของระบบควรมองเห็นได้ง่ายและมีความเหมาะสม

- ความยืดหยุ่นและประสิทธิภาพการใช้งาน (Flexibility and efficiency of use)

อนุญาตให้ผู้ใช้งานเลือกการกระทำที่ทำบ่อยและเหมาะกับผู้ใช้งาน โดยการให้ทางเลือกซึ่งหมายถึง การเข้าถึงและการดำเนินงานสำหรับผู้ใช้งานที่มีความแตกต่างจากผู้ใช้งานปกติ (เช่น ด้านกายภาพ ความสามารถทางสมอง วัฒนธรรม ภาษา เป็นต้น)

- การออกแบบที่สวยงาม (Aesthetic and minimalist design)

ข้อความแจ้งเตือนไม่ควรประกอบด้วยข้อมูลที่ไม่มี ความสอดคล้องหรือไม่มีความต้องการ การแจ้งเตือนที่มีความพิเศษต้องนำมาเทียบกับการแจ้งเตือนแบบปกติเพื่อดูความสอดคล้อง หากไม่มีความจำเป็นควรลด การแจ้งเตือนให้เหลือเพียงแบบปกติ

- การช่วยเหลือโดยการกู้คืนระบบจากข้อผิดพลาด (Help users recognize, diagnose, and recover from errors)

ข้อความแจ้งเตือนควรมีความชัดเจนเพื่อบ่งชี้ปัญหาและการแนะนำอย่างมีโครงสร้างในการแก้ปัญหา ซึ่งควรนำเสนอในภาษาที่เรียบง่าย

- การช่วยเหลือและคู่มือ (Help and documentation)

แม้ว่าจะเป็นข้อดีหากระบบสามารถใช้งานโดยไม่ต้องใช้คู่มือ แต่ก็อาจมีความจำเป็นที่จะต้องให้การช่วยเหลือและให้เอกสารประกอบ ทุกองค์ประกอบของสารสนเทศควรง่ายต่อการค้นหา และเน้นในส่วนเครื่องมือของผู้ใช้งานเป็นหลัก มีการจัดเรียงลำดับในการค้นหาและรายการไม่ควรยาวเกินไป

### 3.3 การประเมินความพึงพอใจของผู้ใช้งานต่อระบบ

การประเมินความพึงพอใจของระบบโดยผู้ใช้งาน เป็นสิ่งสำคัญในการพัฒนาระบบ เนื่องจากระบบตอบสนองการใช้งานจนทำให้ผู้ใช้เกิดความพึงพอใจ จะส่งผลให้ระบบได้รับการสนับสนุนและผลักดันในการนำไปใช้งานจริง ดังนั้นผู้พัฒนาระบบต้องมีการประเมินความพึงพอใจของผู้ใช้งานต่อระบบ (โชติกา ภาชีผล, 2561; ฌภัทร ชัยมงคล, 2558) โดยมีองค์ประกอบ ดังนี้

- ภาพรวมของระบบ คือ ความพึงพอใจในภาพรวมของผู้ใช้งานสำหรับการใช้งานระบบ โดยความพึงพอใจของผู้ใช้งานขึ้นอยู่กับการใช้คำศัพท์ ข้อมูลย้อนกลับจากระบบ การให้ความช่วยเหลือโดยระบบ และการออกแบบหน้าจอ

- หน้าจอของระบบ คือ การการที่ผู้ใช้งานสามารถอ่านรายละเอียดบนหน้าจอของระบบได้ง่าย ชัดเจน รวมถึงความสวยงามของรูปแบบในการจัดวางและการใช้ประโยชน์ ผู้สอบสามารถเรียนรู้เมนูการใช้งานบนหน้าจอได้ด้วยตนเอง และหน้าจอจะต้องกระตุ้นความสนใจของผู้สอบ และดึงดูดความน่าสนใจของผู้สอบให้สนใจอยู่กับการทดสอบตลอดเวลา

- การเรียนรู้ คือ การที่ผู้ใช้งานเกิดความเข้าใจว่าระบบง่ายต่อการเรียนรู้ ไม่ยุ่งยากซับซ้อน รายละเอียดและคำสั่งง่ายต่อการจดจำ

- ความสามารถของระบบ คือ การที่ผู้ใช้งานสามารถใช้งานระบบได้ง่าย พร้อมทั้งสามารถแก้ไขข้อผิดพลาดของระบบได้อย่างรวดเร็ว รวมถึงมีความสัมพันธ์กับระดับประสบการณ์ของผู้ใช้งาน นอกจากนี้ระบบจะต้องมีเครื่องมือที่สนับสนุนการใช้งานที่หลากหลายและมีประสิทธิภาพ

### ตอนที่ 4 มาตรฐานและตัวชี้วัดภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3

ในช่วงที่ผ่านมา ประเทศไทยได้มีความพยายามที่จะเพิ่มขีดความสามารถในการใช้ภาษาอังกฤษด้วยการประกาศนโยบายให้นักเรียนเรียนภาษาอังกฤษเป็นภาษาต่างประเทศที่ 1 โดยกำหนดให้มีการสอนตั้งแต่ระดับประถมศึกษาเป็นต้นไป ในปี 2538 และได้ประกาศใช้หลักสูตรภาษาอังกฤษ พ.ศ.2539 รวมทั้งการส่งเสริมให้มีการเปิดโรงเรียนนานาชาติ และโรงเรียน English Program ซึ่งเป็นโรงเรียนที่จัดการเรียนการสอนตามหลักสูตรกระทรวงศึกษาธิการ โดยใช้ภาษาอังกฤษเป็นสื่อในการจัดการเรียนการสอน แม้จะมีมาตรการดังกล่าวแต่ยังไม่ส่งผลต่อความสามารถในการใช้ภาษาอังกฤษเท่าที่ควร ในปี 2549 กระทรวงศึกษาธิการมีนโยบายที่จะปฏิรูปการเรียนการสอนภาษาอังกฤษทั้งระบบ โดยใช้วิธีการปรับเปลี่ยนกระบวนทัศน์และพัฒนาการเรียนการสอนภาษาอังกฤษแบบสื่อสาร การสร้างความเสมอภาคในโอกาสการเรียนภาษาอังกฤษและการสร้างบรรยากาศและเพิ่มโอกาสในการเรียนรู้ภาษาอังกฤษนอกห้องเรียนและได้เสนอ "แผนยุทธศาสตร์ปฏิรูปการเรียนการสอนภาษาอังกฤษ เพื่อเพิ่มขีดความสามารถในการแข่งขันของประเทศ (พ.ศ.2549-2553)" (กระทรวงศึกษาธิการ, 2553)

ยุทธศาสตร์ที่สำคัญที่กำหนดไว้ในแผนยุทธศาสตร์ปฏิรูปการเรียนการสอนภาษาอังกฤษ (พ.ศ.2549-2553) รวมทั้งคณะอนุกรรมการการศึกษาด้านการต่างประเทศ ได้จัดทำรายงานการศึกษาสภาพปัญหาการเรียนการสอนภาษาต่างประเทศและให้ข้อเสนอแนะไว้คือ การจัดตั้งสถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน เพื่อเป็นหน่วยงานหลักในการทำหน้าที่ส่งเสริมและพัฒนการเรียนการสอนและการใช้ภาษาอังกฤษ โดยเฉพาะในระดับการศึกษาขั้นพื้นฐาน ซึ่งเป็นระดับที่มีความสำคัญอย่างยิ่งในการสร้างทักษะพื้นฐานและเจตคติต่อการเรียนรู้ของนักเรียนและเยาวชนซึ่งจะต้องเรียนรู้และพัฒนาไปสู่ความสามารถในการใช้ภาษาเพื่อการสื่อสารและการแสวงหาความรู้เพื่อการปฏิบัติงานและเป็นพื้นฐานในการศึกษาต่อเนื่องจากสภาพปัจจุบันของการจัดการเรียนการสอนภาษาอังกฤษยังมีปัญหอยู่มาก เกี่ยวกับการพัฒนาหลักสูตร กระบวนการเรียนรู้ สื่อ และการวัดประเมินผลทางภาษา

การจัดตั้งสถาบันภาษาอังกฤษขึ้น เพื่อเป็นหน่วยงานหลักในการส่งเสริมและพัฒนาปัจจัยต่างๆที่ส่งผลต่อคุณภาพการเรียนการสอนและการใช้ภาษาอังกฤษในระดับการศึกษาขั้นพื้นฐานทั้งระบบ ทั้งด้านการส่งเสริม วิจัยและพัฒนาการนำหลักสูตรไปใช้การจัดการเรียนรู้ การวัดประเมินผล สื่อและนวัตกรรมการเรียนรู้ภาษาอังกฤษ กิจกรรมพัฒนาทักษะการใช้ภาษาอังกฤษของนักเรียน การพัฒนาครูและบุคลากรที่เกี่ยวข้อง การพัฒนาสถานศึกษาและเครือข่ายการพัฒนาการเรียนการสอนทุกระดับ องค์การวิชาชีพด้านการเรียนการสอนภาษาอังกฤษ การประสานงานความร่วมมือกับหน่วยงานทั้งภาครัฐและเอกชน ทั้งภายในและต่างประเทศ รวมไปถึงการให้บริการสื่อการเรียนรู้ สื่อ On-line การทดสอบระดับความสามารถด้านภาษาอังกฤษ

สถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ตระหนักถึงความสำคัญในการพัฒนาแบบฝึกทักษะภาษาอังกฤษ ว่าจะเป็นเครื่องมือที่สำคัญหนึ่งที่มีส่วนช่วยให้ผู้เรียนมีความคุ้นเคยรูปแบบข้อสอบ คำสั่งในข้อสอบและการฝึกฝนการประเมิน ส่งผลต่อการยกระดับผลสัมฤทธิ์การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) จึงพัฒนาแบบฝึกทักษะภาษาอังกฤษ ระดับชั้นมัธยมศึกษาปีที่ 3 เพื่อเป็นประโยชน์สำหรับเขตพื้นที่การศึกษา ศึกษาในเทศก์ และครูผู้สอนในการนำไปใช้ เพื่อก่อให้เกิดการพัฒนาอย่างต่อเนื่องต่อไป โดยแบบฝึกทักษะภาษาอังกฤษ ระดับชั้นมัธยมศึกษาปีที่ 3 ที่จัดทำขึ้นมีการวัดใน 4 ด้วย ได้แก่ Conversation, Grammar and Vocabulary, Reading Comprehension และ Miscellaneous ซึ่งออกโดยครูผู้เชี่ยวชาญทางด้านภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ข้อสอบจะสอดคล้องกับ ตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ดังนี้

## **สาระและมาตรฐานการเรียนรู้**

### **สาระที่ 1 ภาษาเพื่อการสื่อสาร**

- มาตรฐาน ต 1.1 เข้าใจและตีความเรื่องที่ฟังและอ่านจากสื่อประเภทต่างๆ และแสดงความคิดเห็นอย่างมีเหตุผล
- มาตรฐาน ต 1.2 มีทักษะการสื่อสารทางภาษาในการแลกเปลี่ยนข้อมูลข่าวสาร แสดงความรู้สึก และความคิดเห็นอย่างมีประสิทธิภาพ
- มาตรฐาน ต 1.3 นำเสนอข้อมูลข่าวสาร ความคิดรวบยอด และความคิดเห็นในเรื่องต่างๆ โดยการพูดและการเขียน

### **สาระที่ 2 ภาษาและวัฒนธรรม**

- มาตรฐาน ต 2.1 เข้าใจความสัมพันธ์ระหว่างภาษากับวัฒนธรรมของเจ้าของภาษา และนำไปใช้ได้อย่างเหมาะสมกับกาลเทศะ
- มาตรฐาน ต 2.2 เข้าใจความเหมือนและความแตกต่างระหว่างภาษาและวัฒนธรรมของเจ้าของภาษากับภาษาและวัฒนธรรมไทย และนำมาใช้อย่างถูกต้องและเหมาะสม

### **สาระที่ 3 ภาษากับความสัมพันธ์กับกลุ่มสาระการเรียนรู้อื่น**

- มาตรฐาน ต 3.1 ใช้ภาษาต่างประเทศในการเชื่อมโยงความรู้กับกลุ่มสาระการเรียนรู้อื่น และเป็นพื้นฐานในการพัฒนา แสวงหาความรู้ และเปิดโลกทัศน์ของตน

### **สาระที่ 4 ภาษากับความสัมพันธ์กับชุมชนและโลก**

- มาตรฐาน ต 4.1 ใช้ภาษาต่างประเทศในสถานการณ์ต่างๆ ทั้งในสถานศึกษา ชุมชน และสังคม
- มาตรฐาน ต 4.2 ใช้ภาษาต่างประเทศเป็นเครื่องมือพื้นฐานในการศึกษาต่อ การประกอบอาชีพ และการแลกเปลี่ยนเรียนรู้กับสังคมโลก

### ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3

#### สาระที่ 1 ภาษาเพื่อการสื่อสาร

มาตรฐาน ต 1.1 เข้าใจและตีความเรื่องที่ฟังและอ่านจากสื่อประเภทต่างๆ และแสดงความคิดเห็นอย่างมีเหตุผล

ตาราง 2.6 ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 1 ภาษาเพื่อการสื่อสาร มาตรฐาน ต 1.1

ชั้น	ตัวชี้วัด	สาระการเรียนรู้แกนกลาง
ม.3	1. ปฏิบัติตามคำขอร้อง คำแนะนำ คำชี้แจง และคำอธิบายที่ฟังและอ่าน	คำขอร้อง คำแนะนำ คำชี้แจง และคำอธิบาย ในการประดิษฐ์ การบอกทิศทาง ป้ายประกาศต่างๆ การใช้อุปกรณ์ - Passive Voice ที่ใช้ในโครงสร้างประโยคง่ายๆ เช่น is/are + past participle - คำสันธาน (conjunction) เช่น and/ but/ or/ before/ after/ because etc. - ตัวเชื่อม (connective words) เช่น First,... Second,... Third,... Fourth,... Next,... Then,... Finally,... etc.
	2. อ่านออกเสียงข้อความ ข่าว โฆษณา และบทร้อยกรองสั้นๆ ถูกต้องตามหลักการอ่าน	ข้อความ ข่าว โฆษณา และบทร้อยกรอง การใช้พจนานุกรม หลักการอ่านออกเสียง เช่น - การออกเสียงพยัญชนะต้นคำและพยัญชนะท้ายคำ สระเสียงสั้น สระเสียงยาว สระประสม - การออกเสียงเน้นหนัก-เบา ในคำและกลุ่มคำ - การออกเสียงตามระดับเสียงสูง-ต่ำ ในประโยค - การออกเสียงเชื่อมโยงในข้อความ - การแบ่งวรรคตอนในการอ่าน - การอ่านบทร้อยกรองตามจังหวะ
	3. ระบุและเขียนสื่อที่ไม่ใช่ความเรียง รูปแบบต่างๆ ให้สัมพันธ์กับประโยค และข้อความที่ฟังหรืออ่าน	ประโยค ข้อความ และความหมายเกี่ยวกับตนเอง ครอบครัว โรงเรียน สิ่งแวดล้อม อาหาร เครื่องดื่ม เวลาว่างและนันทนาการ สุขภาพและสวัสดิการ การซื้อ-ขาย ลมฟ้าอากาศ การศึกษาและอาชีพ การเดินทางท่องเที่ยว การบริการ สถานที่ ภาษา และวิทยาศาสตร์และเทคโนโลยี เป็นวงคำศัพท์สะสมประมาณ 1,400-1,550 คำ (คำศัพท์ที่เป็นรูปธรรมและนามธรรม) การตีความ/ถ่ายโอนข้อมูลให้สัมพันธ์กับสื่อที่ไม่ใช่ความเรียง เช่น สัญลักษณ์ เครื่องหมาย กราฟ แผนภูมิ ตาราง ภาพสัตรี สิ่งของบุคคล สถานที่ต่างๆ โดยใช้ Comparison of adjectives/ adverbs/ Contrast : but, although/ Quantity words เช่น many/ much/ a lot of/ lots of/ some/ any /a few/ few/ a little/ little etc.

ชั้น	ตัวชี้วัด	สาระการเรียนรู้แกนกลาง
	4. เลือก/ระบุหัวข้อเรื่อง ใจความสำคัญ รายละเอียด สนับสนุน และแสดง ความคิดเห็นเกี่ยวกับเรื่องที่ฟังและอ่านจากสื่อประเภทต่างๆ พร้อมทั้งให้เหตุผล และ ยกตัวอย่างประกอบ	<p>การจับใจความสำคัญ เช่น หัวข้อเรื่อง ใจความสำคัญ รายละเอียด สนับสนุน จากสื่อสิ่งพิมพ์และสื่ออิเล็กทรอนิกส์ เช่น หนังสือพิมพ์ วารสาร วิทยุ โทรทัศน์ เว็บไซต์บนอินเทอร์เน็ต</p> <p>คำถามเกี่ยวกับใจความสำคัญของเรื่อง เช่น ใคร ทำอะไร ที่ไหน เมื่อไร อย่างไร ทำไม ใช่หรือไม่</p> <ul style="list-style-type: none"> <li>- Yes/No Question</li> <li>- Wh-Question</li> <li>- Or-Question etc.</li> </ul> <p>ประโยคที่ใช้ในการแสดงความคิดเห็น การให้เหตุผลและการยกตัวอย่าง เช่น I think.../ I feel.../ I believe.../ I agree/disagree.../ I don't believe.../ I have no idea...</p> <ul style="list-style-type: none"> <li>- if clauses</li> <li>- so...that/such...that</li> <li>- คำสันธาน (conjunctions) and/ but/ or/ because/ so/ before/ after etc.- Infinitive pronouns :some/ any/ someone/ anyone/ everyone/ one/ ones etc.</li> <li>- Tenses : present simple/ present continuous/ present perfect/ past simple/ future tense etc.</li> <li>- Simple sentence/ Compound sentence/ Complex sentence</li> </ul>



## สาระที่ 1 ภาษาเพื่อการสื่อสาร

มาตรฐาน ต 1.2 มีทักษะการสื่อสารทางภาษาในการแลกเปลี่ยนข้อมูลข่าวสาร แสดงความรู้สึก และความคิดเห็นอย่างมีประสิทธิภาพ

ตาราง 2.7 ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 1 ภาษาเพื่อการสื่อสาร มาตรฐาน ต 1.2

ชั้น	ตัวชี้วัด	สาระการเรียนรู้แกนกลาง
ม.3	1. สนทนาและเขียนโต้ตอบข้อมูลเกี่ยวกับตนเอง เรื่องต่างๆ ใกล้ตัว สถานการณ์ ข่าว เรื่องที่อยู่ในความสนใจของสังคมและสื่อสารอย่างต่อเนื่องและเหมาะสม	ภาษาที่ใช้ในการสื่อสารระหว่างบุคคล เช่น การทักทาย กล่าวลา ขอบคุณ ขอโทษ ชมเชย การพูดแทรกอย่างสุภาพ การชักชวน การแลกเปลี่ยนข้อมูลเกี่ยวกับตนเอง เรื่องใกล้ตัว สถานการณ์ต่างๆ ในชีวิตประจำวัน การสนทนา/เขียนข้อมูลเกี่ยวกับตนเองและบุคคลใกล้ตัว สถานการณ์ ข่าว เรื่องที่อยู่ในความสนใจในชีวิตประจำวัน
	2. ใช้คำขอร้อง ให้คำแนะนำ คำชี้แจง และคำอธิบายอย่างเหมาะสม	คำขอร้อง คำแนะนำ คำชี้แจง คำอธิบาย ที่มีขั้นตอนซับซ้อน
	3. พูดและเขียนแสดงความต้องการ เสนอ และให้ความช่วยเหลือ ตอบรับและปฏิเสธ การให้ความช่วยเหลือในสถานการณ์ต่างๆ อย่างเหมาะสม	ภาษาที่ใช้ในการแสดงความต้องการ เสนอและให้ความช่วยเหลือ ตอบรับและปฏิเสธการให้ความช่วยเหลือในสถานการณ์ต่างๆ เช่น Please.../..., please./ I'd like.../ I need.../ May/Can/Could...?/ Yes,.../Please do./ Certainly./ Yes, of course./ Sure./ Go right ahead./ Need some help?/ What can I do to help?/ Would you like any help?/ I'm afraid.../ I'm sorry, but.../ Sorry, but...etc.
	4. พูดและเขียนเพื่อขอและให้ข้อมูล อธิบาย เปรียบเทียบ และแสดงความคิดเห็นเกี่ยวกับเรื่องที่ฟังหรืออ่านอย่างเหมาะสม	คำศัพท์ สำนวน ประโยค และข้อความที่ใช้ในการขอและให้ข้อมูล อธิบาย เปรียบเทียบ และแสดงความคิดเห็นเกี่ยวกับเรื่องที่ฟังหรืออ่าน
	5. พูดและเขียนบรรยายความรู้สึก และความคิดเห็นของตนเองเกี่ยวกับ เรื่องต่างๆ กิจกรรม ประสบการณ์ และข่าว/เหตุการณ์ พร้อมทั้งให้เหตุผล ประกอบอย่างเหมาะสม	ภาษาที่ใช้ในการแสดงความรู้สึก ความคิดเห็น และให้เหตุผล ประกอบ เช่น ชอบ ไม่ชอบ ดีใจ เสียใจ มีความสุข เศร้า หิว รสชาติ สวย น่าเกลียด เสียดัง ดี ไม่ดี จากข่าว เหตุการณ์ สถานการณ์ ในชีวิตประจำวัน เช่น Nice./ Very nice./ Well done!/ Congratulations on... / I like...because.../ I love... because... / I feel... because...I think.../ I believe.../ I agree/disagree.../ I'm afraid .../ I don't like... I don't believe.../ I have no idea.../ Oh no! etc.

## สาระที่ 1 ภาษาเพื่อการสื่อสาร

มาตรฐาน ต 1.3 นำเสนอข้อมูลข่าวสาร ความคิดรวบยอด และความคิดเห็นในเรื่องต่างๆ โดยการพูด และการเขียน

ตาราง 2.8 ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 1 ภาษาเพื่อการสื่อสาร มาตรฐาน ต 1.3

ชั้น	ตัวชี้วัด	สาระการเรียนรู้แกนกลาง
ม.3	1. พูดและเขียนบรรยายเกี่ยวกับตนเอง ประสบการณ์ ข่าว/เหตุการณ์ / เรื่อง/ ประเด็นต่างๆ ที่อยู่ในความสนใจของสังคม	การบรรยายเกี่ยวกับตนเอง ประสบการณ์ ข่าว/เหตุการณ์/ ประเด็นที่อยู่ในความสนใจของสังคม เช่น การเดินทาง การรับประทานอาหาร การเล่นกีฬา/ดนตรี การฟังเพลง การอ่าน หนังสือ การท่องเที่ยว การศึกษา สภาพสังคม เศรษฐกิจ
	2. พูดและเขียนสรุปใจความสำคัญ/แก่นสาระ หัวข้อเรื่องที่ได้จากการวิเคราะห์เรื่อง/ ข่าว/เหตุการณ์/สถานการณ์ที่อยู่ในความสนใจของสังคม	การจับใจความสำคัญ/แก่นสาระ หัวข้อเรื่อง การวิเคราะห์เรื่อง/ ข่าว/เหตุการณ์/สถานการณ์ที่อยู่ในความสนใจ เช่น ประสบการณ์ เหตุการณ์ สถานการณ์ต่างๆ ภาพยนตร์ กีฬา ดนตรี เพลง
	3. พูดและเขียนแสดงความคิดเห็นเกี่ยวกับ กิจกรรม ประสบการณ์ และเหตุการณ์ พร้อมทั้งให้เหตุผลประกอบ	การแสดงความคิดเห็น และการให้เหตุผลประกอบเกี่ยวกับกิจกรรม ประสบการณ์ และเหตุการณ์

## สาระที่ 2 ภาษาและวัฒนธรรม

มาตรฐาน ต 2.1 เข้าใจความสัมพันธ์ระหว่างภาษากับวัฒนธรรมของเจ้าของภาษา และนำไปใช้ได้ อย่างเหมาะสมกับกาลเทศะ

ตาราง 2.9 ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 2 ภาษาและวัฒนธรรม มาตรฐาน ต 2.1

ชั้น	ตัวชี้วัด	สาระการเรียนรู้แกนกลาง
ม.3	1. เลือกใช้ภาษา น้ำเสียง และกิริยา ท่าทาง เหมาะกับบุคคลและโอกาส ตาม มารยาทสังคม และวัฒนธรรมของเจ้าของภาษา	การเลือกใช้ภาษา น้ำเสียง และกิริยาท่าทางในการสนทนา ตาม มารยาทสังคมและวัฒนธรรมของเจ้าของภาษา เช่น การขอบคุณ ขอโทษ การชมเชย การใช้สีหน้าท่าทางประกอบ การพูด ขณะ แนะนำตนเอง การสัมผัสมือ การโบกมือ การแสดงความ รู้สึกชอบ/ไม่ชอบ การกล่าวอวยพร การแสดงอาการตอบรับหรือปฏิเสธ
	2. อธิบายเกี่ยวกับชีวิตความเป็นอยู่ ขนบธรรมเนียมและประเพณีของเจ้าของภาษา	ชีวิตความเป็นอยู่ ขนบธรรมเนียมและประเพณีของเจ้าของภาษา
	3. เข้าร่วม/จัดกิจกรรมทางภาษาและ วัฒนธรรมตามความสนใจ	กิจกรรมทางภาษาและวัฒนธรรม เช่น การเล่นเกม การร้องเพลง การเล่านิทาน บทบาทสมมุติ วันขอบคุณพระเจ้า วันคริสต์มาส วันขึ้นปีใหม่ วันวาเลนไทน์

## สาระที่ 2 ภาษาและวัฒนธรรม

มาตรฐาน ต 2.2 เข้าใจความเหมือนและความแตกต่างระหว่างภาษาและวัฒนธรรมของเจ้าของภาษากับภาษาและวัฒนธรรมไทย และนำมาใช้อย่างถูกต้องและเหมาะสม

ตาราง 2.10 ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 2 ภาษาและวัฒนธรรม มาตรฐาน ต 2.2

ชั้น	ตัวชี้วัด	สาระการเรียนรู้แกนกลาง
ม.3	1. เปรียบเทียบและอธิบายความเหมือนและความแตกต่างระหว่างการออกเสียง ประโยคชนิดต่างๆ และการลำดับคำตามโครงสร้างประโยคของภาษาต่างประเทศ และภาษาไทย	การเปรียบเทียบและการอธิบายความเหมือน/ความแตกต่างระหว่างการออกเสียงประโยคชนิดต่างๆ ของเจ้าของภาษากับของไทย การใช้เครื่องหมายวรรคตอนและการลำดับคำตามโครงสร้างประโยคของภาษาต่างประเทศและภาษาไทย
	2. เปรียบเทียบและอธิบายความเหมือนและความแตกต่างระหว่างชีวิตความเป็นอยู่และวัฒนธรรมของเจ้าของภาษากับของไทย และนำไปใช้อย่างเหมาะสม	การเปรียบเทียบและการอธิบายความเหมือนและความแตกต่างระหว่างชีวิตความเป็นอยู่และวัฒนธรรมของเจ้าของภาษากับของไทย การนำวัฒนธรรมของเจ้าของภาษาไปใช้

## สาระที่ 3 ภาษากับความสัมพันธ์กับกลุ่มสาระการเรียนรู้อื่น

มาตรฐาน ต 3.1 ใช้ภาษาต่างประเทศในการเชื่อมโยงความรู้กับกลุ่มสาระการเรียนรู้อื่น และเป็นพื้นฐานในการพัฒนา แสวงหาความรู้ และเปิดโลกทัศน์ของตน

ตาราง 2.11 ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 3 ภาษากับความสัมพันธ์กับกลุ่มสาระการเรียนรู้อื่น มาตรฐาน ต 3.1

ชั้น	ตัวชี้วัด	สาระการเรียนรู้แกนกลาง
ม.3	1. ค้นคว้า รวบรวม และสรุปข้อมูล/ข้อเท็จจริงที่เกี่ยวข้องกับกลุ่มสาระการเรียนรู้จากแหล่งเรียนรู้ และนำเสนอด้วยการพูดและการเขียน	การค้นคว้า การรวบรวม การสรุป และการนำเสนอข้อมูล/ข้อเท็จจริงที่เกี่ยวข้องกับกลุ่มสาระการเรียนรู้

#### สาระที่ 4 ภาษากับความสัมพันธ์กับชุมชนและโลก

มาตรฐาน ต 4.1 ใช้ภาษาต่างประเทศในสถานการณ์ต่างๆ ทั้งในสถานศึกษา ชุมชน และสังคม

ตาราง 2.12 ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 4 ภาษา  
 กับความสัมพันธ์ กับชุมชนและโลก มาตรฐาน ต 4.1

ชั้น	ตัวชี้วัด	สาระการเรียนรู้แกนกลาง
ม.3	1. ใช้ภาษาสื่อสารในสถานการณ์จริง/สถานการณ์จำลองที่เกิดขึ้นในห้องเรียน สถานศึกษา ชุมชน และสังคม	การใช้ภาษาสื่อสารในสถานการณ์จริง/สถานการณ์จำลองที่เกิดขึ้นในห้องเรียน สถานศึกษา ชุมชน และสังคม

#### สาระที่ 4 ภาษากับความสัมพันธ์กับชุมชนและโลก

มาตรฐาน ต 4.2 ใช้ภาษาต่างประเทศเป็นเครื่องมือพื้นฐานในการศึกษาต่อ การประกอบอาชีพ และการแลกเปลี่ยนเรียนรู้กับสังคมโลก

ตาราง 2.13 ตัวชี้วัดและสาระการเรียนรู้แกนกลางระดับชั้นมัธยมศึกษาปีที่ 3 สาระที่ 4 ภาษา  
 กับความสัมพันธ์ กับชุมชนและโลก มาตรฐาน ต 4.2

ชั้น	ตัวชี้วัด	สาระการเรียนรู้แกนกลาง
ม.3	1. ใช้ภาษาต่างประเทศในการสืบค้น/ค้นคว้า รวบรวม และสรุปความรู้/ข้อมูลต่างๆ จากสื่อและแหล่งการเรียนรู้ต่างๆ ในการศึกษาต่อและประกอบอาชีพ	การใช้ภาษาต่างประเทศในการสืบค้น/การค้นคว้าความรู้/ข้อมูลต่างๆ จากสื่อและแหล่งการเรียนรู้ต่างๆ ในการศึกษาต่อและประกอบอาชีพ
	2. เผยแพร่/ประชาสัมพันธ์ ข้อมูล ข่าวสารของโรงเรียน ชุมชน และท้องถิ่น เป็นภาษาต่างประเทศ	การใช้ภาษาต่างประเทศในการเผยแพร่/ประชาสัมพันธ์ ข้อมูลข่าวสารของโรงเรียน ชุมชน และท้องถิ่น เช่น การทำหนังสือเล่มเล็กแนะนำโรงเรียน ชุมชน และท้องถิ่น การทำแผ่นปลิว ป้ายคำขวัญ คำเชิญชวนแนะนำโรงเรียนและสถานที่สำคัญในชุมชนและท้องถิ่น การนำเสนอข้อมูลข่าวสารในโรงเรียน ชุมชน และท้องถิ่น เป็นภาษาอังกฤษ

#### คุณภาพผู้เรียนจบชั้นมัธยมศึกษาปีที่ 3

- ปฏิบัติตามคำขอร้อง คำแนะนำ คำชี้แจง และคำอธิบายที่ฟังและอ่าน อ่านออกเสียงข้อความ ข่าว โฆษณา นิทาน และบทร้อยกรองสั้นๆ ถูกต้องตามหลักการอ่าน ระบุ/เขียนสิ่งที่ไม่ใช่ความเรียงรูปแบบต่างๆ สัมพันธ์กับประโยคและข้อความที่ฟังหรืออ่าน เลือก/ระบุหัวข้อเรื่อง ใจความ

สำคัญ รายละเอียดสนับสนุน และแสดงความคิดเห็นเกี่ยวกับเรื่องที่ฟังและอ่านจากสื่อประเภทต่างๆ พร้อมทั้งให้เหตุผลและยกตัวอย่างประกอบ

- สนทนาและเขียนโต้ตอบข้อมูลเกี่ยวกับตนเองและเรื่องต่างๆ ใกล้ตัว สถานการณ์ ข่าว เรื่องที่อยู่ในความสนใจของสังคมและสื่อสารอย่างต่อเนื่องและเหมาะสม ใช้คำขอร้อง คำชี้แจง และคำอธิบาย ให้คำแนะนำอย่างเหมาะสม พูดและเขียนแสดงความต้องการ เสนอและให้ความช่วยเหลือตอบรับและปฏิเสธ การให้ความช่วยเหลือ พูดและเขียนเพื่อขอและให้ข้อมูล บรรยาย อธิบาย เปรียบเทียบ และแสดงความคิดเห็นเกี่ยวกับเรื่องที่ฟังหรืออ่านอย่างเหมาะสม พูดและเขียนบรรยายความรู้สึกและความคิดเห็นของตนเองเกี่ยวกับเรื่องต่างๆ กิจกรรม ประสบการณ์ และข่าว/เหตุการณ์ พร้อมทั้งให้เหตุผลประกอบอย่างเหมาะสม

- พูดและเขียนบรรยายเกี่ยวกับตนเอง ประสบการณ์ ข่าว/เหตุการณ์/เรื่อง/ประเด็นต่างๆ ที่อยู่ในความสนใจของสังคม พูดและเขียนสรุปใจความสำคัญ/แก่นสาระ หัวข้อเรื่องที่ได้จากการวิเคราะห์เรื่อง/ข่าว/เหตุการณ์/สถานการณ์ที่อยู่ในความสนใจ พูดและเขียนแสดงความคิดเห็นเกี่ยวกับกิจกรรม ประสบการณ์ และเหตุการณ์ พร้อมให้เหตุผลประกอบ

- เลือกใช้ภาษา น้ำเสียง และกิริยาท่าทางเหมาะสมกับบุคคลและโอกาส ตามมารยาทสังคมและวัฒนธรรมของเจ้าของภาษา อธิบายเกี่ยวกับชีวิตความเป็นอยู่ ขนบธรรมเนียมและประเพณี

ของเจ้าของภาษา เข้าร่วม/จัดกิจกรรมทางภาษาและวัฒนธรรมตามความสนใจ

- เปรียบเทียบ และอธิบายความเหมือนและความแตกต่างระหว่างการออกเสียงประโยคชนิดต่างๆ และการลำดับคำตามโครงสร้างประโยคของภาษาต่างประเทศและภาษาไทย เปรียบเทียบ และ อธิบายความเหมือนและความแตกต่างระหว่างชีวิตความเป็นอยู่และวัฒนธรรมของเจ้าของภาษากับของไทย และนำไปใช้อย่างเหมาะสม

- ค้นคว้า รวบรวม และสรุปข้อมูล/ข้อเท็จจริงที่เกี่ยวข้องกับกลุ่มสาระการเรียนรู้อื่นจากแหล่งการเรียนรู้ และนำเสนอด้วยการพูดและการเขียน

- ใช้ภาษาสื่อสารในสถานการณ์จริง/สถานการณ์จำลองที่เกิดขึ้นในห้องเรียน สถานศึกษา ชุมชน และสังคม

- ใช้ภาษาต่างประเทศในการสืบค้น/ค้นคว้า รวบรวม และสรุปความรู้/ข้อมูลต่างๆ จากสื่อและแหล่งการเรียนรู้ต่างๆ ในการศึกษาต่อและประกอบอาชีพ เผยแพร่/ประชาสัมพันธ์ข้อมูลข่าวสารของโรงเรียน ชุมชน และท้องถิ่น เป็นภาษาต่างประเทศ

- มีทักษะการใช้ภาษาต่างประเทศ (เน้นการฟัง-พูด-อ่าน-เขียน) สื่อสารตามหัวเรื่องเกี่ยวกับตนเอง ครอบครัว โรงเรียน สิ่งแวดล้อม อาหาร เครื่องดื่ม เวลาว่างและนันทนาการ สุขภาพและสวัสดิการ การซื้อ-ขาย ลมฟ้าอากาศ การศึกษาและอาชีพ การเดินทางท่องเที่ยว การบริการ สถานที่

ภาษา และวิทยาศาสตร์และเทคโนโลยี ภายในวงคำศัพท์ประมาณ 2,100 – 2,250 คำ (คำศัพท์ที่เป็นนามธรรมมากขึ้น)

- ใช้ประโยคผสมและประโยคซับซ้อน (Complex Sentences) สื่อความหมายตามบริบทต่างๆ ในการสนทนาทั้งที่เป็นทางการและไม่เป็นทางการ

## ตอนที่ 5 กรอบแนวคิดการวิจัย

กรอบแนวคิดในการวิจัย เรื่อง ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ประกอบด้วยหลายส่วนตามระยะการวิจัย เริ่มจากศึกษาหลักการและแนวคิด แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องของการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด จากนั้นนำข้อมูลที่ได้มาพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดเพื่อทดลองระบบและประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีรายละเอียด ดังนี้

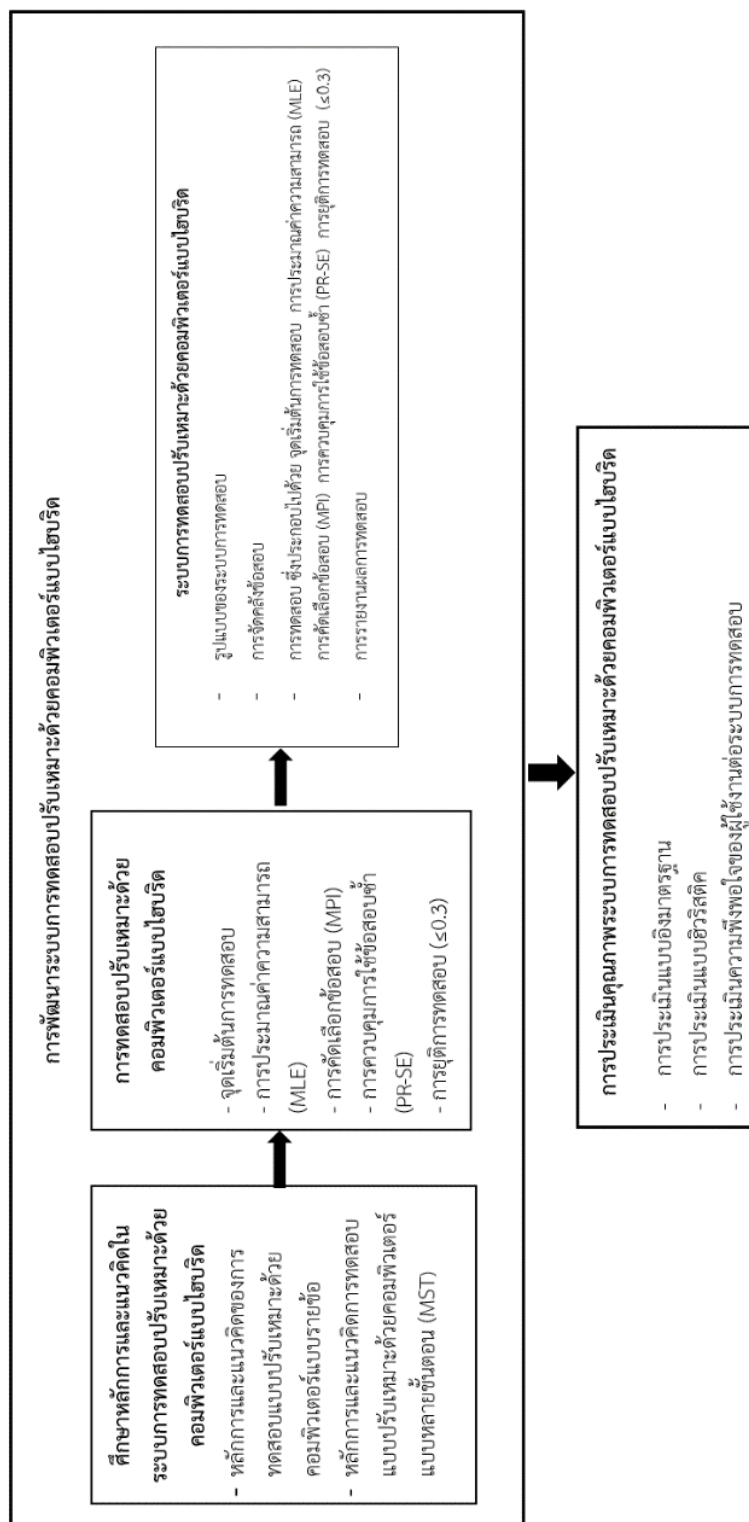
1. ศึกษาหลักการการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด จากสังเคราะห์เอกสาร และงานวิจัยที่เกี่ยวข้อง ซึ่งการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดเป็นการรวมรูปแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดจะประกอบด้วย 2 ส่วน คือ การทดสอบในส่วนแรกใช้ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และในส่วนต่อไปจะใช้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ สำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด จุดเริ่มต้นการทดสอบจะใช้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) โดยการเลือกชุดข้อสอบที่กำหนดไว้ จากนั้นระบบจะคัดเลือกชุดข้อสอบถัดไปที่มีความยากใกล้เคียงกับความสามารถของผู้สอบ เมื่อผู้สอบทำชุดข้อสอบตามระดับความสามารถแล้ว ระบบจะคัดเลือกข้อสอบถัดไปเป็นรายข้อ ซึ่งเป็นส่วนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ จนถึงเกณฑ์การยุติการทดสอบ ตามแนวคิดของ Wang et al. (2016)

การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีการคัดเลือกวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum priority index : MPI) และวิธีการควบคุมการใช้ข้อสอบซ้ำด้วยวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PR-SE) การพัฒนาคัดเลือกข้อสอบจากแบบฝึกพัฒนาภาษาอังกฤษ ของสถาบันภาษาอังกฤษ สำนักงานการศึกษาขั้นพื้นฐาน ที่เก็บรวบรวมข้อมูลจากนักเรียนชั้นมัธยมศึกษาปีที่ 3 มีการตรวจสอบและเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบ และการวิเคราะห์คุณภาพของข้อสอบรายข้อตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ (3PL) เพื่อให้ได้สารสนเทศที่มีความถูกต้อง หลังจากนั้นออกแบบ

และพัฒนาระบบการทดสอบด้วยคอมพิวเตอร์แบบไฮบริดซึ่งประกอบด้วย 1) รูปแบบของระบบการทดสอบ 2) การทดสอบ ซึ่งประกอบไปด้วย จุดเริ่มต้นการทดสอบ การประมาณค่าความสามารถ การคัดเลือกข้อสอบ การควบคุมการใช้ข้อสอบซ้ำ การยุติการทดสอบ และ 3) การรายงานผลการทดสอบ

2. การตรวจสอบคุณภาพของระบบการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ได้ประเมินคุณภาพระบบ 3 แบบ คือ การประเมินแบบอิงมาตรฐาน (evaluation standard) การประเมินระบบแบบฮิวริสติก (heuristic evaluation) โดยผู้เชี่ยวชาญ และการประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยผู้ทดลองใช้ระบบ





ภาพ 2.6 กรอบแนวคิดการวิจัยการพัฒนาระบบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า



### บทที่ 3

#### วิธีดำเนินการวิจัย

การวิจัย เรื่อง การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า ได้ดำเนินการวิจัยจำแนกออกเป็น 2 ระยะ

ระยะที่ 1 การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ระยะที่ 2 การประเมินคุณภาพระบบการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

โดยแต่ละระยะมีรายละเอียดของวิธีดำเนินการวิจัยดังนี้

#### ระยะที่ 1 การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด แบ่งออกเป็น 3 ตอน คือ 1) การพัฒนาค้างข้อสอบ 2) การออกแบบพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด และ 3) การทดลองใช้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีรายละเอียดดังนี้

#### ตอนที่ 1 การพัฒนาค้างข้อสอบ

1. การรวบรวมข้อสอบ ผู้วิจัยได้ทำหนังสือขออนุญาตใช้แบบฝึกพัฒนาทักษะภาษาอังกฤษจากสำนักงานการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ หลังจากนั้นรวบรวมข้อสอบในแบบฝึกพัฒนาทักษะภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่เผยแพร่ในเว็บไซต์ของสถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ตั้งแต่ปี พ.ศ. 2556 – 2559 จำนวน 640 ข้อ ที่มีรูปแบบของคำตอบเป็นแบบ 4 ตัวเลือก โดยมีการมุ่งวัดความรู้ทางภาษาอังกฤษซึ่งประกอบด้วย 4 ด้าน ได้แก่ Conversation, Grammar and Vocabulary, Reading Comprehension และ Miscellaneous ซึ่งมีผังการออกข้อสอบจำแนกตามตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ดังตาราง 3.1

ตาราง 3.1 ผังข้อสอบจำแนกตามตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้  
ภาษาต่างประเทศ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

ด้าน	ภาษาเพื่อการสื่อสาร	ภาษาและวัฒนธรรม	ภาษาและความสัมพันธ์กับ กลุ่มสาระการเรียนรู้ อื่น	ภาษาและความสัมพันธ์กับ ชุมชนและโลก	จำนวน (ข้อ)
Conversation	136	-	-	-	136
Grammar	76	130	-	-	206
Reading	186	-	-	-	186
Miscellaneous	62	50	-	-	112
<b>รวม</b>	<b>460</b>	<b>180</b>	<b>-</b>	<b>-</b>	<b>640</b>

ตัวอย่างข้อสอบภาษาอังกฤษในแต่ละด้าน ดังนี้



### ตัวอย่างข้อสอบด้าน Conversation

Directions: Read each situation and the dialogue carefully and choose the appropriate expression to complete the dialogue.

#### ตัวอย่าง 1

Situation: Two friends meet at the bus stop.

Linda: Hi, how're you doing?

Mark: \_\_\_\_\_ I enjoy my new job.

- A. I'm so-so.
- B. Pretty good.
- C. I'm not well.
- D. Nothing much.

เฉลย B

#### ตัวอย่าง 2

Situation: At the clothes shop.

Jesse: Excuse me. I bought this shirt here yesterday. \_\_\_\_\_

Shop assistant: Why, sir?

Jesse: Because it's too small for me.

Shop assistant: Have you got the receipt?

Jesse: Yes. Here it is. It cost 40 dollars.

- A. Can I change it, please?
- B. May I have a new shirt, please?
- C. Can you give me a change, please?
- D. May I have a small change, please?

เฉลย A

ภาพ 3.1 ตัวอย่างข้อสอบด้าน Conversation

ตัวอย่างข้อสอบด้าน Grammar

Directions: Complete the following passage by choosing the best answer.

ตัวอย่าง 1

Eiffel Tower is one of the world's most popular \_\_\_\_\_. It is named after its architect, Gustave Eiffel, and was in 1889 for the Universal Exposition. The tower is 1,062 feet high and offers fantastic views of Paris. Very often, visitors have to wait in line to take the elevator to the observation deck. At night, the tower itself becomes a very illuminated sculpture.

- A. destiny
- B. demand
- C. descriptions
- D. destinations

เฉลย D

ตัวอย่าง 2

To: SophieK@space.net

Subject: Ha long Bay, Vietnam

Hi, Sophie,

I'm writing this email from Hanoi, the bustling \_\_\_\_\_ of Vietnam. Tomorrow I am going to take a cruise on a junk boat to Ha Long Bay. The bay is one of Southeast Asia's magnificent natural splendors.

Ha Long Bay's beauty comes not only from the blue clear sky and the calm sea, but also from its many caves. Dau Go cave is the largest. Can you imagine yourself in this amazing paradise?

Anna

- A. capital
- B. colony
- C. landmark
- D. landscape

เฉลย A

ภาพ 3.2 ตัวอย่างข้อสอบด้าน Grammar

### ตัวอย่างข้อสอบด้าน Reading

Directions: Read each passage carefully and choose the correct answer the questions.

#### ตัวอย่าง 1

#### Mississippi crash kills three

ST. PAUL PARK. Minnesota – A small airplane trying to return to the airport because of an open cargo door hit some trees on Wednesday and crashed into the Mississippi River, killing the pilot and two passengers.

Two other passengers survived by swimming about 15 metres to an island, where they were rescued by a tugboat.

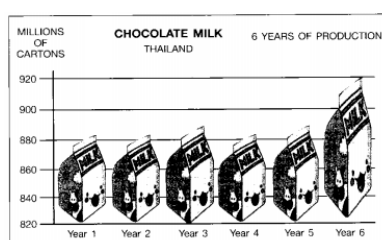
The wreckage of the single-engine Piper Saratoga, with the three bodies inside, was pulled from the river on Wednesday afternoon. The plane was missing a wing. - Associated Press.

What is the news mainly about?

- A. The pilot was killed in the river.
- B. The plane hit some trees at the airport.
- C. The plane fell into the Mississippi River.
- D. The passengers were unable to swim to an island.

เฉลย C

#### ตัวอย่าง 2



According to the chart, which statement is TRUE?

- A. Milk production in Year 7 must drop exactly.
- B. 2 Milk production in Year 1 and 2 are the same amount.
- C. Milk production rose at a steady rate from Year 1 to Year 6.
- D. Milk production dropped continuously from Year 1 t Year 6.

เฉลย B

### ภาพ 3.3 ตัวอย่างข้อสอบด้าน Reading

ตัวอย่างข้อสอบด้าน Miscellaneous

Directions: Choose the correct answer which has the same meaning as the given sentence.

ตัวอย่าง 1

I can speak neither Japanese nor French.

- A. I can speak French.
- B. I can speak Japanese.
- C. I cannot speak Japanese and French.
- D. I cannot speak any languages but French.

เฉลย C

ตัวอย่าง 2

Directions: Choose the answer which is grammatically incorrect.

A dentist in the UK had to take out all of the baby's (A) teeth from a (B) two-year-old child (C) because the teeth (D) are rotten.

เฉลย D

ภาพ 3.4 ตัวอย่างข้อสอบด้าน Miscellaneous

ซึ่งข้อสอบทั้งหมดนี้วัดตามตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 นำข้อสอบมาจัดชุดข้อสอบ จำนวน 9 ชุด (ชุดละ 70 ข้อ จำนวน 8 ชุด และชุดละ 80 ข้อ จำนวน 1 ชุด) โดยจัดชุดแบบสอบไปทดสอบกับนักเรียนกลุ่มตัวอย่างซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา

กำหนดกลุ่มตัวอย่างและการเก็บรวบรวมข้อมูลที่ใช้ในตรวจสอบคุณภาพของข้อสอบเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาด้วยวิธีการสุ่มตัวอย่างหลายขั้นตอน (multistage random) โดยการสุ่มอย่างง่าย จำแนกตามภูมิภาค จากแต่ละภาคสุ่มจังหวัด จากแต่ละจังหวัดเลือกโรงเรียนขนาดใหญ่ มีขั้นตอนการสุ่มดังนี้

1) กำหนดขนาดกลุ่มตัวอย่างนักเรียนที่กำลังศึกษาอยู่ในมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา โดยการวิจัยในครั้งนี้ขึ้นอยู่กับโมเดลการตอบสนอง

ข้อสอบแบบ 3 พารามิเตอร์ ซึ่งมีแนวคิดที่โดยขนาดกลุ่มตัวอย่างที่เหมาะสมและเพียงพอต่อการวิเคราะห์ค่าพารามิเตอร์สำหรับ 3 พารามิเตอร์ จำนวน 1,000 คน (Foley, 2010)

2) จำแนกโรงเรียนตามภูมิภาค ได้ 6 ภูมิภาค ได้แก่ ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคกลาง ภาคตะวันออก ภาคตะวันตก และภาคใต้ จากแต่ละภาคสุ่มจังหวัดอย่างง่าย และจากแต่ละจังหวัดเลือกโรงเรียนขนาดใหญ่ จากแต่ละโรงเรียนสุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 3

**ตาราง 3.2 การดำเนินการทดสอบกับกลุ่มตัวอย่าง**

ภูมิภาค	จังหวัด	โรงเรียน	สำนักงานเขตพื้นที่การศึกษา	จำนวนนักเรียน (คน)
ภาคเหนือ	ลำปาง	โรงเรียนบุญวาทย์วิทยาลัย	35	198
ภาคตะวันออกเฉียงเหนือ	นครราชสีมา	โรงเรียนสุรนารีวิทยา	31	204
ภาคกลาง	กรุงเทพมหานคร	โรงเรียนเศรษฐบุทรบำเพ็ญ	2	178
ภาคตะวันออก	ระยอง	โรงเรียนแก่ง "วิทยสถาวร"	18	140
ภาคตะวันตก	กาญจนบุรี	โรงเรียนท่ามะกาวิทยาคม	8	190
ภาคใต้	ภูเก็ต	โรงเรียนสตรีภูเก็ต	14	199
จำนวนนักเรียน				1,109 คน

3) นัฒหมายกำหนดการในการขอความร่วมมือเพื่อเก็บข้อมูลการวิจัยจากโรงเรียนทั้ง 6 โรงเรียน ได้แก่ โรงเรียนบุญวาทย์วิทยาลัย, โรงเรียนสุรนารีวิทยา, โรงเรียนเศรษฐบุทรบำเพ็ญ, โรงเรียนแก่ง "วิทยสถาวร", โรงเรียนท่ามะกาวิทยาคม และ โรงเรียนสตรีภูเก็ต

4) ออกแบบกระดาษคำตอบภายใต้การใช้โปรแกรมสำเร็จรูปในการตรวจข้อสอบ คือ Zip Grade ซึ่งเป็นกระดาษคำตอบแบบฟลนที่ออกแบบขึ้นภายใต้โปรแกรมสำเร็จรูป จากเว็บไซต์ [www.zipgrade.com](http://www.zipgrade.com) ซึ่งต้องตรวจข้อสอบผ่าน Zip Grade Application บนมือถือ

Name

School

Class

ZIPGRADE.COM

1 (A) (B) (C) (D) 20 (A) (B) (C) (D) 39 (A) (B) (C) (D) 58 (A) (B) (C) (D)

2 (A) (B) (C) (D) 21 (A) (B) (C) (D) 40 (A) (B) (C) (D) 59 (A) (B) (C) (D)

3 (A) (B) (C) (D) 22 (A) (B) (C) (D) 41 (A) (B) (C) (D) 60 (A) (B) (C) (D)

4 (A) (B) (C) (D) 23 (A) (B) (C) (D) 42 (A) (B) (C) (D) 61 (A) (B) (C) (D)

5 (A) (B) (C) (D) 24 (A) (B) (C) (D) 43 (A) (B) (C) (D) 62 (A) (B) (C) (D)

6 (A) (B) (C) (D) 25 (A) (B) (C) (D) 44 (A) (B) (C) (D) 63 (A) (B) (C) (D)

7 (A) (B) (C) (D) 26 (A) (B) (C) (D) 45 (A) (B) (C) (D) 64 (A) (B) (C) (D)

8 (A) (B) (C) (D) 27 (A) (B) (C) (D) 46 (A) (B) (C) (D) 65 (A) (B) (C) (D)

9 (A) (B) (C) (D) 28 (A) (B) (C) (D) 47 (A) (B) (C) (D) 66 (A) (B) (C) (D)

10 (A) (B) (C) (D) 29 (A) (B) (C) (D) 48 (A) (B) (C) (D) 67 (A) (B) (C) (D)

11 (A) (B) (C) (D) 30 (A) (B) (C) (D) 49 (A) (B) (C) (D) 68 (A) (B) (C) (D)

12 (A) (B) (C) (D) 31 (A) (B) (C) (D) 50 (A) (B) (C) (D) 69 (A) (B) (C) (D)

13 (A) (B) (C) (D) 32 (A) (B) (C) (D) 51 (A) (B) (C) (D) 70 (A) (B) (C) (D)

14 (A) (B) (C) (D) 33 (A) (B) (C) (D) 52 (A) (B) (C) (D)

Eng Test 1-1 (0107)

ภาพ 3.5 ตัวอย่างกระดาษคำตอบที่ใช้ในการทดสอบ

5) ดำเนินการจัดการทดสอบ โดยกำหนดให้การเก็บข้อมูลในแต่ละครั้งผู้สอบจะได้รับข้อสอบเพียงคนละ 1 ชุด โดยใช้ระยะเวลาในการทดสอบครั้งละ 60 นาที

6) ดำเนินการเก็บรวบรวมกระดาษคำตอบและชุดข้อสอบกลับคืนเพื่อนำไปวิเคราะห์พารามิเตอร์

2. การตรวจสอบและเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบ เพื่อให้ได้สารสนเทศที่มีความถูกต้อง การตรวจสอบและเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบ 3 โมเดล ได้แก่ 1) 1พารามิเตอร์ (1PL) 2) 2 พารามิเตอร์ (2PL) และ 3) 3 พารามิเตอร์ (3PL) เพื่อยืนยันคุณภาพของข้อสอบ โดยพิจารณาจากค่าสถิติ deviance statistic ( $G^2$ ), ค่า Akaike Information Criterion (AIC) และค่า Bayesian information criterion (BIC) (เมษา นวลศรี, 2559; ชัยวิจิตต์ เขียวชนะ, 2552)

3. ตรวจสอบคุณภาพของข้อสอบที่ได้จากการทดสอบนักเรียนกลุ่มตัวอย่าง มาวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ ได้แก่ ค่าความยากของข้อสอบ (b) ค่าอำนาจจำแนกของข้อสอบ (a) และค่าการเดาของข้อสอบ (c) ดำเนินการวิเคราะห์ข้อสอบด้วยโปรแกรม R ในการวิเคราะห์ข้อสอบตาม



ทฤษฎีการตอบสนองข้อสอบ (item response theory) ในการคัดเลือกข้อสอบเพื่อจัดในคลังมีเกณฑ์ในการคัดเลือกข้อสอบตามค่าพารามิเตอร์ของข้อสอบ (Reckase, 2003; ศิริชัย กาญจนาวาสี, 2555) ได้แก่

- ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง  $\infty$
- ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -4 ถึง 4
- ค่าการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.50

คัดเลือกข้อสอบที่มีคุณภาพตามเกณฑ์การคัดเลือกข้อสอบตามค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และ ค่าการเดาของข้อสอบ (c) จัดเก็บเข้าสู่ระบบคลังข้อสอบ

4. การจัดคลังข้อสอบ ในการจัดคลังข้อสอบจะใช้ข้อมูลจากการทดสอบกับนักเรียนกลุ่มตัวอย่าง หลังจากนั้นนำผลการตอบข้อสอบข้อสอบมาวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) ด้วยโปรแกรม R แบบ 3 พารามิเตอร์ได้แก่ ค่าความยาก (b) อำนาจจำแนก (a) และการเดา (c) ซึ่งมีเกณฑ์ในการคัดเลือกข้อสอบในการจัดทำคลังได้แก่ ค่าความยากอยู่ในช่วง -4 ถึง 4 ค่าอำนาจจำแนกอยู่ระหว่าง 0.50 ถึง  $\infty$  และค่าการเดาไม่เกิน 0.50 หลังจากนั้นนำข้อสอบที่ผ่านเกณฑ์มาจัดเข้าคลัง โดยการจัดข้อสอบเข้าสู่คลังมีรายละเอียด ดังนี้

- ในส่วนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จะมีการจัดเตรียมชุดข้อสอบไว้ล่วงหน้า โดยในขั้นตอนที่ 1 จะจัดชุดข้อสอบที่มีความยากปานกลาง (ระดับปานกลาง) ขั้นตอนที่ 2 จะจัดชุดข้อสอบ โดยแบ่งโมดูลออกเป็น 3 ระดับ ได้แก่ ระดับง่าย ระดับปานกลาง และระดับยาก
- ส่วนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ จะจัดสอบเป็นรายข้อ และใช้คอมพิวเตอร์ในการคัดเลือกข้อสอบแต่ละข้อ โดยเริ่มต้นจากข้อสอบข้อที่ง่าย ปานกลาง และยาก

## ตอนที่ 2 การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด แบ่งออกเป็น 3 ส่วน ได้แก่ รูปแบบของระบบทดสอบ การทดสอบ และการรายงานผลการทดสอบ

1) รูปแบบของระบบ รูปแบบของระบบทดสอบการปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีการกำหนดสิทธิ์ในการเข้าถึงในการใช้งานของระบบ โดยแบ่งออกเป็น ผู้ดูแลระบบ (administrator) และผู้ใช้งานระบบ (user) การออกแบบให้ผู้ใช้งานมีการเข้าถึงข้อมูลที่สะดวก รวดเร็วและทันสมัย มีรายละเอียดของการทดสอบ มีส่วนให้ผู้ใช้งานใส่รหัสผู้เข้าสอบและรหัสผ่าน และจัดทำคู่มือการใช้งานเป็นแนวทางในการนำโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์

แบบไฮบริด ไปใช้งานทำให้ผู้ใช้งานทราบถึงลำดับขั้นตอนการใช้โปรแกรมอย่างละเอียดทุกขั้นตอนในการใช้งาน

2) การทดสอบ สำหรับหน้าจอการทดสอบ ประกอบด้วยองค์ประกอบหลัก คือ ข้อสอบพร้อมตัวเลือกสำหรับการทดสอบแต่ละข้อและระยะเวลาสำหรับการทำแบบทดสอบ โดยมีส่วนสำคัญ 5 ส่วน คือ (1) จุดเริ่มต้นการทดสอบ โดยการกำหนดชุดข้อสอบที่มีความยากปานกลาง (2) การประมาณค่าความสามารถด้วยความเป็นไปได้สูงสุด (maximum likelihood : MLE) (3) การคัดเลือกข้อสอบ โดยวิธีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) (4) การควบคุมการใช้ข้อสอบซ้ำที่มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ โดยวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PRSE) และ (5) การยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) น้อยกว่าหรือเท่ากับ 0.3

3) การรายงานผลการทดสอบ สำหรับการรายงานการทดสอบประกอบด้วยข้อมูลส่วนตัวของกลุ่มตัวอย่างในการวิจัย ได้แก่ ชื่อ – นามสกุล อีเมล (E-mail) และโรงเรียน รายงานการทดสอบแบ่งออกเป็น 2 ส่วน คือ 1) รายงานผลการตอบข้อสอบ จำนวนข้อสอบที่ตอบถูก – ผิด ร้อยละ และ ค่ามาตรฐานที่ (T-score) และ 2) กราฟแสดงความสัมพันธ์ของผลการตอบข้อสอบทั้ง 4 ด้าน

เมื่อได้ระบบการทดสอบต้นแบบแล้ว จะดำเนินการส่งให้ผู้เชี่ยวชาญด้านการทดสอบและผู้เชี่ยวชาญด้านเทคโนโลยีสารสนเทศ จำนวน 5 ท่าน ตรวจสอบความเหมาะสมการทำงานของระบบ โดยใช้แบบประเมินระบบแบบอิงมาตรฐาน (standard evaluation) หลังจากนั้นดำเนินการปรับปรุงตามคำแนะนำของผู้เชี่ยวชาญ และนำระบบไปทดลองใช้กับกลุ่มตัวอย่างในการวิจัยต่อไป

### ตอนที่ 3 การทดลองใช้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

การทดลองใช้ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีรายละเอียดดังนี้

#### 1. ขั้นตอนเตรียมการระบบเพื่อนำไปทดลองใช้ มีการดำเนินงานดังนี้

1) ทำหนังสือขอความอนุเคราะห์ทดลองใช้ระบบการทดสอบที่พัฒนาขึ้น พร้อมแนบคู่มือการใช้งานระบบ

2) ติดต่อประสานงานเพื่อดำเนินการเก็บรวบรวมข้อมูลในการทดลองใช้ระบบการทดสอบ พร้อมนัดหมาย วัน เวลาและสถานที่ในการจัดการทดสอบ

3) ติดต่อเพื่อเข้ารับการทดสอบโดยจัดเตรียมเอกสารคู่มือและตรวจสอบความพร้อมของระบบเพื่อเตรียมนำไปทดลองใช้

2. กำหนดกลุ่มตัวอย่างที่ใช้ทดลองระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนพิชัยรัตนาคาร จังหวัดระนอง สังกัดสำนักงานเขตพื้นที่ การศึกษามัธยมศึกษา เขต 14 จำนวน 74 คน

3. ดำเนินการทดสอบและประเมินระบบ มีการดำเนินการทดสอบในห้องคอมพิวเตอร์ของ โรงเรียน โดยใช้คอมพิวเตอร์ 1 เครื่อง ต่อ 1 คน มีขั้นตอนดังนี้

1) อธิบายหลักเกณฑ์ เงื่อนไข และวิธีการทดสอบโดยใช้วิธีการทดสอบปรับเหมาะด้วย คอมพิวเตอร์วิธีไฮบริดผ่านระบบคอมพิวเตอร์ พร้อมทั้งแจกคู่มือการใช้งานที่จัดทำขึ้นก่อนดำเนินการ ทดสอบ

2) นักเรียนกลุ่มตัวอย่างทำการทดสอบโดยใช้ระบบการทดสอบปรับเหมาะด้วย คอมพิวเตอร์ด้วยวิธีไฮบริด โดยทำการลงทะเบียนผ่านระบบ เริ่มดำเนินการทดสอบ จนถึงการยุติ การทดสอบ

การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ผู้วิจัยมีการวิเคราะห์ ข้อมูลมีรายละเอียดดังต่อไปนี้

1) วิเคราะห์คุณภาพของข้อสอบรายข้อ ตามทฤษฎีตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ ได้แก่ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c) ด้วยโปรแกรม R แพคเกจ ltm และการวิเคราะห์ข้อมูลของคะแนนสอบ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่า การเดา (c) โดยใช้สถิติบรรยายได้แก่ ค่าเฉลี่ย ค่าสูงสุด ค่าต่ำสุด ส่วนเบี่ยงเบนมาตรฐาน

2) วิเคราะห์เพื่อตรวจสอบและเปรียบเทียบความสอดคล้องกลมกลืนของโมเดล การตอบสนองข้อสอบ เพื่อให้ได้สารสนเทศที่มีความถูกต้อง การตรวจสอบและเปรียบเทียบ ความสอดคล้องกลมกลืนของโมเดล การตอบสนองข้อสอบ โดยพิจารณาจากค่าสถิติ deviance statistic ( $G^2$ ), ค่า Akaike Information Criterion (AIC) และ ค่า Bayesian information criterion (BIC) เกณฑ์การพิจารณาจากโมเดลใดมีค่าน้อยกว่าจะแสดงถึงโมเดลมีความสอดคล้องกับ ข้อมูลเชิงประจักษ์มากกว่า และยังเป็นตัวบ่งชี้ถึงโมเดลที่คาดว่าดีที่สุดจากการพิจารณาจาก ความแตกต่างระหว่างโมเดล (เมษา นวลศรี, 2559; ชัยวิจิต เชียรชนะ, 2552)

3) วิเคราะห์ข้อมูลพื้นฐานของกลุ่มตัวอย่าง ได้แก่ ความถี่ ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้และความโด่ง

## ระยะที่ 2 การประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ในการประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด แบ่งออกเป็น 3 ตอน ได้แก่ ตอนที่ 1 การประเมินแบบอิงมาตรฐาน (evaluation standard) ตอนที่ 2 การประเมินระบบแบบฮิวริสติก (heuristic evaluation) และ ตอนที่ 3 การประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยมีรายละเอียดดังนี้

### ตอนที่ 1 การประเมินระบบแบบอิงมาตรฐาน (evaluation standard)

การประเมินระบบจากผู้เชี่ยวชาญก่อนนำระบบไปใช้งานด้วยแบบประเมินอิงมาตรฐานตามแนวคิดของ Stufflebeam และคณะ ประกอบด้วยเนื้อหาหลักในการประเมินทั้งสิ้น 4 หมวด ได้แก่ ความเป็นประโยชน์ (utility) ความเป็นไปได้ (feasibility) ความเหมาะสม (propriety) และความถูกต้อง (accuracy) โดยแบบประเมินจะนำไปใช้ประเมินความเหมาะสมของระบบก่อนนำระบบไปใช้จริง ซึ่งมีรายละเอียด ดังนี้ (ศิริชัย กาญจนวาสี, 2554)

- 1) ความเป็นประโยชน์ (utility) เป็นการประกันความเป็นประโยชน์ของระบบในการตอบสนองต่อความต้องการในการใช้สารสนเทศของผู้ที่เกี่ยวข้องให้ครอบคลุม ทันเวลา และมีผลต่อการนำไปใช้
- 2) ความเป็นไปได้ (feasibility) เป็นการประกันความสอดคล้องของระบบกับสภาพความเป็นจริง เหมาะสมกับสถานการณ์ ปฏิบัติได้ ยอมรับได้ ประหยัดและคุ้มค่า
- 3) ความเหมาะสม (propriety) เป็นการประกันว่าระบบมีความเหมาะสมซึ่งเป็นไปตามกฎ ระเบียบ จรรยาบรรณ คำนึงถึงสวัสดิภาพของผู้ที่เกี่ยวข้องและผู้ที่ได้รับผลกระทบ
- 4) ความถูกต้อง (accuracy) เป็นการประกันว่าระบบมีการใช้เทคนิคที่เหมาะสม ให้สารสนเทศที่เพียงพอ

ซึ่งแบบประเมินจะนำไปใช้ในการประเมินระบบหลังจากนำระบบไปใช้ในการทดสอบจริงเรียบร้อยแล้ว

เครื่องมือในการประเมินเป็นรูปแบบมาตรประมาณค่า (rating scale) 5 ระดับ ได้แก่ 1 = น้อยที่สุด 2 = น้อย 3 = ปานกลาง 4 = มาก และ 5 = มากที่สุด โดยมีเกณฑ์ในการแปลผลดังนี้

- |                         |                                                             |
|-------------------------|-------------------------------------------------------------|
| คะแนนเฉลี่ย 1.00 – 1.49 | หมายถึง มีประโยชน์/ เป็นไปได้/ เหมาะสม/ ถูกต้อง/ น้อยที่สุด |
| คะแนนเฉลี่ย 1.50 – 2.49 | หมายถึง มีประโยชน์/ เป็นไปได้/ เหมาะสม/ ถูกต้อง/ น้อย       |
| คะแนนเฉลี่ย 2.50 – 3.49 | หมายถึง มีประโยชน์/ เป็นไปได้/ เหมาะสม/ ถูกต้อง/ ปานกลาง    |
| คะแนนเฉลี่ย 3.50 – 4.49 | หมายถึง มีประโยชน์/ เป็นไปได้/ เหมาะสม/ ถูกต้อง/ มาก        |
| คะแนนเฉลี่ย 4.50 – 5.00 | หมายถึง มีประโยชน์/ เป็นไปได้/ เหมาะสม/ ถูกต้อง/ มากที่สุด  |

แบบประเมินอิงมาตรฐานมีลักษณะเป็นมาตรประมาณค่า 5 ระดับ ประกอบด้วยข้อคำถามจำนวน 14 ข้อ ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินอิงมาตรฐานพบว่า ผ่านเกณฑ์การประเมิน 14 ข้อ โดยมีค่าดัชนี IOC อยู่ระหว่าง 0.6 – 1.00 ทั้งนี้ผู้เชี่ยวชาญด้านการศึกษาจำนวน 5 ท่าน ให้ข้อเสนอแนะในการปรับปรุงข้อคำถามดังรายละเอียดในตาราง 3.3

ตาราง 3.3 ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินอิงมาตรฐาน

ข้อคำถาม	IOC	ข้อเสนอแนะ
<b>ความมีประโยชน์ (utility)</b>		
1. ระบบตอบสนองความต้องการของผู้ใช้งาน	1	
2. การรายงานผลมีความชัดเจน เข้าใจง่าย	1	
3. ระบบการทดสอบสามารถนำไปใช้ประโยชน์ในการพัฒนาทางด้านทักษะภาษาอังกฤษของผู้สอบได้	0.6	
<b>ความเป็นไปได้ (feasibility)</b>		
1. กระบวนการและขั้นตอนของระบบสามารถนำไปปฏิบัติได้จริง	1	
2. ระบบมีการเรียกใช้งานง่าย สะดวก และรวดเร็ว	1	
3. ผลที่ได้จากการทดสอบมีความคุ้มค่า	0.8	
4. ระบบสามารถไปใช้ในการเตรียมความพร้อมก่อนการทดสอบภาษาอังกฤษได้	1	คำถามควรเปลี่ยนเป็นระบบมีความสอดคล้องตามสถานการณ์ปัจจุบัน
<b>ความเหมาะสม (propriety)</b>		
1. มีการระบุการใช้ระบบไว้อย่างชัดเจน	0.8	ระบบการทดสอบมีคู่มือแนะนำการใช้งานที่เหมาะสมเข้าใจง่าย
2. ขอบเขตของเนื้อหาในการทดสอบมีความเหมาะสม	0.8	
3. ระยะเวลาในการทดสอบมีความเหมาะสม	0.8	
4. การรายงานผลการทดสอบมีความครบถ้วน ครอบคลุม	0.8	
<b>ความถูกต้อง (accuracy)</b>		
1. ระบบจัดการทดสอบได้ตรงตามวัตถุประสงค์	1	
2. กระบวนการทดสอบของระบบมีความถูกต้อง	1	
3. การรายงานผลคะแนนมีความถูกต้องตรงตามคุณลักษณะที่แท้จริง	1	

ผู้วิจัยปรับปรุงแก้ไขข้อคำถามตามข้อเสนอแนะของผู้ทรงคุณวุฒิ จากนั้นจึงจัดพิมพ์แบบประเมินอิงมาตรฐานเพื่อให้ผู้เชี่ยวชาญด้านเทคโนโลยีสารสนเทศการศึกษาประเมินคุณภาพ

ของระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ โดยแบบประเมินอิงมาตรฐานมีลักษณะดัง  
ภาคผนวก ค

## ตอนที่ 2 การประเมินระบบแบบฮิวริสติก (heuristic evaluation)

การประเมินระบบจากผู้เชี่ยวชาญก่อนนำระบบไปใช้งานด้วยแบบประเมินฮิวริสติก  
ประกอบด้วยเนื้อหาหลักในการประเมินทั้งสิ้น 10 หมวด ได้แก่

- ความชัดเจนของระบบ (Visibility of system status) ระบบควรจะมีการแจ้งเตือน  
ผู้ใช้งานอยู่ตลอดเวลาว่ากำลังดำเนินการอะไรอยู่ โดยมีการให้ข้อมูลย้อนกลับที่เหมาะสมภายใน  
ระยะเวลาที่สมเหตุสมผล

- ความสอดคล้องของระบบและการใช้งานจริง (Match between system and the  
real world) ระบบควรพูดภาษาเดียวกับผู้ใช้งาน ด้วยคำ ประโยค และแนวคิดซึ่งผู้ใช้งานมี  
ความคุ้นเคยมากกว่าการใช้คำศัพท์ที่เป็นของระบบ โดยให้เป็นไปตามแบบแผนของโลกความเป็นจริง  
ซึ่งการจัดทำข้อมูลต้องมีลำดับที่เป็นธรรมชาติและมีตรรกะ

- การควบคุมและอิสระของผู้ใช้งาน (User control and freedom) ผู้ใช้งานควรมี  
อิสระในการพัฒนาแนวทางในการใช้งานของตนเอง เลือกและเรียงลำดับเครื่องมือ และสามารถ  
ยกเลิกหรือทำกิจกรรมซ้ำที่ได้ทำไปแล้ว มากกว่าที่จะกำหนดระบบให้ผู้ใช้งาน

- ความสอดคล้องและมาตรฐาน (Consistency and standards) ผู้ใช้งานไม่ควรสับสน  
กับคำพูด สถานการณ์ หรือการกระทำที่มีความแตกต่างกัน แต่มีความหมายเหมือนกัน และระบบ  
ควรมีการทำงานที่เป็นแบบแผน

- การป้องกันความผิดพลาด (Error prevention) ข้อความแจ้งเตือนความผิดพลาดควร  
การออกแบบอย่างระมัดระวังเพื่อป้องกันปัญหาที่เกิดขึ้นในการใช้งานครั้งแรก

- การจำข้อมูลมากกว่าการเรียกคืนข้อมูล (Recognition rather than recall)  
การสร้างวัตถุ การกระทำ และตัวเลือกในการใช้งาน ผู้ใช้งานไม่ควรที่จะต้องจำข้อมูลจากส่วนหนึ่ง  
เพื่อไปใช้งานอีกส่วนหนึ่ง คำสั่งในการใช้งานของระบบควรมองเห็นได้ง่ายและมีความเหมาะสม

- ความยืดหยุ่นและประสิทธิภาพการใช้งาน (Flexibility and efficiency of use)  
อนุญาตให้ผู้ใช้งานเลือกการกระทำที่ทำบ่อยและเหมาะกับผู้ใช้งาน โดยการให้ทางเลือกซึ่งหมายถึง  
การเข้าถึงและการดำเนินงานสำหรับผู้ใช้งานที่มีความแตกต่างจากผู้ใช้งานปกติ (เช่น ด้านกายภาพ  
ความสามารถทางสมอง วัฒนธรรม ภาษา เป็นต้น)

- การออกแบบที่สวยงาม (Aesthetic and minimalist design) ข้อความแจ้งเตือนไม่  
ควรประกอบด้วยข้อมูลที่ไม่มีความสอดคล้องหรือไม่มีความต้องการ การแจ้งเตือนที่มีความพิเศษต้อง

นำมาเทียบกับการแจ้งเตือนแบบปกติเพื่อตรวจสอบความสอดคล้อง หากไม่มีความจำเป็นควรลดการแจ้งเตือนให้เหลือเพียงแบบปกติ

- การช่วยเหลือโดยการกู้คืนระบบจากข้อผิดพลาด (Help users recognize, diagnose, and recover from errors) ข้อความแจ้งเตือนควรมีความชัดเจนเพื่อบ่งชี้ปัญหาและการแนะนำอย่างมีโครงสร้างในการแก้ปัญหา ซึ่งควรนำเสนอในภาษาที่เรียบง่าย

- การช่วยเหลือและคู่มือ (Help and documentation) แม้ว่าจะเป็นข้อดีหากระบบสามารถใช้งานโดยไม่ต้องใช้คู่มือ แต่ก็อาจมีความจำเป็นที่จะต้องให้การช่วยเหลือและให้เอกสารประกอบ ทุกองค์ประกอบของสารสนเทศควรง่ายต่อการค้นหาและเน้นในส่วนเครื่องมือของผู้ใช้งานเป็นหลัก มีการจัดเรียงลำดับในการค้นหาและรายการไม่ควรยาวเกินไป

เครื่องมือในการประเมินเป็นรูปแบบมาตราประมาณค่า (rating scale) 5 ระดับ ได้แก่ 1 = น้อยที่สุด 2 = น้อย 3 = ปานกลาง 4 = มาก และ 5 = มากที่สุด โดยมีเกณฑ์ในการแปลผลดังนี้

คะแนนเฉลี่ย 1.00 – 1.49 หมายถึงมีความเหมาะสมระดับน้อยที่สุด

คะแนนเฉลี่ย 1.50 – 2.49 หมายถึงมีความเหมาะสมระดับน้อย

คะแนนเฉลี่ย 2.50 – 3.49 หมายถึงมีความเหมาะสมระดับปานกลาง

คะแนนเฉลี่ย 3.50 – 4.49 หมายถึงมีความเหมาะสมระดับมาก

คะแนนเฉลี่ย 4.50 – 5.00 หมายถึงมีความเหมาะสมระดับมากที่สุด

แบบประเมินอิวิริสติกมีลักษณะเป็นมาตราประมาณค่า 5 ระดับ ประกอบด้วยข้อคำถามจำนวน 13 ข้อ ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินอิงมาตรฐานพบว่า ผ่านเกณฑ์การประเมิน 13 ข้อ โดยมีค่าดัชนี IOC อยู่ระหว่าง 0.6 – 1.00 ทั้งนี้ผู้เชี่ยวชาญด้านการศึกษากำหนด 5 ท่าน ให้ข้อเสนอแนะในการปรับปรุงข้อคำถามดังรายละเอียดในตาราง 3.4

ตาราง 3.4 ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินอิวิริสติก

ข้อคำถาม	IOC	ข้อเสนอแนะ
1. การมองเห็นสถานะของระบบ	0.6	มีการแสดงสถานะของระบบอย่างชัดเจน
2. ความสอดคล้องระหว่างระบบกับโลกของความเป็นจริง	1	ระบบมีความสอดคล้องกับโลกของความเป็นจริง
3. การควบคุมและอิสระของผู้ใช้งาน	0.8	การควบคุมระบบมีประสิทธิภาพและผู้ใช้งานมีความสะดวกในการเข้าถึง
4. ความคงเส้นคงวาและได้มาตรฐานของระบบ	0.8	การทำงานของระบบมีความเสถียร
5. ระบบมีการป้องกันความผิดพลาด	1	
6. ระบบใช้การจำข้อมูลมากกว่าการเรียกคืนข้อมูล	0.8	ระบบมีความสามารถในการจดจำข้อมูลมากกว่าการเรียกข้อมูล
7. ระบบมีความยืดหยุ่นและประสิทธิภาพของการใช้งาน	1	
8. ระบบมีรูปแบบสวยงามและเรียบง่าย	1	
9. ระบบมีการช่วยเหลือผู้ใช้งานระดับตน วินิจฉัยและกู้คืนจากข้อผิดพลาด	1	
10. ระบบมีการช่วยเหลือและคู่มือการใช้งาน	1	ระบบมีการช่วยเหลือและคู่มือการใช้งาน
11. ระบบมีการสนับสนุนและเพิ่มทักษะของผู้ใช้งาน	1	
12. ความพึงพอใจและการยอมรับปฏิสัมพันธ์กับผู้ใช้งาน	0.8	ระบบมีการตรวจสอบความพึงพอใจและการยอมรับปฏิสัมพันธ์กับผู้ใช้งาน
13. ระบบมีการป้องกันข้อมูลส่วนบุคคล	1	

ผู้วิจัยปรับปรุงแก้ไขข้อคำถามตามข้อเสนอแนะของผู้ทรงคุณวุฒิ จากนั้นจึงจัดพิมพ์แบบประเมินอิวิริสติกเพื่อให้ผู้ทรงคุณวุฒิด้านเทคโนโลยีสารสนเทศการศึกษาและผู้ทรงคุณวุฒิด้านการศึกษาประเมินคุณภาพของระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ โดยแบบประเมินอิงมาตรฐานมีลักษณะดังภาพผนวก

### ตอนที่ 3 การประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

การประเมินความพึงพอใจของผู้ใช้งานต่อระบบจะใช้แบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ประกอบด้วยเนื้อหาหลักในการประเมิน 4 องค์ประกอบ ได้แก่ 1) ภาพรวมของระบบ 2) หน้าจอของระบบ 3) การเรียนรู้ และ 4) ความสามารถของระบบ โดยผู้วิจัยได้ดำเนินการพัฒนาแบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดและได้ดำเนินการวิเคราะห์การตรวจสอบความตรงเชิงเนื้อหาด้วยค่าดัชนีความสอดคล้อง (IOC) โดยในการตรวจสอบคุณภาพ



แบบประเมินอาศัยดุลพินิจจากผู้เชี่ยวชาญด้านการศึกษากำหนด 5 ท่าน ซึ่งผู้เชี่ยวชาญได้ตัดสินรายการในรูปแบบตัวเลข โดย 1 = เหมาะสม 0 = ไม่แน่ใจ และ -1 = ไม่เหมาะสม จากนั้นนำค่าที่ได้มาคำนวณหาค่าดัชนีความสอดคล้อง (IOC) สำหรับเกณฑ์ที่ใช้ตัดสินของรายการประเมินคือ ค่าดัชนีความสอดคล้อง (IOC) ต้องมีค่าอยู่ในช่วงระหว่าง 0.5 -1.00 สำหรับรายการประเมินที่มีค่าดัชนีความสอดคล้อง (IOC) ต่ำกว่า 0.50 ผู้วิจัยจะดำเนินการปรับปรุงแก้ไขรายการประเมินดังกล่าวให้มีความเหมาะสมตามข้อเสนอแนะ ในกรณีที่ค่า IOC มีค่ามากกว่า 0.50 แต่มีข้อเสนอแนะจากผู้เชี่ยวชาญ ดำเนินการปรับปรุงแบบประเมินตามข้อเสนอแนะเพื่อให้แบบประเมินมีความสมบูรณ์มากยิ่งขึ้นและนำเสนอให้อาจารย์ที่ปรึกษาพิจารณา หลังจากอาจารย์ที่ปรึกษาพิจารณาเรียบร้อยแล้ว นำแบบประเมินไปใช้ในการประเมินคุณภาพระบบต่อไป

แบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดมีลักษณะเป็นมาตราประมาณค่า 5 ระดับ ประกอบด้วยข้อคำถาม 10 ข้อ ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด พบว่า ผ่านเกณฑ์การประเมิน 10 ข้อ โดยมีค่าดัชนีความสอดคล้อง (IOC) อยู่ระหว่าง 0.8 – 1.00 ทั้งนี้ผู้เชี่ยวชาญด้านการศึกษากำหนด 5 ท่าน ให้ข้อเสนอแนะในการปรับปรุงข้อคำถาม ดังรายละเอียดในตาราง 3.5

**ตาราง 3.5 ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด**

ข้อคำถาม	IOC	ข้อเสนอแนะ
1. การทำงานของระบบสะดวก ง่าย ต่อการใช้งาน และไม่ซับซ้อน	1	
2. ผู้ใช้งานสามารถเรียนรู้การใช้งานและ การทำงานของระบบด้วยตนเอง	0.8	
3. กระบวนการทำงานของระบบมีความรวดเร็วในการเรียกใช้บริการ	1	
4. รูปแบบของข้อความ ส่วนประกอบต่าง ๆ บนหน้าจอมีความเหมาะสม	1	
5. ลำดับการแสดงผลภาพและส่วนประกอบต่าง ๆ บนหน้าจอ	0.8	ลำดับการแสดงผลภาพและส่วนประกอบต่าง ๆ บนหน้าจอเป็นลำดับขั้นตอน
6. การออกแบบการใช้สีที่เหมาะสมต่อการใช้งาน	1	
7. ความรวดเร็วในการรายงานผลการทดสอบ	1	
8. การรายงานผลการทดสอบมีความชัดเจน เข้าใจง่าย	1	
9. ผลการทดสอบนำไปใช้ในการเตรียมความพร้อมก่อนการทดสอบภาษาอังกฤษได้	1	
10. ความรู้สึกโดยรวมที่มีต่อระบบ	1	

ผู้วิจัยปรับปรุงแก้ไขข้อคำถามตามข้อเสนอแนะของผู้ทรงคุณวุฒิ จากนั้นจึงจัดพิมพ์แบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เพื่อให้ผู้ใช้งานได้ประเมินหลังจากทดลองใช้ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ดังภาคผนวก

การประเมินระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ ผู้วิจัยมีวิธีการเก็บรวบรวมและวิเคราะห์ข้อมูล ซึ่งมีรายละเอียดดังต่อไปนี้

- การเก็บรวบรวมข้อมูล มีรายละเอียดดังนี้

1) ดำเนินการเก็บรวบรวมผลการประเมินระบบก่อนนำระบบไปใช้ ด้วยการประเมินระบบแบบอิงมาตรฐานและแบบประเมินอิวิริสติก จากผู้เชี่ยวชาญด้านการทดสอบ จำนวน 3 ท่าน และผู้เชี่ยวชาญด้านเทคโนโลยีสารสนเทศ จำนวน 2 ท่าน

2) ดำเนินการเก็บรวบรวมผลการประเมินความพึงพอใจของผู้ใช้งานต่อระบบ ด้วยแบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด จากนักเรียนผู้ใช้งานระบบ จำนวน 74 คน

- การวิเคราะห์ข้อมูล มีรายละเอียดดังนี้

1) การวิเคราะห์ค่าความตรงตามเนื้อหา (IOC) มีสูตรการคำนวณความสอดคล้องระหว่างข้อคำถามกับองค์ประกอบในการวัด (Item Objective Congruence Index: IOC) ดังนี้

$$IOC = \frac{\sum R}{N}$$

$\sum R$  = ผลรวมคะแนนความคิดเห็นของผู้ทรงคุณวุฒิ

เกณฑ์การพิจารณา

$IOC \geq$  แสดงว่าข้อคำถามวัดวัตถุประสงค์ข้อนั้นจริง หมายความว่า ข้อนั้นมี ความตรงตามเนื้อหา

$IOC <$  แสดงว่าข้อคำถามไม่ได้วัดวัตถุประสงค์ข้อนั้นจริง หมายความว่า ข้อนั้นไม่มี ความตรงตามเนื้อหา

2) การวิเคราะห์ข้อมูลใช้การวิเคราะห์ด้วยสถิติบรรยาย ซึ่งสถิติที่ใช้ ได้แก่ ความถี่ ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้และความโด่ง

### ระยะที่ 1 การพัฒนาระบบการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

1. การพัฒนาคลังข้อสอบ
  - การรวบรวมข้อสอบ
  - การตรวจสอบและเปรียบเทียบความสอดคล้องกลไกของโมเดลการตอบสนองข้อสอบ
  - การวิเคราะห์คุณภาพของข้อสอบรายข้อตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์
  - การจัดคลังข้อสอบ
2. การพัฒนาระบบการทดสอบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด
  - รูปแบบของระบบทดสอบ
  - การทดสอบ
    - 1) จุดเริ่มต้นการทดสอบ
    - 2) การประมาณค่าความสามารถ (MLE)
    - 3) การคัดเลือกข้อสอบ (MPI)
    - 4) การควบคุมการใช้ข้อสอบซ้ำ (PR-SE)
    - 5) การยุติการทดสอบ ( $\leq 0.3$ )
  - การรายงานผลการทดสอบ
3. การทดลองใช้ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด
  - เตรียมการระบบเพื่อนำไปทดลองใช้
  - กำหนดกลุ่มตัวอย่างที่ใช้ในการทดลองระบบ
  - ดำเนินการทดสอบและประเมินระบบ

### ระยะที่ 2 การประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

- การประเมินระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด
- การประเมินแบบอิงมาตรฐานโดยผู้เชี่ยวชาญ
  - การประเมินแบบฮิวริสติก โดยผู้เชี่ยวชาญ
  - การประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดโดยผู้ใช้งาน

ภาพ 3.6 ขั้นตอนการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

## บทที่ 4

### ผลการวิเคราะห์ข้อมูล

การวิจัย เรื่อง การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า มีวัตถุประสงค์ 2 ประการ ได้แก่ 1) เพื่อพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า และ 2) เพื่อประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยการประเมินแบบอิงมาตรฐาน การประเมินอิวิริสติก และการประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบ โดยมีการดำเนินการวิจัย 2 ระยะ ได้แก่ ระยะที่ 1 การพัฒนาระบบการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด และระยะที่ 2 การประเมินคุณภาพระบบการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ผู้วิจัยได้แบ่งการนำเสนอผลการวิเคราะห์ข้อมูลที่ศึกษา ออกเป็น 2 ระยะ ดังนี้

ระยะที่ 1 ผลพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ตอนที่ 1 ผลการพัฒนาค้างข้อสอบ

ตอนที่ 2 ผลการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด และตรวจสอบคุณภาพของระบบการทดสอบ

ตอนที่ 3 ผลการทดลองใช้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ระยะที่ 2 ผลการประเมินคุณภาพระบบการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ตอนที่ 1 ผลการประเมินแบบอิงมาตรฐาน

ตอนที่ 2 ผลการประเมินระบบแบบอิวิริสติก

ตอนที่ 3 ผลการแบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

#### ระยะที่ 1 ผลการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ผลการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด แบ่งออกเป็น 3 ตอน คือ 1) ผลการพัฒนาค้างข้อสอบ 2) ผลการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด และ 3) ผลการทดลองใช้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีรายละเอียดดังนี้

### ตอนที่ 1 การพัฒนาค้างข้อสอบ

ผลการพัฒนาค้างข้อสอบแบบฝึกทักษะภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 มีรายละเอียดดังนี้

1) ผู้วิจัยได้รวบรวมข้อสอบในแบบฝึกพัฒนาทักษะภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่เผยแพร่ในเว็บไซต์ของสถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ตั้งแต่ปี พ.ศ. 2556 – 2559 จำนวน 640 ข้อ ที่มีรูปแบบของคำตอบเป็นแบบ 4 ตัวเลือก โดยมีการมุ่งวัดความรู้ทางภาษาอังกฤษ ซึ่งประกอบด้วย 4 ด้าน ได้แก่ ด้าน Conversation จำนวน 136 ข้อ, ด้าน Grammar and Vocabulary จำนวน 206 ข้อ, ด้าน Reading Comprehension จำนวน 186 ข้อ และ ด้าน Miscellaneous จำนวน 112 ข้อ ดังตาราง 4.1 ซึ่งข้อสอบทั้งหมดวัดตามตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

ตาราง 4.1 จำนวนข้อสอบแบบฝึกทักษะภาษาอังกฤษ

ด้าน	จำนวนข้อสอบ (ข้อ)
Conversation	136
Grammar	206
Reading	186
Miscellaneous	112
รวม	640

การตรวจสอบคุณภาพข้อสอบโดยนำแบบฝึกทักษะภาษาอังกฤษที่จัดชุดเรียบร้อยแล้วไปทดสอบกับนักเรียนกลุ่มตัวอย่างซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา จำนวน 6 โรงเรียน ได้แก่ โรงเรียนบุญวาทย์วิทยาลัย, โรงเรียนสุนารีวิทยา, โรงเรียนเศรษฐบุตธบ่าเพ็ญ, โรงเรียนแก่ง "วิทยสถาวร", โรงเรียนท่ามะกาวิทยาคม และโรงเรียนสตรีภูเก็ต จำนวน 1,109 คน และตรวจสอบความเรียบร้อยของชุดข้อสอบและกระดาษคำตอบก่อนนำตรวจสอบคุณภาพของข้อสอบต่อไป

2) การตรวจสอบและเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบ ก่อนจะทำการวิเคราะห์คุณภาพของข้อสอบรายข้อตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) เพื่อให้ได้สารสนเทศที่มีความถูกต้อง ผู้วิจัยได้ตรวจสอบและเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบ 3 โมเดล ได้แก่ 1 พารามิเตอร์ (1PL)

2 พารามิเตอร์ (2PL) และ 3 พารามิเตอร์ (3PL) เพื่อยืนยันคุณภาพของข้อสอบ โดยพิจารณาจากค่าสถิติ  $G^2$  (deviance statistic) ค่า AIC (Akaike Information Criterion) ค่า BIC (Bayesian information criterion) เกณฑ์การพิจารณาจากโมเดลใดมีค่าน้อยกว่าจะแสดงถึงโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า และยังเป็นตัวบ่งชี้ถึงโมเดลที่คาดว่าจะดีที่สุดจากการพิจารณาจากความแตกต่างระหว่างโมเดล (เมษา นวลศรี, 2559; ชัยวิจิตต์ เขียวชนะ, 2552) ผลการเปรียบเทียบความสอดคล้องกลมกลืนของข้อสอบ พบว่า ค่าดัชนี deviance ( $G^2$ ) ค่า AIC และค่า BIC ของ โมเดล 3 พารามิเตอร์ มีค่าต่ำที่สุด (3PL) โดย แสดงว่าโมเดล 3 พารามิเตอร์ (3PL) สอดคล้องกลมกลืนกับข้อสอบมากที่สุด ดังตาราง 4.2 แล้วนำไปวิเคราะห์คุณภาพข้อสอบต่อไป โดยดำเนินการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ ด้วยโปรแกรม R แพคเกจ ltm (ณภัทร ชัยมงคล และศิริชัย กาญจนวาสี, 2563)

ตาราง 4.2 ผลการเปรียบเทียบความสอดคล้องกลมกลืนระหว่างโมเดล 1PL, 2PL, และ 3PL

	โมเดล 1 PL	โมเดล 2PL	โมเดล 3PL
n	1109	1109	1109
p	640	1280	1920
$G^2$	863691.4	840653.8	833014.2
AIC	864971.4	843213.8	836854.2
BIC	868183.6	849628.2	846475.7

3) ผลการวิเคราะห์คุณภาพของข้อสอบรายข้อตามทฤษฎีการตอบสนองข้อสอบ การตรวจสอบคุณภาพของข้อสอบที่ได้จากการทดสอบนักเรียนกลุ่มตัวอย่าง มาวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ ได้แก่ ค่าความยากของข้อสอบ (b) ค่าอำนาจจำแนกของข้อสอบ (a) และค่าการเดาของข้อสอบ (c) ดำเนินการวิเคราะห์ข้อสอบด้วยโปรแกรม R แพคเกจ ltm ในการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (item response theory) ในการคัดเลือกข้อสอบเพื่อจัดในคลังมีเกณฑ์ในการคัดเลือกข้อสอบตามค่าพารามิเตอร์ของข้อสอบ ได้แก่ 1) ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง  $\infty$  2) ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -4 ถึง 4 และ 3) ค่าการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.50 ตัวอย่างข้อสอบที่ผ่านการคัดเลือกและคัดออกของด้าน Conversation ด้าน Grammar ด้าน Reading และด้าน Miscellaneous ตาราง 4.3 – 4.6

หลังจากทำการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ พบว่า ข้อสอบจำนวน 640 ข้อมีค่าสถิติพื้นฐานดังนี้ คุณภาพข้อสอบมีค่าอำนาจจำแนกของข้อสอบ (a) มีค่าเฉลี่ย 2.36

ค่าส่วนเบี่ยงเบนมาตรฐาน 5.84 ค่าต่ำสุด -33.56 ค่าสูงสุด 33.71 ค่าความยากของข้อสอบ (b) มีค่าเฉลี่ย -1.86 ค่าส่วนเบี่ยงเบนมาตรฐาน 8.66 ค่าต่ำสุด -84.02 ค่าสูงสุด 5.13 และค่าการเดาของข้อสอบ (c) มีค่าเฉลี่ย -0.17 ค่าส่วนเบี่ยงเบนมาตรฐาน 0.13 ค่าต่ำสุด 0.00 ค่าสูงสุด 0.57 ดังตาราง 4.7

และข้อสอบที่ผ่านเกณฑ์ในการคัดเลือกข้อสอบ จำนวน 476 ข้อ มีรายละเอียด ดังนี้ ข้อสอบด้าน Conversation ที่ผ่านเกณฑ์การคัดเลือกข้อสอบ จำนวน 107 ข้อ ข้อสอบด้าน Grammar ที่ผ่านเกณฑ์การคัดเลือกข้อสอบ จำนวน 133 ข้อ ข้อสอบด้าน Reading ที่ผ่านเกณฑ์การคัดเลือกข้อสอบ จำนวน 149 ข้อ และ ข้อสอบด้าน Miscellaneous ที่ผ่านเกณฑ์การคัดเลือกข้อสอบ จำนวน 87 ข้อ คุณภาพข้อสอบมีค่าอำนาจจำแนกของข้อสอบ (a) มีค่าเฉลี่ย 3.72 ค่าส่วนเบี่ยงเบนมาตรฐาน 4.22 ค่าต่ำสุด 0.51 ค่าสูงสุด 33.71 ค่าความยากของข้อสอบ (b) มีค่าเฉลี่ย 0.07 ค่าส่วนเบี่ยงเบนมาตรฐาน 1.67 ค่าต่ำสุด -3.86 ค่าสูงสุด 4.00 และค่าการเดาของข้อสอบ (c) มีค่าเฉลี่ย 0.18 ค่าส่วนเบี่ยงเบนมาตรฐาน 0.12 ค่าต่ำสุด 0.00 ค่าสูงสุด 0.49 ดังตาราง 4.8 – 4.9

ตาราง 4.3 ตัวอย่างข้อสอบที่ผ่านการคัดเลือกและคัดออกของด้าน Conversation

ตัวอย่างข้อสอบที่ผ่านการคัดเลือก ด้าน Conversation	ตัวอย่างข้อสอบที่ผ่านการคัดออก ด้าน Conversation
<p><b>Directions:</b> Read each situation and the dialogue carefully and choose the appropriate expression to complete the dialogue.</p> <p><b>Situation:</b> Kitty is going abroad by plane.</p> <p>George says “_____”</p> <p>A. Have a good flight. B. Enjoy your cruise. C. Have a safe drive. D. Have a nice ride.</p>	<p><b>Directions:</b> Read each situation and the dialogue carefully and choose the appropriate expression to complete the dialogue.</p> <p><b>Situation:</b> At the café.</p> <p>Chris: Would you like some coffee? Jemma: _____ I don't like it.</p> <p>A. No, thanks. B. Yes, please. C. Not just now. D. I would love to.</p>
พารามิเตอร์อำนาจจำแนกของข้อสอบ (a) = 6.213	พารามิเตอร์อำนาจจำแนกของข้อสอบ (a) = 8.497
พารามิเตอร์ความยาก (b) = 2.904	พารามิเตอร์ความยาก (b) = 4.602
พารามิเตอร์การเดา (c) = 0.362	พารามิเตอร์การเดา (c) = 0.380

ตาราง 4.4 ตัวอย่างข้อสอบที่ผ่านการคัดเลือกและคัดออกของด้าน Grammar

ตัวอย่างข้อสอบที่ผ่านการคัดเลือก ด้าน Grammar	ตัวอย่างข้อสอบที่ผ่านการคัดออก ด้าน Grammar
<p><b>Directions:</b> Complete the following passage by choosing the best answer.</p> <p>Eiffel Tower is one of the world's most popular destinations. It is named after its architect, Gustave Eiffel, and _____ in 1889 for the Universal Exposition. The tower is 1,062 feet high and offers fantastic views of Paris. Very often, visitors have to wait in line to take the elevator to the observation deck. At night, the tower itself becomes a very illuminated sculpture.</p> <p>A .is B .was C .are D .were</p>	<p><b>Directions:</b> Complete the following passage by choosing the best answer.</p> <p>Google is a search site that helps internet users find information quickly. _____ to other sites, Google is faster, and it becomes the most popular search engine .Every second, thousands of people visit the site. That is millions of search results a day-a number that increases daily.</p> <p>A. Compare B. Compared C. Comparing D. To compare</p>
พารามิเตอร์อำนาจจำแนกของข้อสอบ (a) = 4.804	พารามิเตอร์อำนาจจำแนกของข้อสอบ (a) = - 0.411
พารามิเตอร์ความยาก (b) = 1.395	พารามิเตอร์ความยาก (b) = -1.634
พารามิเตอร์การเดา (c) = 0.279	พารามิเตอร์การเดา (c) = 0.000



ตาราง 4.5 ตัวอย่างข้อสอบที่ผ่านการคัดเลือกและคัดออกของด้าน Reading

ตัวอย่างข้อสอบที่ผ่านการคัดเลือก ด้าน Reading	ตัวอย่างข้อสอบที่ผ่านการคัดออก ด้าน Reading														
<p><b>Directions:</b> Read each passage carefully and choose the correct answer the questions.</p> <p style="text-align: center;"><b>GM Foods</b></p> <p>GM foods are especially popular in the US where up to seventy percent of food products in supermarkets could contain a crop that has been genetically modified. However, in many other countries, GM foods can be controversial. In fact, many people protest and concerns are so great that some governments have taken action.</p> <p style="text-align: center;"><b>Which part of the newspaper can you find this news?</b></p> <p>A. Comics B. Business C. Editorials D. World news</p>	<p><b>Directions:</b> Read each passage carefully and choose the correct answer the questions.</p> <div data-bbox="882 562 1342 797" data-label="Figure"> <table border="1"> <caption>CHOCOLATE MILK THAILAND 6 YEARS OF PRODUCTION</caption> <thead> <tr> <th>Year</th> <th>Production (Millions of Cartons)</th> </tr> </thead> <tbody> <tr> <td>Year 1</td> <td>~850</td> </tr> <tr> <td>Year 2</td> <td>~860</td> </tr> <tr> <td>Year 3</td> <td>~870</td> </tr> <tr> <td>Year 4</td> <td>~880</td> </tr> <tr> <td>Year 5</td> <td>~890</td> </tr> <tr> <td>Year 6</td> <td>~900</td> </tr> </tbody> </table> </div> <p style="text-align: center;"><b>Which year was chocolate milk produced the most?</b></p> <p>A. Year 1 B. Year 3 C. Year 4 D. Year 6</p>	Year	Production (Millions of Cartons)	Year 1	~850	Year 2	~860	Year 3	~870	Year 4	~880	Year 5	~890	Year 6	~900
Year	Production (Millions of Cartons)														
Year 1	~850														
Year 2	~860														
Year 3	~870														
Year 4	~880														
Year 5	~890														
Year 6	~900														
พารามิเตอร์อำนาจจำแนกของข้อสอบ (a) = 2.185	พารามิเตอร์อำนาจจำแนกของข้อสอบ (a) = 11.976														
พารามิเตอร์ความยาก (b) = - 0.306	พารามิเตอร์ความยาก (b) = 4.101														
พารามิเตอร์การเดา (c) = 0.201	พารามิเตอร์การเดา (c) = 0.262														

ตาราง 4.6 ตัวอย่างข้อสอบที่ผ่านการคัดเลือกและคัดออกของด้าน Miscellaneous

ตัวอย่างข้อสอบที่ผ่านการคัดเลือก ด้าน Miscellaneous	ตัวอย่างข้อสอบที่ผ่านการคัดออก ด้าน Miscellaneous
<p><b>Directions:</b> Choose the correct answer which has the same meaning as the given sentence.</p> <p><b>Both Mary and John go out at night.</b></p> <p>A. John doesn't go out at night. B. John and Mary don't go out at night. C. Mary goes out at night, so does John. D. Mary goes out at night but John doesn't.</p>	<p><b>Directions:</b> Choose the answer which is grammatically incorrect.</p> <p>Nike introduced the very first self-tying shoe which you just (A) <u>put on</u>, and when your heel (B) <u>is</u> in, the shoe (C) <u>automatic</u> adjusts to a (D) <u>comfortable</u> fit.</p>
พารามิเตอร์อำนาจจำแนกของข้อสอบ (a) = 6.359	พารามิเตอร์อำนาจจำแนกของข้อสอบ (a) = 27.496
พารามิเตอร์ความยาก (b) = 2.715	พารามิเตอร์ความยาก (b) = -5.584
พารามิเตอร์การเดา (c) = 0.258	พารามิเตอร์การเดา (c) = 0.222

ตาราง 4.7 ค่าสถิติพื้นฐานของพารามิเตอร์ของข้อสอบ

ค่าพารามิเตอร์ของข้อสอบ	M	SD	Min	Max
ค่าอำนาจจำแนกของข้อสอบ (a)	2.36	5.84	-33.56	33.71
ค่าความยากของข้อสอบ (b)	-1.86	8.66	-84.02	5.13
ค่าการเดาของข้อสอบ (c)	0.17	0.13	0.00	0.57

ตาราง 4.8 จำนวนข้อสอบที่ผ่านการคัดเลือก

ด้าน	จำนวนข้อสอบที่ผ่านการคัดเลือก (ร้อยละ)	จำนวนข้อสอบที่ไม่ผ่านการคัดเลือก (ร้อยละ)	จำนวนข้อสอบทั้งหมด (ร้อยละ)
Conversation	107 (78.68)	29 (21.32)	136 (100.00)
Grammar	133 (64.56)	73 (35.44)	206 (100.00)
Reading	149 (80.11)	37 (19.89)	186 (100.00)
Miscellaneous	87 (77.68)	25 (22.32)	112 (100.00)
รวม	476 (74.37)	164 (25.63)	640 (100.00)

ตาราง 4.9 ค่าสถิติพื้นฐานของพารามิเตอร์ของข้อสอบที่ผ่านการคัดเลือก

ค่าพารามิเตอร์ของข้อสอบ	M	SD	Min	Max
ค่าอำนาจจำแนกของข้อสอบ (a)	3.72	4.22	0.51	33.17
ค่าความยากของข้อสอบ (b)	0.07	1.67	-3.86	4.00
ค่าการเดาของข้อสอบ (c)	0.18	0.12	0.00	0.49

#### 4) การจัดคลังข้อสอบ

สำหรับการทดสอบในขั้นแรก (stage) ดำเนินการจัดชุดข้อสอบไว้ล่วงหน้า จำนวน 1 ชุด โดยเป็นข้อสอบที่มีความยาก (b) ปานกลาง หรือมีค่าอยู่ระหว่าง -0.5 ถึง 0.5 ซึ่งข้อสอบแต่ละชุดประกอบด้วยเนื้อหาครบทั้ง 4 ด้าน เพื่อให้ผู้เข้ารับการทดสอบได้รับชุดข้อสอบชุดแรกที่มีเนื้อหาครบถ้วน ซึ่งระบบการสร้างชุดข้อสอบ ประกอบด้วยฐานข้อมูล ซึ่งมีข้อสอบในคลัง จำนวน 476 ข้อ

ในขั้นที่สอง ดำเนินการจัดชุดข้อสอบไว้ล่วงหน้า จำนวน 3 ชุด ซึ่งข้อสอบแต่ละชุดแบ่งเป็น ชุดง่าย ชุดปานกลาง และชุดยาก ซึ่งข้อสอบแต่ละชุดประกอบด้วยเนื้อหาครบทั้ง 4 ด้าน รายละเอียดดังภาพ 4.1

จัดการข้อมูลต่างๆ   ออกรายงานต่างๆ   ออกจากระบบ				
ข้อมูล - ข้อสอบ				
ข้อสอบ				
ชุดที่ 1				
1	2	3	4	5
ชุดที่ 2				
Low Level				
1	2	3	4	5
Medium Level				
1	2	3	4	5
High Level				
1	2	3	4	5

ภาพ 4.1 ชุดข้อสอบสำหรับการทดสอบขั้นแรกและขั้นที่สอง

ชุดที่ 3									
Low Level									
1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
Medium Level									
1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170
171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190
191	192	193	194	195	196	197	198	199	200
201	202	203	204	205	206	207	208	209	210
211	212	213	214	215	216	217	218	219	220
221	222	223	224	225	226	227	228	229	230
231	232	233	234	235	236	237	238	239	240
241	242	243	244	245	246	247	248	249	250
251	252	253	254	255	256	257	258	259	260
High Level									
1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110

ภาพ 4.2 ชุดข้อสอบสำหรับการสอบขั้นที่ 3

## ข้อที่ 1

ประเภทข้อสอบ : C=Conversation

ระดับข้อสอบ : M=Medium

ค่า a : 1.413

ค่า b : 0.054

ค่า c : 0.268

1. Read each situation and the dialogue carefully and choose the appropriate expression to complete the dialogue.

**Situation:** Your brother has won a scholarship to study in Canada.  
You congratulate him by saying, “\_\_\_\_\_”

- A. How pretty you look!
- B. Very well. How did it go?
- C. That's great! How lucky you are!
- D. Amazing! I never expected so much.

Answer : C

แก้ไขข้อสอบ

ภาพ 4.3 ตัวอย่างข้อสอบที่บันทึกลงในฐานข้อมูล

### ตอนที่ 2 การพัฒนาระบบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด

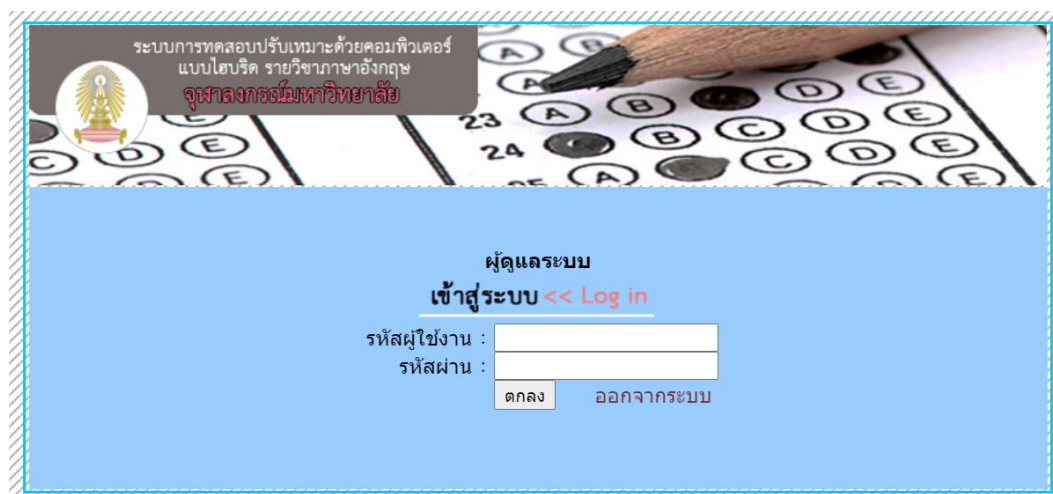
การพัฒนาระบบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด แบ่งออกเป็น 3 ส่วน ได้แก่ 1) รูปแบบของระบบทดสอบ 2) การทดสอบ และ 3) การรายงานผลการทดสอบ

#### 2.1 รูปแบบของระบบทดสอบ มีรายละเอียด ดังนี้

รูปแบบของระบบทดสอบการปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดเป็นรูปแบบออนไลน์ มีการกำหนดสิทธิ์ในการเข้าถึงในการใช้งานของระบบ และจัดทำคู่มือการใช้งานเป็นแนวทางในการนำโปรแกรมการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด ไปใช้งานทำให้ผู้ใช้งานทราบถึงลำดับขั้นตอนการใช้โปรแกรมอย่างละเอียดทุกขั้นตอนในการใช้งาน มีรายละเอียดดังนี้

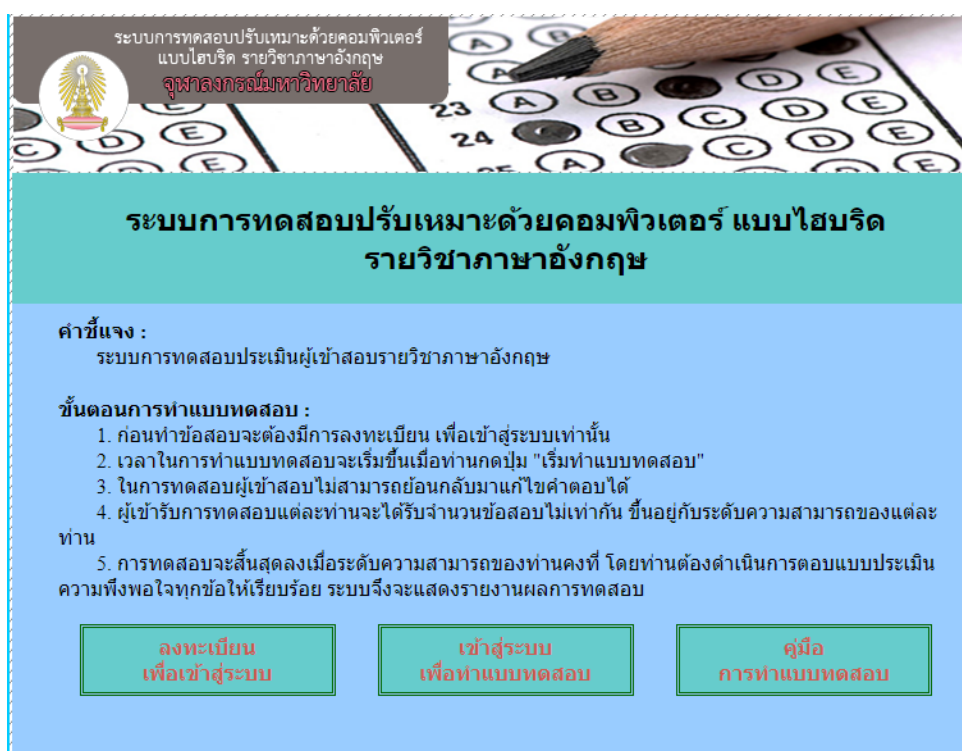
- 1) การกำหนดสิทธิ์ในการเข้าถึงในการใช้งานของระบบ โดยแบ่งออกเป็น ผู้จัดการระบบ (admin) และผู้ใช้งาน (user) การออกแบบให้ผู้ใช้งานมีการเข้าถึงข้อมูลที่สะดวก รวดเร็วและทันสมัย มีรายละเอียดของการทดสอบ

- ผู้จัดการระบบ (admin) ผู้ดูแลระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีหน้าที่จัดการข้อมูลต่าง ๆ ของระบบ ได้แก่จัดการข้อมูลโรงเรียน จัดการข้อมูลนักเรียน ผู้เข้าสอบ จัดการข้อมูลข้อสอบ และออกรายงานต่าง ๆ ได้แก่ ผลการตอบแบบสอบถาม และผลการสอบ โดยการเข้าสู่ระบบจะต้องใช้รหัสผู้ใช้งานและรหัสผ่านที่ได้รับจากผู้พัฒนาโปรแกรม ดังภาพ 4.4



ภาพ 4.4 การเข้าระบบของผู้จัดการระบบ

- ผู้ใช้งาน (user) ผู้ที่ใช้งานเพื่อเข้าระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เพื่อทดสอบทักษะภาษาอังกฤษ ต้องมีการลงทะเบียนเพื่อสมัครสมาชิกเข้าสู่ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยต้องดำเนินการลงทะเบียนก่อนเข้ารับ การทดสอบ โดยเปิดเว็บไซต์ <http://www.hybrid-cat.com> จะปรากฏข้อมูลดังภาพ 4.5 หลังจากนั้นเลือกเมนู ลงทะเบียนเพื่อเข้าสู่ระบบเพื่อลงทะเบียน โดยผู้เข้ารับการทดสอบต้องกรอกอีเมล (e-mail) รหัสผ่าน ยืนยันรหัสผ่าน ชื่อ นามสกุล เบอร์โทรศัพท์ และโรงเรียนของผู้ทดสอบ โดยรายละเอียดแสดงดังภาพ 4.6 เมื่อดำเนินการกรอกข้อมูลเรียบร้อยแล้ว กดปุ่ม บันทึกลงทะเบียน และการเข้าสู่ระบบการทดสอบ ผู้ใช้งานกรอกรหัสผู้ใช้งาน และรหัสผ่านให้ถูกต้อง เพื่อเข้าสู่ระบบ ดังภาพ 4.7



ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

## ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ แบบไฮบริด รายวิชาภาษาอังกฤษ

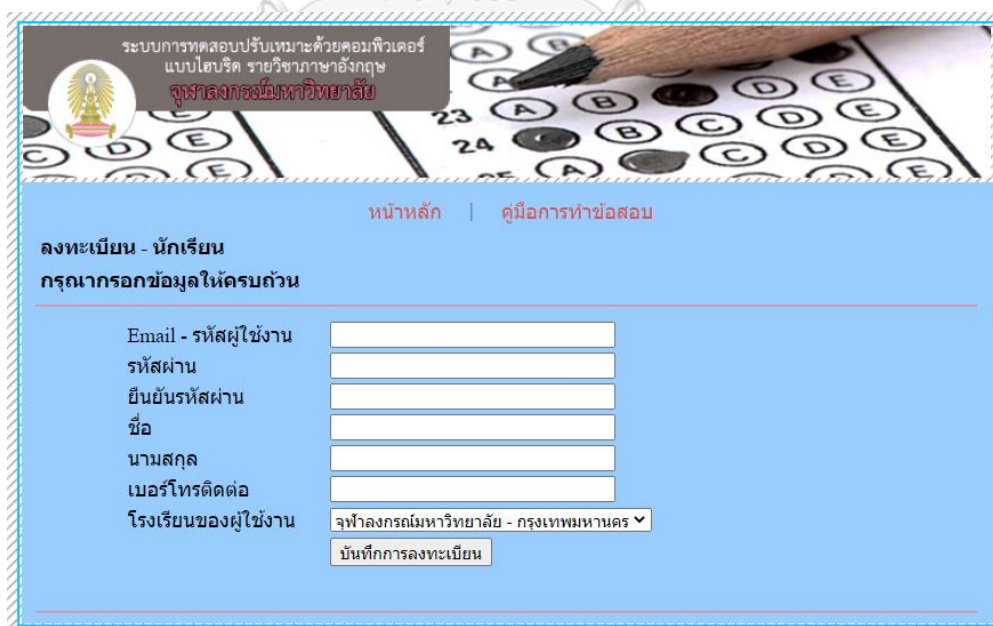
**คำชี้แจง :**  
ระบบการทดสอบประเมินผู้เข้าสอบรายวิชาภาษาอังกฤษ

**ขั้นตอนการทำแบบทดสอบ :**

1. ก่อนทำข้อสอบจะต้องมีการลงทะเบียน เพื่อเข้าสู่ระบบเท่านั้น
2. เวลาในการทำแบบทดสอบจะเริ่มขึ้นเมื่อท่านกดปุ่ม "เริ่มทำแบบทดสอบ"
3. ในการทดสอบผู้เข้าสอบไม่สามารถย้อนกลับมาแก้ไขคำตอบได้
4. ผู้เข้ารับการทดสอบแต่ละท่านจะได้รับจำนวนข้อสอบไม่เท่ากัน ขึ้นอยู่กับระดับความสามารถของแต่ละท่าน
5. การทดสอบจะสิ้นสุดลงเมื่อระดับความสามารถของท่านคงที่ โดยท่านต้องดำเนินการตอบแบบประเมินความพึงพอใจทุกข้อให้เรียบร้อย ระบบจึงจะแสดงรายงานผลการทดสอบ

**ลงทะเบียนเพื่อเข้าสู่ระบบ**      **เข้าสู่ระบบเพื่อทำแบบทดสอบ**      **คู่มือการทำแบบทดสอบ**

ภาพ 4.5 หน้าหลักของระบบการทดสอบ



ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

หน้าหลัก | คู่มือการทำข้อสอบ

### ลงทะเบียน - นักเรียน

กรณารอกข้อมูลให้ครบถ้วน

Email - รหัสผู้ใช้งาน

รหัสผ่าน

ยืนยันรหัสผ่าน

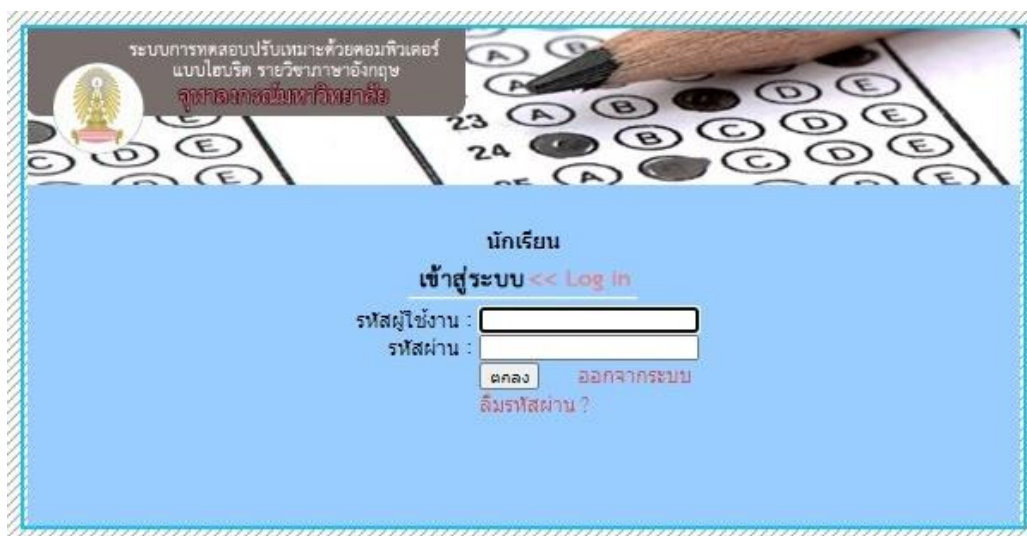
ชื่อ

นามสกุล

เบอร์โทรศัพท์

โรงเรียนของผู้ใช้งาน

ภาพ 4.6 การลงทะเบียน



ภาพ 4.7 การเข้าระบบของผู้ใช้งาน

2) คู่มือการใช้ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

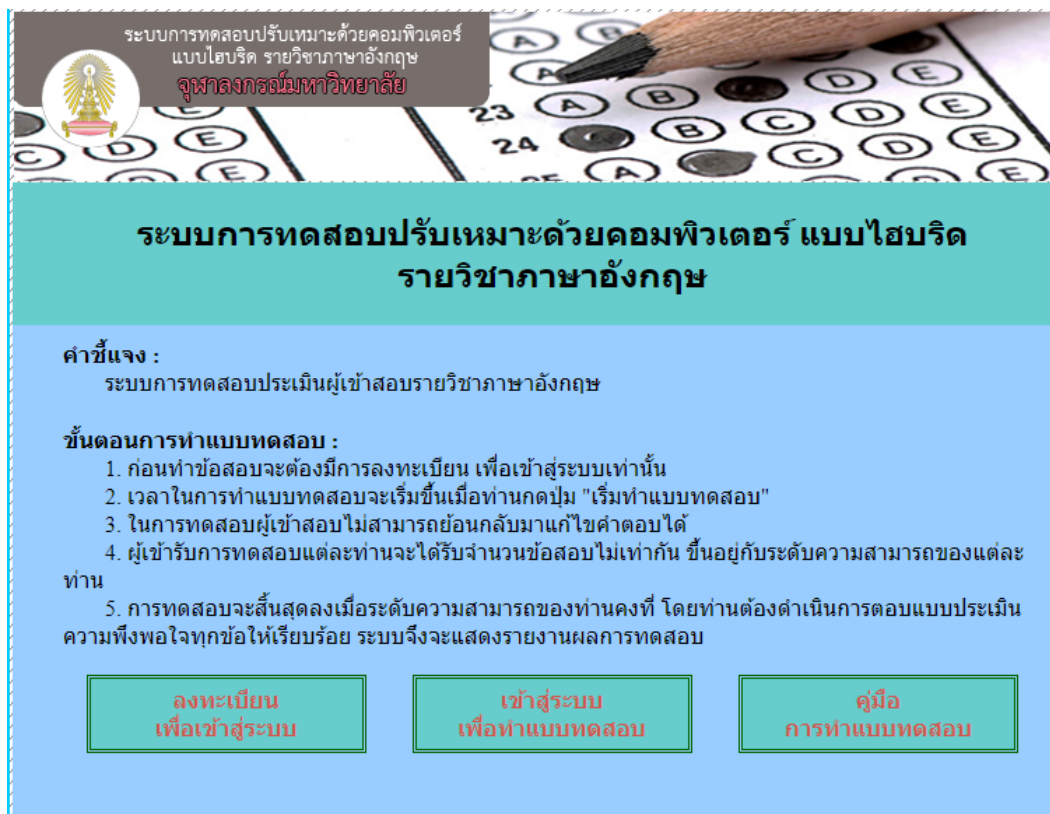
สำหรับคู่มือการใช้ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด แบ่งออกเป็น 2 ส่วน ได้แก่ คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป (user) และคู่มือการใช้งานสำหรับผู้จัดการระบบ (admin) โดยแต่ละส่วนมีรายละเอียดดังนี้

- คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป (user) เพื่อเข้าสู่ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีรายละเอียดขั้นตอนการใช้งานระบบ ได้แก่ การลงทะเบียนเข้าสู่ระบบการทดสอบ และการรายงานผลการทดสอบ



### คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป (user)

ผู้ใช้งานสามารถเข้าสู่ระบบการทดสอบโดยพิมพ์ Url: <http://www.hybrid-cat.com> ซึ่งจะ  
เจอหน้าจอหลัก ดังภาพ 4.8



ภาพ 4.8 หน้าจอหลัก

หน้าจอหลักของระบบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด สำหรับผู้ใช้งาน  
ทั่วไป ประกอบด้วย

1. คำชี้แจง ระบบการทดสอบประเมินผู้เข้าสอบรายวิชาภาษาอังกฤษ
2. ขั้นตอนการทำแบบทดสอบ มีรายละเอียดดังนี้
  - 1) ก่อนทำข้อสอบจะต้องมีการลงทะเบียน เพื่อเข้าสู่ระบบเท่านั้น
  - 2) เวลาในการทำแบบทดสอบจะเริ่มขึ้นเมื่อท่านกดปุ่ม "เริ่มทำแบบทดสอบ"
  - 3) ในการทดสอบผู้เข้าสอบไม่สามารถย้อนกลับมาแก้ไขคำตอบได้
  - 4) ผู้เข้ารับการทดสอบแต่ละท่านจะได้รับจำนวนข้อสอบไม่เท่ากัน ขึ้นอยู่กับระดับ  
ความสามารถของแต่ละท่าน
  - 5) การทดสอบจะสิ้นสุดลงเมื่อระดับความสามารถของท่านคงที่ โดยท่านต้องดำเนินการ  
ตอบแบบประเมินความพึงพอใจทุกข้อให้เรียบร้อย ระบบจึงจะแสดงรายงานผลการทดสอบ

3. ลงทะเบียนเพื่อเข้าสู่ระบบ
4. เข้าสู่ระบบเพื่อทำแบบทดสอบ

#### การลงทะเบียน

ในหน้าจอหลัก ผู้ใช้งานคลิก ลงทะเบียนเพื่อเข้าสู่ระบบ เพื่อลงทะเบียน ดังภาพ 4.9 โดยเมื่อกดปุ่มลงทะเบียนเพื่อเข้าสู่ระบบ จะให้ผู้ใช้งานกรอกข้อมูลส่วนบุคคล ได้แก่ E-mail – รหัสผู้ใช้งาน รหัสผ่าน ยืนยันรหัสผ่าน ชื่อ นามสกุล เบอร์โทรศัพท์ และโรงเรียนของผู้ใช้งาน จะปรากฏหน้าจอ ดังภาพ 4.10

ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

**ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ แบบไฮบริด  
รายวิชาภาษาอังกฤษ**

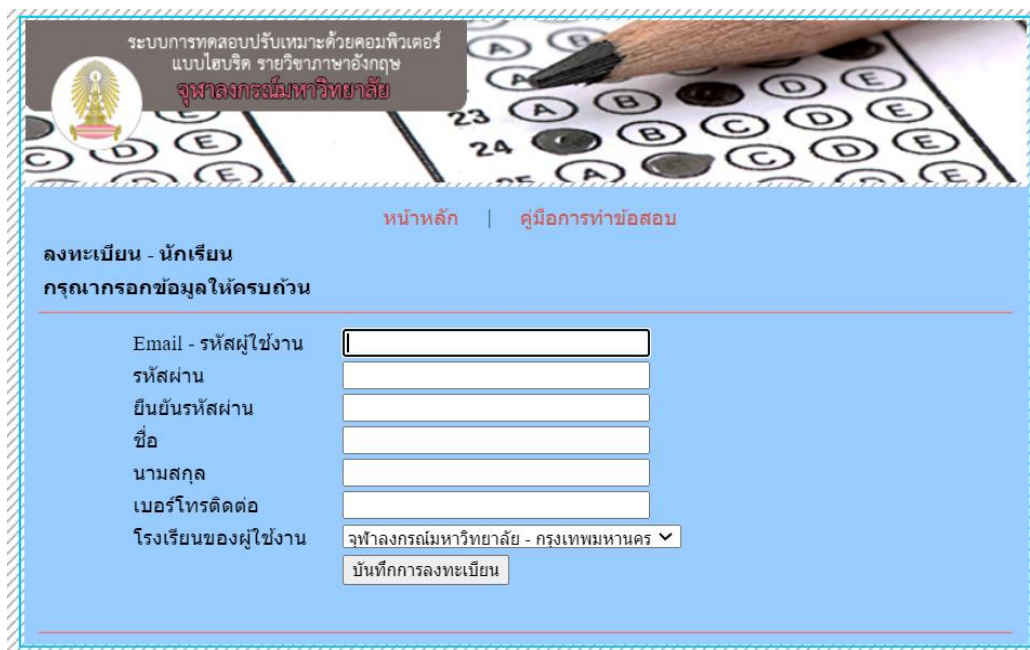
**คำชี้แจง :**  
ระบบการทดสอบประเมินผู้เข้าสอบรายวิชาภาษาอังกฤษ

**ขั้นตอนการทำแบบทดสอบ :**

1. ก่อนทำข้อสอบจะต้องมีการลงทะเบียน เพื่อเข้าสู่ระบบเท่านั้น
2. เวลาในการทำแบบทดสอบจะเริ่มขึ้นเมื่อท่านกดปุ่ม "เริ่มทำแบบทดสอบ"
3. ในการทดสอบผู้เข้าสอบไม่สามารถย้อนกลับมาแก้ไขคำตอบได้
4. ผู้เข้ารับการทดสอบแต่ละท่านจะได้รับจำนวนข้อสอบไม่เท่ากัน ขึ้นอยู่กับระดับความสามารถของแต่ละท่าน
5. การทดสอบจะสิ้นสุดลงเมื่อระดับความสามารถของท่านคงที่ โดยท่านต้องดำเนินการตอบแบบประเมินความพึงพอใจทุกข้อให้เรียบร้อย ระบบจึงจะแสดงรายงานผลการทดสอบ

**ลงทะเบียนเพื่อเข้าสู่ระบบ**      **เข้าสู่ระบบเพื่อทำแบบทดสอบ**      **คู่มือการทำแบบทดสอบ**

ภาพ 4.9 หน้าจอหลักระบบการทดสอบ



ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

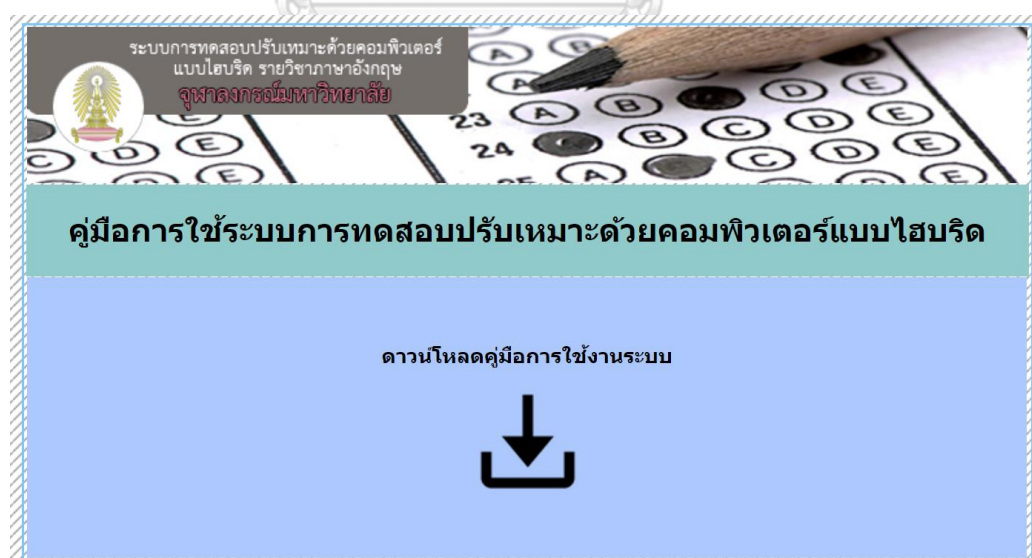
หน้าหลัก | คู่มือการทำข้อสอบ

ลงทะเบียน - นักเรียน  
กรุณากรอกข้อมูลให้ครบถ้วน

Email - รหัสผู้ใช้งาน	<input type="text"/>
รหัสผ่าน	<input type="password"/>
ยืนยันรหัสผ่าน	<input type="password"/>
ชื่อ	<input type="text"/>
นามสกุล	<input type="text"/>
เบอร์โทรศัพท์	<input type="text"/>
โรงเรียนของผู้ใช้งาน	จุฬาลงกรณ์มหาวิทยาลัย - กรุงเทพมหานคร ▼
	บันทึกการลงทะเบียน

ภาพ 4.10 การลงทะเบียนผู้เข้าสอบ

ในหน้าจอหลัก หากผู้ใช้งานเลือก “คู่มือการทำข้อสอบ” และคลิกที่ดาวน์โหลดคู่มือการใช้งานระบบ จะมีรายละเอียดคู่มือการใช้งานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด หน้าจอ ดังภาพ 4.11



ภาพ 4.11 หน้าจอคู่มือการใช้งานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

เมื่อผู้ใช้งานลงทะเบียนเป็นสมาชิกสมบูรณ์แล้ว จะดำเนินการเข้าสู่ระบบการทดสอบในหน้าแรก โดยผู้ใช้งานดำเนินการกรอกรหัสผู้ใช้งาน และรหัสผ่านให้ถูกต้อง หลังจากนั้นคลิกปุ่ม ตกลง เพื่อเข้าสู่ระบบ จะปรากฏหน้าจอ ดังภาพ 4.12

ภาพ 4.12 การเข้าสู่ระบบการทดสอบ

จากภาพ 5 เมื่อผู้ใช้งานเข้าสู่ระบบแล้ว จะปรากฏหน้าจอที่แสดงรายละเอียดเกี่ยวกับผู้เข้ารับการทดสอบ ประกอบด้วยเมนู แก้ไขข้อมูลส่วนตัว ประวัติการทำแบบทดสอบ คู่มือการทำแบบทดสอบ และคู่มือการทำแบบทดสอบ ดังภาพ 4.13

แก้ไขข้อมูลส่วนตัว | ประวัติการทำแบบทดสอบ | คู่มือการทำแบบทดสอบ | ออกจากระบบ

**คำชี้แจง :**  
ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ แบบไฮบริด รายวิชาภาษาอังกฤษ

**ขั้นตอนการทำแบบทดสอบ :**

1. ก่อนทำข้อสอบจะต้องมีการลงทะเบียน เพื่อเข้าสู่ระบบเท่านั้น
2. เวลาในการทำแบบทดสอบจะเริ่มขึ้นเมื่อท่านกดปุ่ม "เริ่มทำแบบทดสอบ"
3. ในการทดสอบผู้เข้าสอบไม่สามารถย้อนกลับมาแก้ไขคำตอบได้
4. ผู้เข้ารับการทดสอบแต่ละท่านจะได้รับจำนวนข้อสอบไม่เท่ากัน ขึ้นอยู่กับระดับความสามารถของแต่ละท่าน
5. การทดสอบจะสิ้นสุดลงเมื่อระดับความสามารถของท่านคงที่ โดยท่านต้องดำเนินการตอบแบบประเมินความพึงพอใจทุกข้อให้เรียบร้อย ระบบจึงจะแสดงรายงานผลการทดสอบ

**เริ่มทำแบบทดสอบ**

ภาพ 4.13 หน้าจอแสดงรายละเอียดเกี่ยวกับผู้เข้ารับการทดสอบ

จากภาพ 4.13 เมื่อใช้งานต้องการแก้ไขข้อมูลส่วนตัว สามารถกดที่เมนู แก้ไขข้อมูลส่วนตัว จะปรากฏหน้าจอ ดังภาพ 4.14 หลังจากนั้นผู้ใช้งานสามารถแก้ไขข้อมูล เมื่อดำเนินการแก้ไขเรียบร้อยแล้ว คลิก “บันทึกการแก้ไขข้อมูล”

ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

แก้ไขข้อมูลส่วนตัว | ประวัติการทำข้อสอบ | คู่มือการทำข้อสอบ | ออกจากระบบ

**แก้ไขข้อมูลการลงทะเบียน - นักเรียน**  
กรณเภากรอกข้อมูลให้ครบถ้วน

ลำดับที่	107
Email - รหัสผู้ใช้งาน	rungsimaporn.n@gmail.com
รหัสผ่าน	<input type="password"/>
ยืนยันรหัสผ่าน	<input type="password"/>
ชื่อ	Rungsimaporn
นามสกุล	Noonoi
เบอร์โทรศัพท์	-
โรงเรียนของผู้ใช้งาน	จุฬาลงกรณ์มหาวิทยาลัย - กรุงเทพมหานคร ▼

ภาพ 4.14 หน้าจอแก้ไขข้อมูลส่วนตัว

จากภาพ 4.13 เมื่อผู้ใช้งานต้องการทราบข้อมูลเกี่ยวกับประวัติการทำข้อสอบ สามารถกดที่เมนู “ประวัติการทำแบบทดสอบ” จะแสดงผลการทดสอบของผู้ใช้งาน จะปรากฏหน้าจอ ดังภาพ 4.15

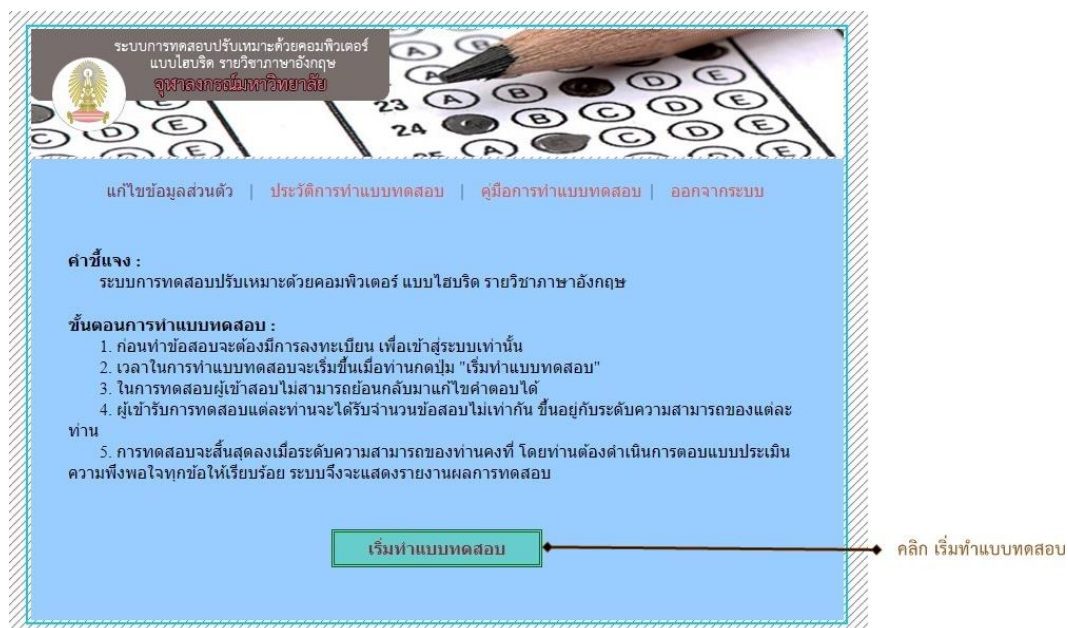


ครั้งที่ 2 วันที่ 04 / 06 / 2020						
ด้าน	ผลการทำแบบทดสอบ (จำนวนข้อ)			ร้อยละของ ระดับความสามารถ	T-Score	ผล
	ทั้งหมด	ถูก	ผิด			
Conversation	3	3	0	100	57.8	ดีมาก
Reading	5	5	0	100	57.8	ดีมาก
Grammar	2	1	1	50	36.9	พอใช้
Miscellaneous	4	3	1	75	47.4	ดี
รวม	14	12	2	85.71		
ระดับความสามารถจากการทดสอบครั้งนี้ ดีมาก						
ท่านได้คะแนนสอบภาษาอังกฤษ ด้าน Conversation มากที่สุด คิดเป็นร้อยละ 100 ของข้อที่ทำได้ ด้าน Reading มากที่สุด คิดเป็นร้อยละ 100 ของข้อที่ทำได้ ด้าน Grammar น้อยที่สุด คิดเป็นร้อยละ 50 ของข้อที่ทำได้						
ครั้งที่ 1 วันที่ 04 / 06 / 2020						
ด้าน	ผลการทำแบบทดสอบ (จำนวนข้อ)			ร้อยละของ ระดับความสามารถ	T-Score	ผล
	ทั้งหมด	ถูก	ผิด			
Conversation	4	4	0	100	57.4	ดีมาก
Reading	5	4	1	80	48.9	ดีมาก
Grammar	2	1	1	50	36.2	พอใช้
Miscellaneous	3	3	0	100	57.4	ดีมาก
รวม	14	12	2	85.71		
ระดับความสามารถจากการทดสอบครั้งนี้ ดีมาก						
ท่านได้คะแนนสอบภาษาอังกฤษ ด้าน Conversation มากที่สุด คิดเป็นร้อยละ 100 ของข้อที่ทำได้ ด้าน Miscellaneous มากที่สุด คิดเป็นร้อยละ 100 ของข้อที่ทำได้ ด้าน Grammar น้อยที่สุด คิดเป็นร้อยละ 50 ของข้อที่ทำได้						

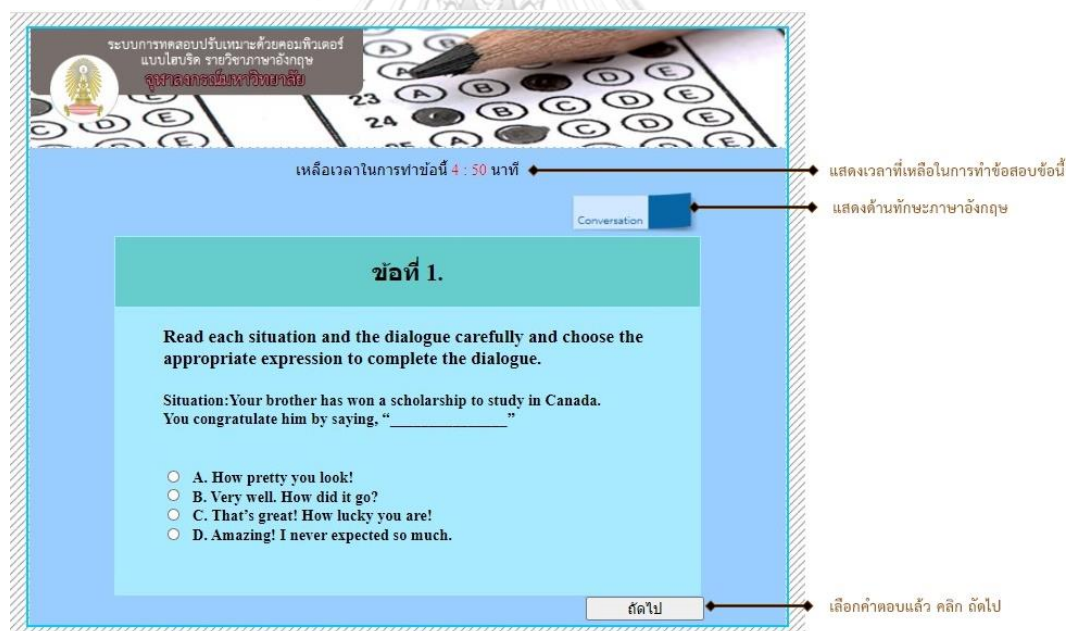
ภาพ 4.15 หน้าจอแสดงประวัติการทำข้อสอบ

การทดสอบ

จากภาพ 4.13 เมื่อผู้ใช้งานดำเนินการเข้าระบบเรียบร้อยแล้ว จะพบหน้าต่างชี้แจง และ ขั้นตอน การทำแบบทดสอบ เมื่อคลิก “เริ่มทำแบบทดสอบ” ดังภาพ 4.16 จะเข้าสู่ห้องสอบ จะปรากฏหน้าจอ ดังภาพ 4.17 โดยแสดงข้อสอบทีละข้อ จะมีการแสดงเวลาที่เหลือในการทำข้อสอบ และแสดงเนื้อหาทักษะภาษาอังกฤษของข้อสอบ เมื่อผู้เข้ารับการทดสอบเลือกคำตอบแล้ว คลิกที่ปุ่ม “ถัดไป” เพื่อไปทำในข้อถัดไปเรื่อย ๆ จนกระทั่งถึงเกณฑ์การยุติการทดสอบ



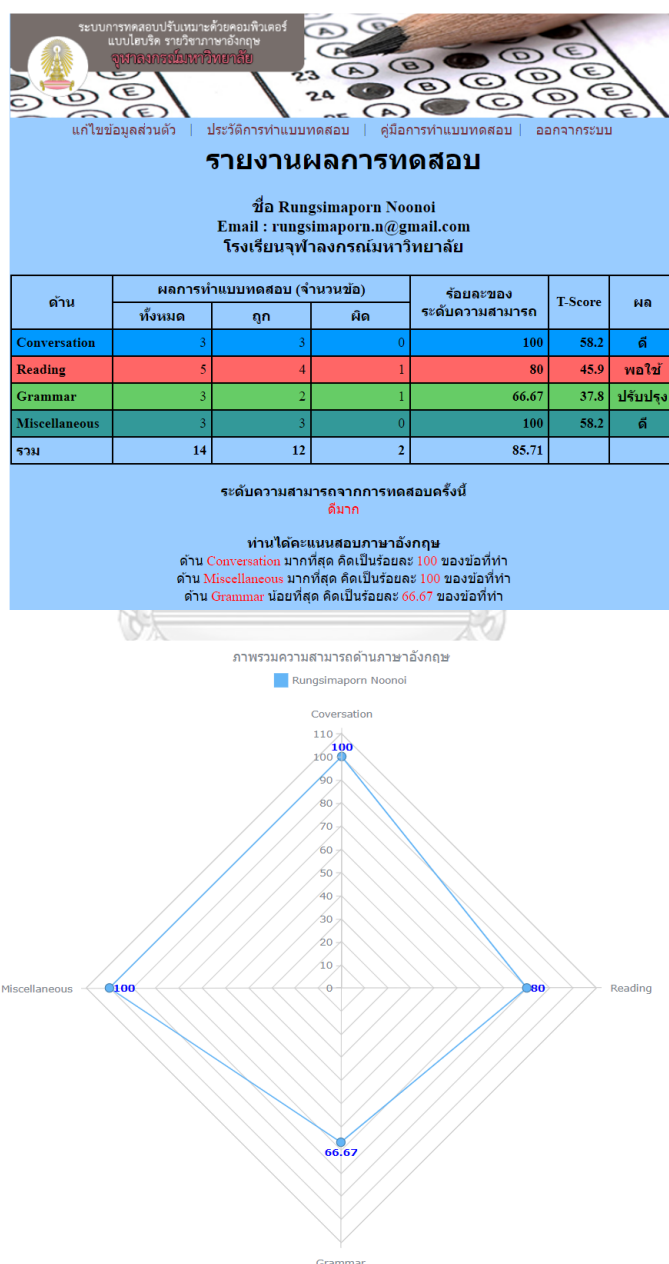
ภาพ 4.16 หน้าจอแสดงรายละเอียดคำชี้แจง



ภาพ 4.17 หน้าจอข้อสอบ

## การรายงานผลการทดสอบ

เมื่อผู้เข้ารับการทดสอบทำแบบทดสอบครบถ้วนจนถึงการยุติการทดสอบตามเกณฑ์ที่ระบบกำหนดไว้ จะขึ้นหน้าจอแสดงผลการทดสอบ โดยจะมีการรายงานด้านเนื้อหาในการทำการทดสอบ ผลการทำแบบทดสอบ ประกอบไปด้วย ข้อสอบทั้งหมด จำนวนข้อถูก จำนวนข้อผิด ร้อยละของระดับความสามารถของผู้สอบ ค่า T-score และผลการประเมิน และรวมทั้งนำเสนอผลการประเมินในรูปของกราฟ ดังแสดงในภาพ 4.18



ภาพ 4.18 หน้าจอรายงานผลการทดสอบ



จากภาพ 4.18 เมื่อทำการทดสอบเสร็จเรียบร้อยแล้ว ผู้เข้ารับการทดสอบสามารถประเมินความพึงพอใจที่มีต่อการใช้ระบบ โดยทำ “แบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด” ดังภาพ 4.19

ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด รายวิชาภาษาอังกฤษ จุฬาลงกรณ์มหาวิทยาลัย

**แบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด**

ขอให้นักเรียนเลือกตอบตรงความพึงพอใจของท่าน

ข้อที่	คำถาม	ความพึงพอใจ				
		น้อยที่สุด	น้อย	ปานกลาง	มาก	มากที่สุด
1	การทำงานของระบบสะดวก ง่าย ต่อการใช้งาน และไม่ซับซ้อน	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	ผู้ใช้งานสามารถเรียนรู้การใช้งานและ การทำงานของระบบด้วยตนเอง	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	กระบวนการทำงานของระบบมีความรวดเร็วในการเรียกใช้บริการ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	รูปแบบของข้อความ ส่วนประกอบต่างๆ บนหน้าจอมีความเหมาะสม	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	ลำดับการแสดงผลภาพและส่วนประกอบต่างๆ บนหน้าจอ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	การออกแบบการใช้สีที่เหมาะสมต่อการใช้งาน	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	ความรวดเร็วในการรายงานผลการทดสอบ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8	การรายงานผลการทดสอบมีความชัดเจน เข้าใจง่าย	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9	ผลการทดสอบนำไปใช้ในการเตรียมความพร้อมก่อนการทดสอบภาษาอังกฤษได้	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10	ความรู้สึกรวมที่มีต่อระบบ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

บันทึกข้อมูล

ภาพ 4.19 แบบประเมินความพึงพอใจต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

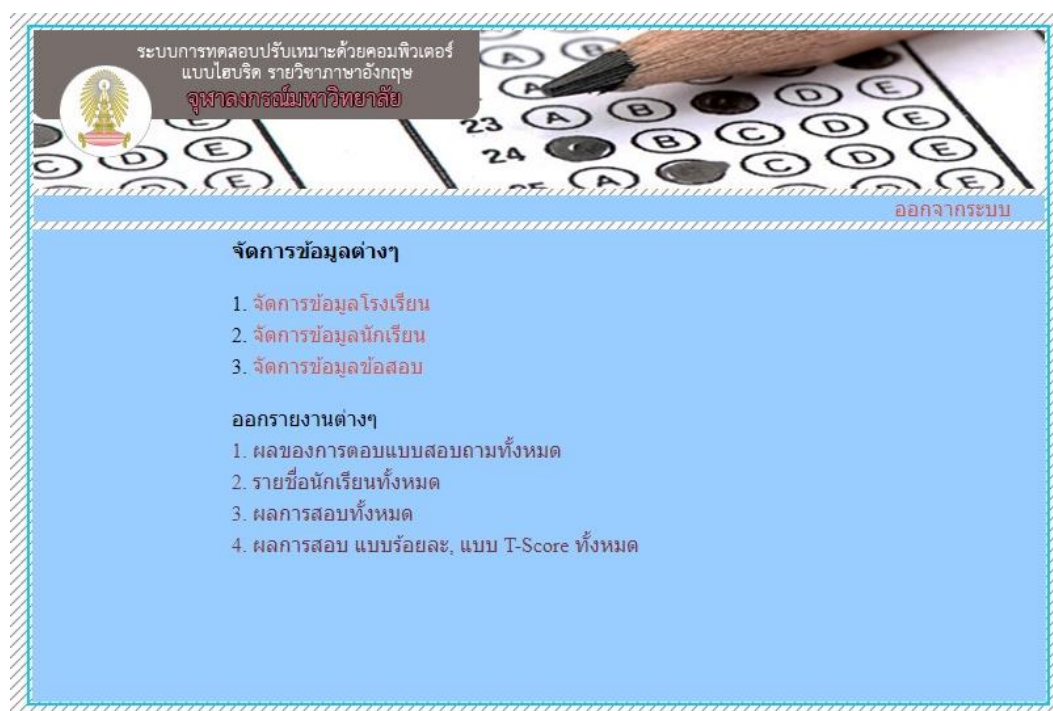
- คู่มือการใช้งานสำหรับผู้จัดการระบบ (admin) เพื่อดูแลระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีรายละเอียดขั้นตอนการใช้งานระบบ ได้แก่การจัดการข้อมูลโรงเรียน การจัดการข้อมูลนักเรียนผู้เข้ารับการทดสอบ การจัดการข้อมูลข้อสอบ และการออกรายงานผลการตอบแบบประเมินความพึงพอใจ และผลการสอบ รายละเอียดดังภาคผนวก

### คู่มือการใช้งานสำหรับผู้จัดการระบบ (admin)

คู่มือการใช้งานสำหรับผู้จัดการระบบ (admin) เพื่อใช้สำหรับจัดการคลังข้อสอบ และจัดการประวัติ การทำแบบทดสอบ การเข้าสู่ระบบจะต้องใช้รหัสผู้ใช้งานและรหัสผ่านที่ได้รับจากผู้พัฒนาโปรแกรม ดังภาพ 4.20 เมื่อเข้าสู่ระบบเรียบร้อยแล้ว หน้าจอหลักของผู้จัดการระบบ ประกอบด้วยข้อมูล การจัดการข้อมูลต่าง ๆ ออกรายงานต่าง ๆ ปรากฏรายละเอียดดังภาพ 4.21 โดยมีรายละเอียดดังนี้

1. การจัดการข้อมูลต่างๆ มีรายละเอียดดังนี้
  - จัดการข้อมูลโรงเรียน : เพิ่ม/แก้ไขข้อมูลโรงเรียน
  - จัดการข้อมูลนักเรียน : จัดการข้อมูลของผู้เข้ารับการทดสอบ
  - จัดการข้อมูลข้อสอบ : เพิ่ม/แก้ไข ข้อสอบในระบบ
- 2) ออกรายงานต่าง ๆ มีรายละเอียดดังนี้
  - ผลการตอบแบบประเมินทั้งหมด
  - รายชื่อนักเรียนทั้งหมด
  - ผลการสอบทั้งหมด
  - ผลการสอบแบบร้อยละ, T-score ทั้งหมด

ภาพ 4.20 การเข้าสู่ระบบของผู้จัดการระบบ

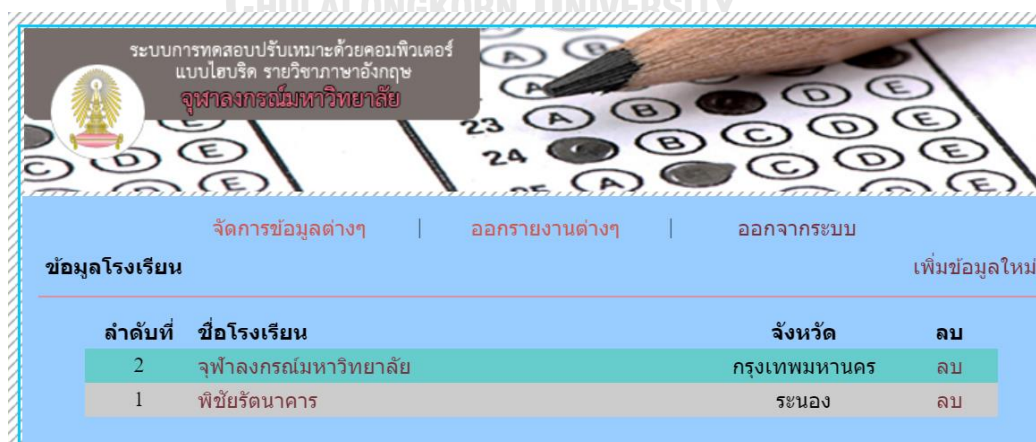


ภาพ 4.21 หน้าจอหลักของผู้จัดการระบบ

### การจัดการข้อมูล

#### 1. การจัดการข้อมูลโรงเรียน

ในส่วนของหน้าจอจัดการข้อมูลต่าง ๆ เมื่อเลือกในส่วนจัดการข้อมูลโรงเรียน จะปรากฏหน้าจอ ดังภาพ 4.22 ถ้าผู้จัดการระบบต้องการจะเพิ่มโรงเรียน ให้คลิก “เพิ่มข้อมูลใหม่” ถ้าผู้จัดการระบบจะลบโรงเรียน ให้เลือกโรงเรียน แล้ว คลิก “ลบ” โรงเรียนที่ต้องการจะลบ



ภาพ 4.22 การจัดการข้อมูลโรงเรียน

## 2. การจัดการข้อสอบ

ส่วนของการจัดการข้อสอบ ผู้จัดการระบบสามารถ เพิ่ม ซึ่งมีรายละเอียด ดังนี้ ประเภทข้อสอบ ระดับข้อสอบ ค่าพารามิเตอร์อำนาจจำแนก (a) ค่าพารามิเตอร์ความยาก (b) ค่าพารามิเตอร์การเดา (c) คำสั่ง การทดสอบ การเพิ่มรูปภาพ รายละเอียดประกอบสำหรับทักษะ Reading คำถามของข้อสอบ ตัวเลือกคำตอบ และคำตอบที่ถูกต้อง ได้ตามความต้องการ แล้วคลิก “บันทึกข้อสอบ” ปรากฏดังภาพ 4.24

จัดการข้อมูลต่างๆ		ออกรายงานต่างๆ		ออกจากระบบ	
<b>ข้อมูล - ข้อสอบ</b>					
<b>ข้อสอบ</b>					
ชุดที่ 1					
1	2	3	4	5	
ชุดที่ 2					
Low Level					
1	2	3	4	5	
Medium Level					
1	2	3	4	5	
High Level					
1	2	3	4	5	

ภาพ 4.23 ตัวอย่างชุดข้อสอบสำหรับการทดสอบขั้นแรกและขั้นที่สอง

**ข้อที่ 109**

เลือกประเภทข้อสอบ	ประเภทข้อสอบ	<input checked="" type="radio"/> C=Conversation <input type="radio"/> R=Reading <input type="radio"/> G=Grammar <input type="radio"/> M=Miscellaneous <input type="radio"/> L=Low <input type="radio"/> M=Medium <input checked="" type="radio"/> H=High
เลือกระดับของข้อสอบ	ระดับข้อสอบ	<input type="text" value="0"/> <input type="text" value="0"/> <input type="text" value="0"/> <input type="text" value="0"/>
กรอกคำพารามิเตอร์อ่านจำแนก	คำ a	
กรอกคำพารามิเตอร์ความยาก	คำ b	
กรอกคำพารามิเตอร์การเดา	คำ c	
กรอกคำสั่งของข้อสอบ	คำสั่ง	
กรอกรายละเอียดประกอบสำหรับทักษะ Reading	รูปโจทย์ Reading	
กรอกคำถามของข้อสอบ	คำถาม	
	ตัวเลือก A	
กรอกตัวเลือกคำตอบ	ตัวเลือก B	
	ตัวเลือก C	
	ตัวเลือก D	
เลือกคำตอบที่ถูกต้อง	คำตอบ	<input checked="" type="radio"/> A <input type="radio"/> B <input type="radio"/> C <input type="radio"/> D <input type="button" value="บันทึกข้อสอบ"/>
อัปโหลดรูปภาพ	โจทย์ที่มีรูปประกอบ	Upload รูปภาพ <input type="button" value="เลือกไฟล์"/> <input type="button" value="ไม่ได้เลือกไฟล์ใด"/> (*.jpg เท่านั้น) <input type="button" value="บันทึกรูปภาพ"/>

ภาพ 4.24 การเพิ่มข้อสอบ

ผู้จัดการระบบสามารถแก้ไขข้อสอบจากหน้าจอหลักผู้จัดการระบบ โดย คลิก “จัดการข้อมูลข้อสอบ” ตามภาพ 4.21 จะปรากฏภาพ 4.25 เลือกข้อสอบที่จะแก้ไข แล้วคลิก “แก้ไขข้อสอบ” ผู้จัดการระบบสามารถแก้ไขข้อสอบตามต้องการ ดังภาพ 4.26 แล้วคลิก “บันทึกข้อสอบ”

ระบบการทดสอบปรับคะแนนด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

จัดการข้อมูลต่างๆ | ออกรายงานต่างๆ | ออกจากระบบ

ข้อสอบ

## ชุดที่ 1

### ข้อที่ 1

ประเภทข้อสอบ : C=Conversation  
ระดับข้อสอบ : M=Medium  
ค่า a : 1.413  
ค่า b : 0.054  
ค่า c : 0.268

1. Read each situation and the dialogue carefully and choose the appropriate expression to complete the dialogue.

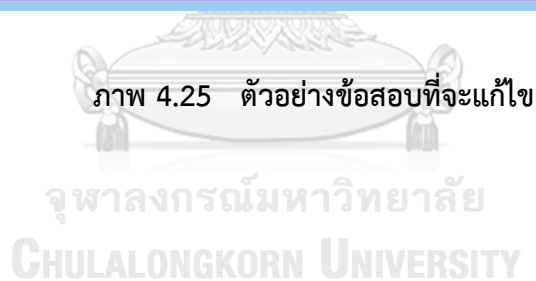
Situation: Your brother has won a scholarship to study in Canada.  
You congratulate him by saying, "\_\_\_\_\_"

A. How pretty you look!  
B. Very well. How did it go?  
C. That's great! How lucky you are!  
D. Amazing! I never expected so much.

Answer : C

แก้ไขข้อสอบ

คลิก เพื่อแก้ไขข้อสอบ





**ข้อที่ 1**

ประเภทข้อสอบ ☒ C=Conversation ☐ R=Reading ☐ G=Grammar ☐ M=Miscellaneous

ระดับข้อสอบ ☐ L=Low ☒ M=Medium ☐ H=High

ค่า a 1.413  
ค่า b 0.054  
ค่า c 0.268

คำสั่ง Read each situation and the dialogue carefully and choose the appropriate expression to complete the dialogue.

รูปโจทย์ Reading

คำถาม Situation: Your brother has won a scholarship to study in Canada. <br> You congratulate him by saying, "\_\_\_\_\_"

ตัวเลือก A How pretty you look!

ตัวเลือก B Very well. How did it go?

ตัวเลือก C That's great! How lucky you are!

ตัวเลือก D Amazing! I never expected so much.

คำตอบ ☐ A ☐ B ☒ C ☐ D

บันทึกข้อสอบ

โจทย์ที่มีรูปประกอบ

Upload รูปภาพ   (\*.jpg เท่านั้น)

แก้ไขข้อมูลตามต้องการ

กดบันทึกข้อสอบเมื่อแก้ไขข้อมูลเสร็จเรียบร้อยแล้ว

ภาพ 4.26 ตัวอย่างการแก้ไขข้อสอบ

### 3. การจัดการสมาชิก

จากหน้าจอหลักภาพ 4.21 ผู้จัดการระบบต้องการ เพิ่ม ลบ หรือแก้ไขข้อมูลของผู้เข้ารับการทดสอบ โดย คลิก “จัดการข้อมูลนักเรียน” เลือกสมาชิกที่ต้องการแก้ไข ระบบจะแสดงข้อมูลของสมาชิก ผู้จัดการระบบสามารถแก้ไขข้อมูลสมาชิก เมื่อแก้ไขข้อมูลสมาชิกเสร็จแล้ว คลิก “บันทึกการแก้ไขข้อมูล” ดังภาพ 4.27

ภาพ 4.27 ตัวอย่างการแก้ไขข้อมูลของสมาชิก

### การรายงานผลการทดสอบ

ผู้จัดการระบบสามารถดูรายงานผลการทดสอบของสมาชิกได้ โดยจากหน้าจอหลักภาพ 4.21 เลือกออกรายงานต่างๆ ซึ่งมีรายละเอียด ดังนี้

- ผลการตอบแบบสอบถามทั้งหมด
- รายชื่อนักเรียนทั้งหมด
- ผลการสอบทั้งหมด
- ผลการสอบแบบร้อยละ, T-score ทั้งหมด

ผู้จัดการระบบสามารถเลือกการออกรายงานตามต้องการ โดยการคลิกรายงานที่ต้องการ ข้อมูลจะเป็นไฟล์ Excel



## 2.2 การทดสอบ

การทดสอบ สำหรับหน้าจอการทดสอบเมื่อผู้รับการทดสอบกดปุ่มเข้าสู่ระบบการทดสอบ ระบบจะแสดงคำชี้แจงรายละเอียดเกี่ยวกับการทดสอบ ดังภาพ 4.16 เมื่อกดปุ่มเริ่มทำข้อสอบแล้วระบบจะแสดงข้อสอบครั้งละ 1 ข้อ พร้อมตัวเลือกสำหรับการทำข้อสอบแต่ละข้อและระยะเวลาสำหรับการทำแบบทดสอบ ดังภาพ 4.17 ในการทดสอบผู้เข้ารับการทดสอบจะได้รับข้อสอบครั้งละจำนวน 1 ข้อ โดยมีการมุ่งวัดความรู้ทางภาษาอังกฤษ ซึ่งประกอบด้วย 4 ด้าน ได้แก่ Conversation, Grammar and Vocabulary, Reading Comprehension และ Miscellaneous โดยรายละเอียดดังกล่าวจะแสดงอยู่ด้านบนของแบบทดสอบ พร้อมระยะเวลาในการทดสอบ สำหรับระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดมีส่วนสำคัญ 5 ส่วน คือ 1) จุดเริ่มต้นการทดสอบ 2) การประมาณค่าความสามารถของผู้สอบ 3) การคัดเลือกข้อสอบ 4) การควบคุมการใช้ข้อสอบซ้ำ และ 5) การยุติการทดสอบ

### 1) จุดเริ่มต้นการทดสอบ

ในการเริ่มต้นการทดสอบด้วยการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด จะเริ่มต้นด้วยการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) โดยชุดข้อสอบที่กำหนดขึ้น (initial item) (MST ขั้นที่ 1) ที่มีความยากปานกลาง (ค่า  $b$  อยู่ระหว่าง -0.5 ถึง 0.5) เมื่อไม่ทราบค่าสารสนเทศเบื้องต้นของผู้สอบ โดยมีชุดข้อสอบที่กำหนดขึ้นมีข้อสอบครอบคลุมความรู้ทางภาษาอังกฤษทั้ง 4 ด้าน ได้แก่ Conversation, Grammar and Vocabulary, Reading Comprehension และ Miscellaneous

### 2) การประมาณค่าความสามารถของผู้สอบ (ability estimation)

สำหรับการประมาณค่าความสามารถของผู้สอบใช้วิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (maximum likelihood estimation: MLE) คัดเลือกข้อสอบที่มีตำแหน่งสารสนเทศสูงสุด ( $m_i$ ) สอดคล้องกับระดับความสามารถที่ประมาณได้ ( $\theta$ ) [Match  $m_i$  to  $\theta$ ] โดยระบบจะทำการวิเคราะห์ความสามารถของผู้สอบ ( $\theta$ ) ด้วยความเป็นไปได้สูงสุด (maximum likelihood : MLE) ที่มีความยากใกล้เคียงกับความสามารถของผู้สอบ และระบบจะคัดเลือกข้อสอบชุดถัดไปที่มีความยากใกล้เคียงกับความสามารถของผู้สอบ (MST ขั้นที่ 2) ในขั้นตอนนี้จะมี การแบ่งโมดูลข้อสอบออกเป็น 3 ระดับ ได้แก่ ระดับง่าย ระดับปานกลาง และระดับยาก เมื่อผู้สอบทำชุดข้อสอบตามระดับความสามารถ ระบบจะประมาณค่าความสามารถของผู้สอบจากการตอบข้อสอบในชุดที่ 2 เพื่อคัดเลือกข้อสอบ

### 3) การคัดเลือกข้อสอบ (item selection)

กระบวนการคัดเลือกข้อสอบสามารถนำมาใช้กับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ซึ่งในแต่ละขั้น (stage) จะทำการเลือกข้อสอบที่มีสารสนเทศสูงสุดภายหลังการประมาณค่าความสามารถของผู้สอบ การคัดเลือกข้อสอบเป็นสิ่งสำคัญในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด วิธีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) เป็นวิธีการคัดเลือกข้อสอบที่เพิ่มประสิทธิภาพจากการใช้ค่าฟังก์ชันสารสนเทศสูงสุด (maximum item information function) แบบดั้งเดิม เป็นการกำหนดค่าถ่วงน้ำหนักตามเงื่อนไขที่กำหนด ทำให้การคัดเลือกข้อสอบมีความแม่นยำมากยิ่งขึ้น โดยการคัดเลือกข้อสอบวิธีลำดับความสำคัญสูงสุด ระบบจะคัดเลือกชุดข้อสอบและคัดเลือกข้อสอบที่ละข้อ ซึ่งมีการจัดระดับข้อสอบ ออกเป็น 3 ระดับ ได้แก่ ระดับง่าย ระดับปานกลาง และระดับยาก ตามระดับความสามารถของผู้สอบ

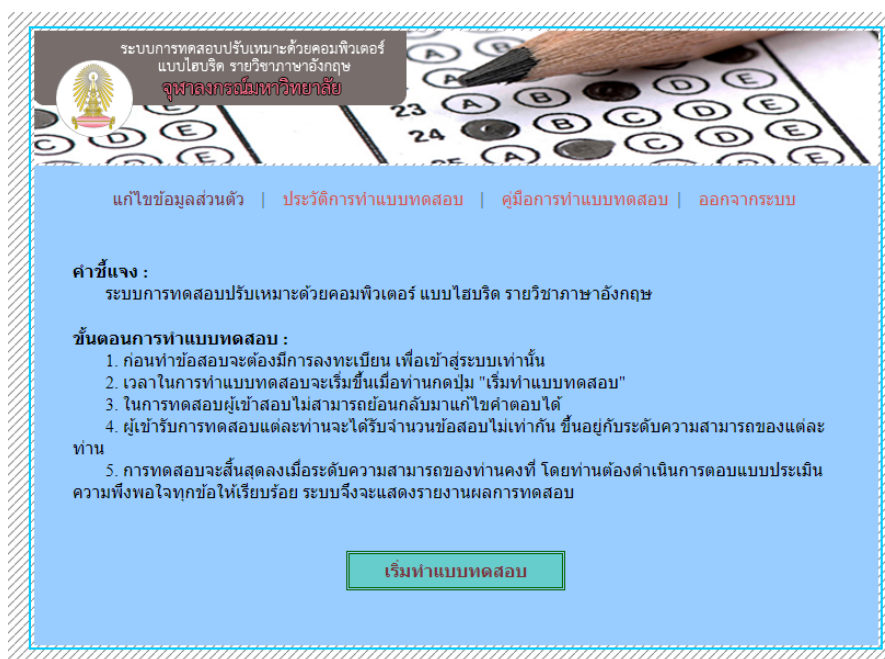
### 4) การควบคุมการใช้ข้อสอบซ้ำ (exposure control)

ในการควบคุมการเปิดเผยของข้อสอบสามารถควบคุมได้ด้วยกระบวนการอัตโนมัติ เหมือนกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ โดยเราสามารถป้องกันข้อสอบที่ถูกนำมาใช้บ่อย (over-exposed) จากกรณีที่มีการเปิดเผยและมีการแบ่งปันข้อสอบกับผู้สอบอื่น วิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PR-SE) เป็นวิธีที่นำอัตราส่วนของความคลาดเคลื่อนมาตรฐาน (SE) มาใช้ในกฎการหยุด โดยการใช้ความคลาดเคลื่อนมาตรฐาน (SE) เพื่อตรวจสอบองค์ประกอบการสุ่ม หรือ สารสนเทศที่มีอิทธิพลต่อการถ่วงน้ำหนัก และเหมาะสมกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบความยาวแปรผันในงานวิจัยนี้ เนื่องจากจะกำหนดค่าความคลาดเคลื่อนมาตรฐาน (SE) ของการประมาณค่าความสามารถผู้สอบ โดยมีอัตราการใช้ข้อสอบซ้ำสูงสุด เท่ากับ 20 เปอร์เซ็นต์

### 5) การยุติการทดสอบ (Stopping criteria)

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เป็นการทดสอบแบบความยาวแปรผัน (variable length) พิจารณาจากค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) น้อยกว่าหรือเท่ากับ 0.3 ในการทดสอบผู้เข้ารับการทดสอบต้องทำข้อสอบ จำนวน 11 ข้อ เป็นอย่างน้อย หากเมื่อผู้เข้ารับการทดสอบมีผลการทดสอบในการทดสอบมีความคงเส้นคงวา โดยที่ค่าความคลาดเคลื่อนมาตรฐาน น้อยกว่าหรือเท่ากับ 0.3 ระบบจะทำการยุติการทดสอบและจะแสดงหน้า แบบประเมินความพึงพอใจของระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยผู้เข้ารับการทดสอบต้องทำการประเมินคุณภาพทุกคน หากไม่ทำการประเมินระบบจะไม่ออกรายงานผลให้ผู้เข้ารับการทดสอบหากผู้เข้ารับการทดสอบดำเนินการทำข้อสอบแล้วปรากฏว่าค่าความคลาดเคลื่อนมาตรฐานยังมากกว่า 0.3 ระบบจะทำการเลือกข้อสอบให้ผู้เข้ารับการทดสอบทำทีละข้อไปเรื่อยๆ จนกว่าผู้เข้ารับการทดสอบจะมีค่าความคลาดเคลื่อน

มาตรฐาน น้อยกว่าหรือเท่ากับ 0.3 หลังจากนั้นผู้เข้ารับการทดสอบต้องทำการประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดเพื่อเข้าดูรายงานผลการทดสอบ



ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

แก้ไขข้อมูลส่วนตัว | ประวัติการทำแบบทดสอบ | คู่มือการทำแบบทดสอบ | ออกจากระบบ

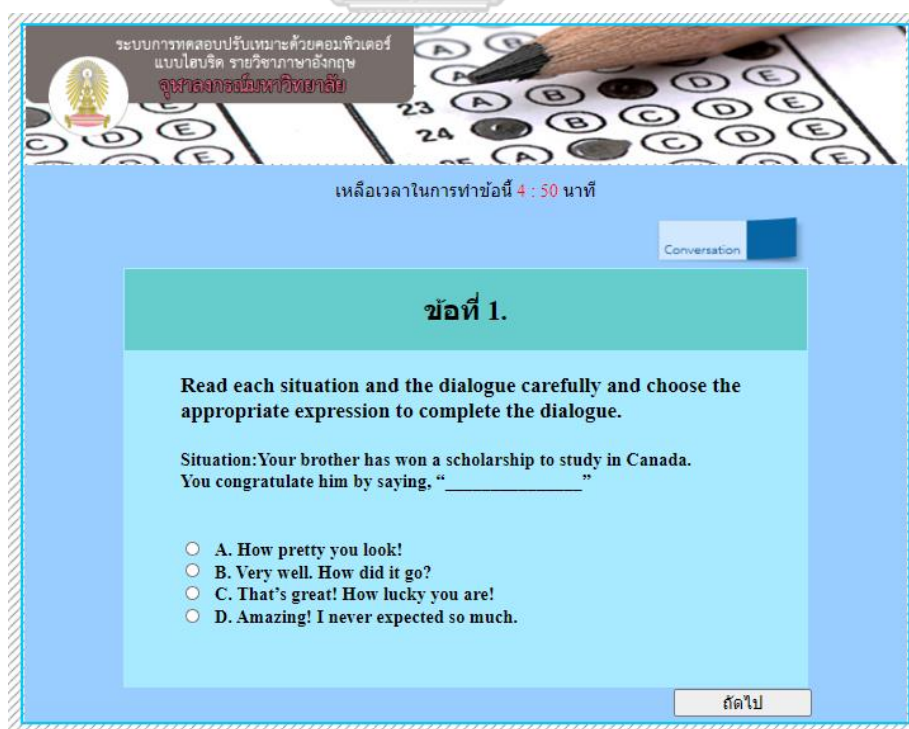
**คำชี้แจง :**  
ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ แบบไฮบริด รายวิชาภาษาอังกฤษ

**ขั้นตอนการทำแบบทดสอบ :**

1. ก่อนทำข้อสอบจะต้องมีการลงทะเบียน เพื่อเข้าสู่ระบบเท่านั้น
2. เวลาในการทำแบบทดสอบจะเริ่มขึ้นเมื่อท่านกดปุ่ม "เริ่มทำแบบทดสอบ"
3. ในการทดสอบผู้เข้าสอบไม่สามารถย้อนกลับมาแก้ไขคำตอบได้
4. ผู้เข้ารับการทดสอบแต่ละท่านจะได้รับจำนวนข้อสอบไม่เท่ากัน ขึ้นอยู่กับระดับความสามารถของแต่ละท่าน
5. การทดสอบจะสิ้นสุดลงเมื่อระดับความสามารถของท่านคงที่ โดยท่านต้องดำเนินการตอบแบบประเมินความพึงพอใจทุกข้อให้เรียบร้อย ระบบจึงจะแสดงรายงานผลการทดสอบ

**เริ่มทำแบบทดสอบ**

ภาพ 4.28 หน้าจอแสดงรายละเอียดคำชี้แจง



ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

แก้ไขข้อมูลส่วนตัว | ประวัติการทำแบบทดสอบ | คู่มือการทำแบบทดสอบ | ออกจากระบบ

เหลือเวลาในการทำข้อนี้ 4:50 นาที

Conversation

**ข้อที่ 1.**

Read each situation and the dialogue carefully and choose the appropriate expression to complete the dialogue.

Situation: Your brother has won a scholarship to study in Canada.  
You congratulate him by saying, "\_\_\_\_\_"

☐ A. How pretty you look!

☐ B. Very well. How did it go?

☐ C. That's great! How lucky you are!

☐ D. Amazing! I never expected so much.

**ถัดไป**

ภาพ 4.29 หน้าจอข้อสอบ



ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย



### แบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

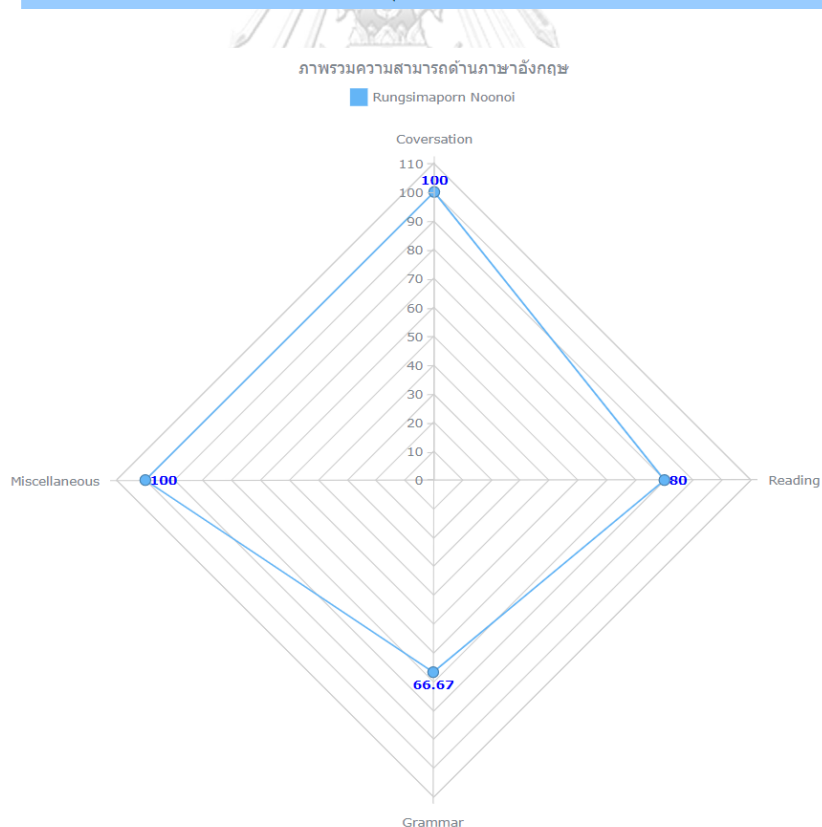
ขอให้นักเรียนเลือกตอบตรงความพึงพอใจของท่าน

ข้อ ที่	คำถาม	ความพึงพอใจ				
		น้อยที่สุด	น้อย	ปานกลาง	มาก	มากที่สุด
1	การทำงานของระบบสะดวก ง่าย ต่อการใช้งาน และ ไม่ซับซ้อน	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	ผู้ใช้งานสามารถเรียนรู้การใช้งานและ การทำงาน ของระบบด้วยตนเอง	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	กระบวนการทำงานของระบบมีความรวดเร็วในการ เรียกใช้บริการ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	รูปแบบของข้อความ ส่วนประกอบต่างๆ บนหน้าจรมี ความเหมาะสม	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	ลำดับการแสดงผลภาพและส่วนประกอบต่างๆ บนหน้า จอ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	การออกแบบการใช้สีที่เหมาะสมต่อการใช้งาน	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	ความรวดเร็วในการรายงานผลการทดสอบ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8	การรายงานผลการทดสอบมีความชัดเจน เข้าใจง่าย	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9	ผลการทดสอบนำไปใช้ในการเตรียมความพร้อม ก่อนการทดสอบภาษาอังกฤษได้	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10	ความรู้สึกรวมที่มีต่อระบบ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

บันทึกข้อมูล

ภาพ 4.30 แบบประเมินความพึงพอใจต่อระบบการทดสอบปรับเหมาะ  
ด้วยคอมพิวเตอร์แบบไฮบริด

2.3 การรายงานผลการทดสอบ การรายงานการทดสอบประกอบด้วยข้อมูลส่วนตัว  
ของกลุ่มตัวอย่างในการวิจัย ได้แก่ ชื่อ – นามสกุล อีเมล (E-mail) และโรงเรียน รายงานการทดสอบ  
แบ่งออกเป็น 2 ส่วน คือ 1) รายงานผลการตอบข้อสอบ ร้อยละ และ ค่ามาตรฐานที (T-score) และ  
2) กราฟแสดงความสัมพันธ์ของผลการทดสอบในแต่ละด้าน ดังภาพ 4.31



ภาพ 4.31 รายงานผลการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

### 2.3 การทดลองใช้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ผู้วิจัยทดลองใช้ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนพิชัยรัตนาคาร จังหวัดระนอง จำนวน 74 คน โดยผู้ทดลองใช้ระบบเป็นเพศชาย จำนวน 42 คน คิดเป็น 56.76 และเพศหญิง 32 คน คิดเป็นร้อยละ 43.24 ดังตาราง 4.10

**ตาราง 4.10 ข้อมูลพื้นฐานของนักเรียนที่ทดลองใช้ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด**

เพศ	จำนวน	ร้อยละ
ชาย	42	56.76
หญิง	32	43.24
รวม	74	100.00

โดยมีรายละเอียด ดังนี้

#### 1) ผลการวัดความสามารถด้านทักษะภาษาอังกฤษ

นักเรียนผู้ทดลองใช้ระบบมีความสามารถด้านทักษะภาษาอังกฤษ เฉลี่ยเท่ากับ  $-0.18$  ( $SD=0.85$ ) มีการแจกแจงแบบเบ้ซ้าย ( $sk = -1.33$ ) และมีความโด่งมากกว่าโค้งปกติ ( $ku = 1.46$ ) แสดงว่าความสามารถด้านทักษะภาษาอังกฤษของนักเรียนมีการกระจายมากกว่าโค้งปกติ พบว่านักเรียนชายมีความสามารถด้านทักษะภาษาอังกฤษเฉลี่ยเท่ากับ  $-0.21$  ( $0.67$ ) มีการแจกแจงแบบเบ้ซ้าย ( $sk = -1.35$ ) และมีความโด่งมากกว่าโค้งปกติ ( $ku = 1.25$ ) แสดงว่าความสามารถด้านทักษะภาษาอังกฤษของนักเรียนชายมีการกระจายมากกว่าโค้งปกติ และ นักเรียนหญิงมีความสามารถด้านทักษะภาษาอังกฤษเฉลี่ยเท่ากับ  $-0.14$  ( $0.70$ ) มีการแจกแจงแบบเบ้ซ้าย ( $sk = -1.13$ ) และมีความโด่งมากกว่าโค้งปกติ ( $ku = 1.17$ ) แสดงว่าความสามารถด้านทักษะภาษาอังกฤษของนักเรียนหญิงมีการกระจายมากกว่าโค้งปกติ ดังตาราง 4.11

**ตาราง 4.11 ค่าสถิติพื้นฐานของความสามารถด้านทักษะภาษาอังกฤษของนักเรียนระดับชั้นมัธยมศึกษาชั้นปีที่ 3**

เพศ	M	SD	Min	Max	Sk(SE)	Ku(SE)
ชาย	-0.21	0.67	-2.29	0.51	-1.35 (0.37)	1.25 (0.72)
หญิง	-0.14	0.70	-2.05	0.95	-1.13 (0.41)	1.17 (0.81)
รวม	-0.18	0.85	-2.05	0.95	-1.33 (0.28)	1.46 (0.55)

ค่าสถิติพื้นฐานของคะแนนทักษะภาษาอังกฤษ แต่ละด้านดังตาราง 4.12

ตาราง 4.12 ค่าสถิติพื้นฐานของคะแนนทักษะภาษาอังกฤษ

ด้าน	M	SD	Min	Max	Sk(SE)	Ku(SE)
Conversation	60.33	23.60	0.00	100.00	-0.25 (0.27)	-0.44 (0.54)
Grammar and Vocabulary	54.96	19.94	11.11	100.00	-0.01 (0.27)	-0.51 (0.54)
Reading Comprehension	67.50	28.23	5.56	100.00	-0.29 (0.27)	-1.15 (0.54)
Miscellaneous	58.13	29.66	0.00	100.00	0.10 (0.27)	-1.06 (0.54)
รวม	59.62	19.18	21.43	85.71	-0.39 (0.27)	-1.03 (0.54)

## ระยะที่ 2 ผลการประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ผลการประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด แบ่งออกเป็น 3 ตอน ได้แก่ ตอนที่ 1 ผลการประเมินแบบอิงมาตรฐาน (evaluation standard) ตอนที่ 2 การประเมินระบบแบบฮิวริสติก (heuristic evaluation) และ ตอนที่ 3 การประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยมีรายละเอียดดังนี้

### ตอนที่ 1 ผลการประเมินแบบอิงมาตรฐาน

ผลการประเมินแบบอิงมาตรฐานของระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยผู้เชี่ยวชาญด้านเทคโนโลยี มีรายละเอียดดังนี้

ผลการตรวจสอบคุณภาพของระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ในการอิงมาตรฐาน ด้วยแบบประเมินแบบอิงมาตรฐานมีลักษณะเป็นมาตรประมาณค่า 5 ระดับ ประกอบด้วยเนื้อหา 4 หมวด ได้แก่ ความมีประโยชน์ (utility) ความเป็นไปได้ (feasibility) ความเหมาะสม (propriety) และความถูกต้อง (accuracy) ประเมินโดยผู้เชี่ยวชาญด้านเทคโนโลยี จำนวน 5 ท่าน พบว่า ภาพรวมระบบ การทดสอบมีคุณภาพอยู่ในระดับมากที่สุด ( $M = 4.59$ ,  $SD = 0.35$ ) หากพิจารณาคุณภาพของระบบการทดสอบในแต่ละด้าน พบว่า ด้านความมีประโยชน์ ( $M = 4.53$ ,  $SD = 0.56$ ) ด้านความเป็นไปได้ ( $M = 4.65$ ,  $SD = 0.42$ ) ด้านความเหมาะสม ( $M = 4.60$ ,  $SD = 0.22$ ) และด้านความถูกต้อง ( $M = 4.53$ ,  $SD = 0.65$ ) มีคุณภาพอยู่ในระดับมากที่สุดทั้ง 4 ด้าน โดยมีผลการวิเคราะห์ดังตาราง 4.13

ตาราง 4.13 ผลการตรวจสอบคุณภาพของระบบการทดสอบในการประเมินอิงมาตรฐาน

ข้อคำถาม	ระดับความคิดเห็น				
	น้อยที่สุด n (%)	น้อย n (%)	ปานกลาง n (%)	มาก n (%)	มากที่สุด n (%)
<b>ความมีประโยชน์ (utility) (M = 4.53, SD = 0.56)</b>					
1. ระบบตอบสนองความต้องการของผู้ใช้งาน	-	-	-	2 (40.00)	3 (60.00)
2. การรายงานผลมีความชัดเจน เข้าใจง่าย	-	-	-	3 (60.00)	2 (40.00)
3. ระบบการทดสอบสามารถนำไปใช้ประโยชน์ในการพัฒนาทางด้านทักษะภาษาอังกฤษของผู้สอบได้	-	-	-	1 (20.00)	4 (80.00)
<b>ความเป็นไปได้ (feasibility) (M = 4.65, SD = 0.42)</b>					
4. กระบวนการและขั้นตอนของระบบสามารถนำไปปฏิบัติได้จริง	-	-	-	2 (40.00)	3 (60.00)
5. ระบบมีการเรียกใช้งานง่าย สะดวก และรวดเร็ว	-	-	-	2 (40.00)	3 (60.00)
6. ผลที่ได้จากการทดสอบมีความคุ้มค่า	-	-	-	1 (20.00)	4 (80.00)
7. คำถามควรเปลี่ยนเป็นระบบมีความสอดคล้องตามสถานการณ์ปัจจุบัน	-	-	-	2 (40.00)	3 (60.00)
<b>ความเหมาะสม (propriety) (M = 4.60, SD = 0.22)</b>					
8. ระบบการทดสอบมีคู่มือแนะนำการใช้งานที่เหมาะสมเข้าใจง่าย	-	-	-	1 (20.00)	4 (80.00)
9. ขอบเขตของเนื้อหาในการทดสอบมีความเหมาะสม	-	-	-	1 (20.00)	4 (80.00)
10. ระยะเวลาในการทดสอบมีความเหมาะสม	-	-	1 (20.00)	2 (40.00)	2 (40.00)
11. การรายงานผลการทดสอบมีความครบถ้วนครอบคลุม	-	-	-	2 (40.00)	3 (60.00)
<b>ความถูกต้อง (accuracy) (M = 4.53, SD = 0.65)</b>					
12. ระบบจัดการทดสอบได้ตรงตามวัตถุประสงค์	-	-	-	2 (40.00)	3 (60.00)
13. กระบวนการทดสอบของระบบมีความถูกต้อง	-	1 (20.00)	-	1 (20.00)	3 (60.00)
14. การรายงานผลคะแนนมีความถูกต้องตรงตามคุณลักษณะที่แท้จริง	-	-	-	1 (20.00)	4 (80.00)
<b>รวม (M = 4.59, SD = 0.35)</b>	-	1 (1.43)	1 (1.43)	23 (32.85)	45 (64.29)



## ตอนที่ 2 ผลการประเมินระบบแบบฮิวริสติก

ผลการตรวจสอบคุณภาพของระบบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด ในการประเมินฮิวริสติก ด้วยแบบประเมินฮิวริสติกมีลักษณะเป็นมาตราประมาณค่า 5 ระดับ ประกอบด้วยเนื้อหาหลักในการประเมินทั้งสิ้น 10 หมวด ได้แก่ ความชัดเจนของระบบ ความสอดคล้องของระบบและการใช้งานจริง การควบคุมและอิสระของผู้ใช้งาน ความสอดคล้องและมาตรฐาน การป้องกันความผิดพลาด การจำข้อมูลมากกว่าการเรียกคืนข้อมูล ความยืดหยุ่นและประสิทธิภาพการใช้งาน การออกแบบที่สวยงาม การช่วยเหลือโดยการกู้คืนระบบจากข้อผิดพลาด และการช่วยเหลือและคู่มือ ประเมินโดยผู้เชี่ยวชาญด้านเทคโนโลยี จำนวน 5 ท่าน พบว่า ภาพรวมระบบการทดสอบมีคุณภาพอยู่ในระดับมาก ( $M = 4.42$ ,  $SD = 0.33$ ) หากพิจารณาคุณภาพของระบบการทดสอบในแต่ละข้อ พบว่า การทำงานของระบบมีความเสถียร ( $M = 4.80$ ,  $SD = 0.55$ ) ระบบมีความยืดหยุ่นและประสิทธิภาพของการใช้งาน ( $M = 5.00$ ,  $SD = 0.00$ ) ระบบมีการตรวจสอบความพึงพอใจและการยอมรับปฏิสัมพันธ์กับผู้ใช้งาน ( $M = 4.80$ ,  $SD = 0.45$ ) มีการแสดงสถานะของระบบอย่างชัดเจน ( $M = 4.60$ ,  $SD = 0.55$ ) ระบบมีความสอดคล้องกับโลกของความเป็นจริง ( $M = 4.60$ ,  $SD = 0.55$ ) การควบคุมระบบมีประสิทธิภาพและผู้ใช้งานมีความสะดวกในการเข้าถึงจริง ( $M = 4.60$ ,  $SD = 0.55$ ) ระบบมีการป้องกันข้อมูลส่วนบุคคล ( $M = 4.60$ ,  $SD = 0.55$ ) อยู่ในระดับมากที่สุด ระบบมีการป้องกันความผิดพลาด ( $M = 4.40$ ,  $SD = 0.55$ ) ระบบมีความสามารถในการจดจำข้อมูลมากกว่าการเรียกข้อมูล ( $M = 4.20$ ,  $SD = 0.45$ ) ระบบมีรูปแบบสวยงามและเรียบง่าย ( $M = 3.60$ ,  $SD = 0.89$ ) ระบบมีการช่วยเหลือผู้ใช้งานระบุตัวตน วินิจฉัยและกู้คืนจากข้อผิดพลาด ( $M = 4.40$ ,  $SD = 0.89$ ) ระบบมีการช่วยเหลือและคู่มือการใช้งาน ( $M = 4.40$ ,  $SD = 0.55$ ) ระบบมีการสนับสนุนและเพิ่มทักษะของผู้ใช้งาน ( $M = 4.40$ ,  $SD = 0.55$ ) อยู่ในระดับมาก ผลการวิเคราะห์ดังตาราง 4.14

ตาราง 4.14 ผลการตรวจสอบคุณภาพของระบบการทดสอบในการประเมินอิทธิพล

ข้อคำถาม	ระดับความคิดเห็น					M	SD
	น้อยที่สุด n (%)	น้อย n (%)	ปานกลาง n (%)	มาก n (%)	มากที่สุด n (%)		
1. มีการแสดงสถานะของระบบอย่างชัดเจน	-	-	-	2 (40.00)	3 (60.00)	4.60	0.55
2. ระบบมีความสอดคล้องกับโลกของความเป็นจริง	-	-	-	2 (40.00)	3 (60.00)	4.60	0.55
3. การควบคุมระบบมีประสิทธิภาพและผู้ใช้งานมีความสะดวกในการเข้าถึง	-	-	-	2 (40.00)	3 (60.00)	4.60	0.55
4. การทำงานของระบบมีความเสถียร	-	-	-	1 (20.00)	4 (80.00)	4.80	0.45
5. ระบบมีการป้องกันความผิดพลาด	-	-	-	3 (60.00)	2 (40.00)	4.40	0.55
6. ระบบมีความสามารถในการจดจำข้อมูลมากกว่าการเรียกข้อมูล	-	-	-	4 (80.00)	1 (20.00)	4.20	0.45
7. ระบบมีความยืดหยุ่นและประสิทธิภาพของการใช้งาน	-	-	-	-	5 (100.00)	5.00	0.00
8. ระบบมีรูปแบบสวยงามและเรียบง่าย	-	-	-	1 (20.00)	4 (80.00)	3.60	0.89
9. ระบบมีการช่วยเหลือผู้ใช้งานระบบตัวตน วินิจฉัยและกู้คืนจากข้อผิดพลาด	-	-	1 (20.00)	1 (20.00)	3 (60.00)	4.40	0.89
10. ระบบมีการช่วยเหลือและคู่มือการใช้งาน	-	-	-	3 (60.00)	2 (40.00)	4.40	0.55
11. ระบบมีการสนับสนุนและเพิ่มทักษะของผู้ใช้งาน	-	-	-	3 (60.00)	2 (40.00)	4.40	0.55
12. ระบบมีการตรวจสอบความพึงพอใจและการยอมรับปฏิสัมพันธ์กับผู้ใช้งาน	-	-	-	1 (20.00)	4 (80.00)	4.80	0.45
13. ระบบมีการป้องกันข้อมูลส่วนบุคคล	-	-	-	2 (40.00)	3 (60.00)	4.60	0.55
รวม	-	-	1 (1.54)	25 (38.46)	39 (60.00)	4.42	0.33

จากผลการตรวจสอบคุณภาพของระบบการทดสอบที่เหมาะสมด้วยคอมพิวเตอร์แบบไฮบริด โดยผู้เชี่ยวชาญด้านเทคโนโลยี มีข้อเสนอแนะและการปรับปรุงแก้ไขตามข้อเสนอแนะดังตาราง 4.15

ตาราง 4.15 การปรับปรุงแก้ไขตามข้อเสนอแนะของผู้เชี่ยวชาญ

ข้อเสนอแนะ	การปรับปรุงแก้ไขตามข้อเสนอแนะ
1) ควรเพิ่มเติมการแสดงผลระหว่างการทำทดสอบ	1) เพิ่มการแสดงผลระยะเวลาให้ผู้รับการทดสอบทราบในระหว่างการทดสอบแต่ละข้อ
2) ควรเพิ่มคำชี้แจงให้ผู้สอบทราบว่าในการทดสอบไม่สามารถย้อนกลับมาแก้ไขคำตอบได้	2) เพิ่มรายละเอียดในคำชี้แจง ให้ผู้เข้ารับการทดสอบทราบเรื่องการทดสอบผู้เข้าสอบไม่สามารถย้อนกลับมาแก้ไขคำตอบได้
3) พื้นหลังบางข้อให้สีเข้มทำให้อ่านไม่สบายตา ควรเน้นสีโทนอ่อน	3) ปรับสีพื้นหลังให้อ่อนลง
ควรลดเวลาให้เหมาะสมกับความยาวของโจทย์	4) ปรับแก้ไขโดน ข้อสอบด้าน Conversation ,Grammar and Vocabulary และ Miscellaneous มีระยะเวลาในการสอบ 5 นาที ข้อสอบด้าน Reading Comprehension ระยะเวลาในการสอบ 10 นาที
5) ควรปรับสีรายงานให้สว่างขึ้นเพื่อให้ดูง่าย	5) ปรับสีรายงานผลการทดสอบให้สว่างขึ้น

### ตอนที่ 3 ผลการประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ผลการศึกษาความพึงพอใจของผู้ใช้ที่มีต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดโดยผู้ใช้งาน ด้วยแบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีรูปแบบเป็นมาตราประมาณค่า 5 ระดับ ประกอบด้วยเนื้อหาหลักในการประเมิน 4 องค์ประกอบ ได้แก่ 1) ภาพรวมของระบบ 2) หน้าจอของระบบ 3) การเรียนรู้ และ 4) ความสามารถของระบบ ซึ่งประเมินโดยนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนพิชัยรัตนาคาร จำนวน 74 คนที่เป็นผู้ใช้งานระบบ พบว่า โดยภาพรวมนักเรียนมีความพึงพอใจระดับมากต่อระบบการทดสอบ ( $M=3.91$ ,  $SD= 0.50$ ) หากพิจารณาความพึงพอใจในแต่ละข้อพบว่า การทำงานของระบบสะดวก ง่าย ต่อการใช้งาน และไม่ซับซ้อน ( $M=4.03$ ,  $SD=0.72$ ) ผู้ใช้งานสามารถเรียนรู้การใช้งานและการทำงานของระบบด้วยตนเอง ( $M=4.12$ ,  $SD= 0.68$ ) กระบวนการทำงานของระบบมีความรวดเร็วในการเรียกใช้บริการ ( $M=3.82$ ,  $SD=0.69$ ) รูปแบบของข้อความ ส่วนประกอบต่าง ๆ บนหน้าจอมีความเหมาะสม ( $M=3.85$ ,  $SD=0.68$ ) ลำดับการแสดงผลและส่วนประกอบต่างๆ บนหน้าจอเป็นลำดับขั้นตอน ( $M=3.80$ ,  $SD=0.64$ ) การออกแบบการใช้สีที่เหมาะสมต่อการใช้งาน ( $M=3.78$ ,  $SD=0.75$ ) ความรวดเร็วในการรายงานผลการทดสอบ ( $M=3.97$ ,  $SD=0.74$ ) การรายงานผลการทดสอบมีความชัดเจน เข้าใจง่าย ( $M=3.89$ ,  $SD=0.71$ ) ผลการทดสอบนำไปใช้ในการเตรียมความพร้อมก่อนการทดสอบภาษาอังกฤษได้ ( $M=3.92$ ,  $SD=0.72$ ) และความรู้สึกโดยรวมที่มีต่อระบบ ( $M=3.91$ ,  $SD=0.75$ ) มีคุณภาพอยู่ในระดับมากทั้ง 10 ข้อ ซึ่งมีผลการวิเคราะห์ดังตาราง 4.16

ตาราง 4.16 ผลการศึกษาความพึงพอใจของผู้ใช้ที่มีต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

คำถาม	ระดับความพึงพอใจ					M	SD
	น้อยที่สุด n (%)	น้อย n (%)	ปานกลาง n (%)	มาก n (%)	มากที่สุด n (%)		
1. การทำงานของระบบสะดวกง่ายต่อการใช้งาน และไม่ซับซ้อน	-	-	18 (24.32)	36 (48.65)	20 (27.03)	4.03	0.72
2. ผู้ใช้งานสามารถเรียนรู้การใช้งานและการทำงานของระบบด้วยตนเอง	-	-	13 (17.57)	39 (52.70)	22 (29.73)	4.12	0.68
3. กระบวนการทำงานของระบบมีความรวดเร็วในการเรียกใช้บริการ	-	-	25 (33.78)	37 (50.00)	12 (16.22)	3.82	0.69
4. รูปแบบของข้อความส่วนประกอบต่าง ๆ บนหน้าจอมีความเหมาะสม	-	-	23 (31.08)	39 (52.70)	12 (16.22)	3.85	0.68
5. ลำดับการแสดงผลและส่วนประกอบต่างๆ บนหน้าจอเป็นลำดับขั้นตอน	-	-	24 (32.43)	41 (55.41)	9 (12.16)	3.80	0.64
6. การออกแบบการใช้สีที่เหมาะสมต่อการใช้งาน	-	3 (4.05)	21 (28.38)	39 (52.70)	11 (14.86)	3.78	0.75
7. ความรวดเร็วในการรายงานผลการทดสอบ	-	-	21 (28.38)	34 (45.95)	19 (25.68)	3.97	0.74
8. การรายงานผลการทดสอบมีความชัดเจนเข้าใจง่าย	-	-	23 (31.08)	36 (48.65)	15 (20.27)	3.89	0.71
9. ผลการทดสอบนำไปใช้ในการเตรียมความพร้อมก่อนการทดสอบภาษาอังกฤษได้	-	-	22 (29.73)	36 (48.65)	16 (21.62)	3.92	0.72
10. ความรู้สึกโดยรวมที่มีต่อระบบ	-	-	24 (32.43)	32 (43.24)	18 (24.32)	3.92	0.75

## บทที่ 5

### สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยเรื่อง การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า มีวัตถุประสงค์การวิจัย ได้แก่ เพื่อพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้วิธีดัชนีลำดับความสำคัญสูงสุด และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า และเพื่อประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยการประเมินแบบอิงมาตรฐาน การประเมินอิวิริสติก และการประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบ ในการดำเนินการวิจัยแบ่งการดำเนินงานออกเป็น 2 ระยะ ได้แก่ ระยะที่ 1 การพัฒนาระบบการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด และ ระยะที่ 2 การประเมินคุณภาพระบบการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีรายละเอียดดังต่อไปนี้

ระยะที่ 1 การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด แบ่งการดำเนินงานออกเป็น 3 ตอน คือ ตอนที่ 1 การพัฒนาล้างข้อสอบ โดยการรวบรวมข้อสอบแบบฝึกพัฒนาทักษะภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่เผยแพร่ในเว็บไซต์ของสถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ตั้งแต่ปี พ.ศ. 2556 – 2559 จำนวน 640 ข้อ ที่มีรูปแบบของคำตอบเป็นแบบ 4 ตัวเลือก โดยมีการมุ่งวัดความรู้ทางภาษาอังกฤษ ซึ่งประกอบด้วย 4 ด้าน ได้แก่ Conversation, Grammar and Vocabulary, Reading Comprehension และ Miscellaneous โดยทดสอบกับนักเรียนกลุ่มตัวอย่างซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา จำนวน 1,109 คน ตรวจสอบและเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบ เพื่อยืนยันคุณภาพของข้อสอบ มีเกณฑ์การพิจารณาจากค่าสถิติ  $G^2$  (deviance statistic) ค่า AIC (Akaike Information Criterion) ค่า BIC (Bayesian information criterion) และตรวจสอบคุณภาพของข้อสอบรายข้อตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ได้แก่ ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -4 ถึง 4, ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง  $\infty$  และค่าการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.50 ดำเนินการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ ด้วยโปรแกรม R แพคเกจ ltm ตอนที่ 2 การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด แบ่งออกเป็น 7 ส่วน ได้แก่ 1) รูปแบบของระบบทดสอบ เป็นรูปแบบออนไลน์ มีการกำหนดสิทธิ์ในการเข้าถึงในการใช้งานของระบบ โดยแบ่งออกเป็น ผู้จัดการระบบ (admin) และผู้ใช้งาน (user) และจัดทำคู่มือการใช้งานเป็นแนวทางในการนำโปรแกรมการทดสอบปรับเหมาะด้วย

คอมพิวเตอร์แบบไฮบริดแบ่งออกเป็น 2 ส่วน คือ คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป (user) และ คู่มือการใช้งานสำหรับผู้จัดการระบบ (admin) 2) การทดสอบ สำหรับหน้าจอการทดสอบจะแสดงข้อสอบพร้อมตัวเลือกสำหรับการทำทดสอบแต่ละข้อ ระยะเวลาสำหรับการทำแบบทดสอบและด้านเนื้อหาของข้อสอบ โดยมีส่วนสำคัญ 5 ส่วน คือ (1) จุดเริ่มต้นการทดสอบ (2) การประมาณค่าความสามารถของผู้สอบ (3) การคัดเลือกข้อสอบ (4) การควบคุมการใช้ข้อสอบซ้ำ และ (5) การยุติการทดสอบ และ 3) การรายงานผลการทดสอบ และตอนที่ 3 การทดลองใช้การทดสอบปรับแก้ด้วยคอมพิวเตอร์แบบไฮบริด กับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนพิชัยรัตนาคาร จังหวัดระนอง สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 14 จำนวน 74 คน

ระยะที่ 2 การประเมินคุณภาพระบบการทดสอบปรับแก้ด้วยคอมพิวเตอร์แบบไฮบริดแบ่งออกเป็น 3 ตอน ได้แก่ ตอนที่ 1 การประเมินแบบอิงมาตรฐาน (standard evaluation) ด้วยแบบประเมินแบบอิงมาตรฐาน ประกอบด้วยเนื้อหา 4 หมวด ได้แก่ ความมีประโยชน์ (utility) ความเป็นไปได้ (feasibility) ความเหมาะสม (propriety) และความถูกต้อง (accuracy) ประเมินโดยผู้เชี่ยวชาญด้านเทคโนโลยี จำนวน 5 ท่าน ตอนที่ 2 การประเมินระบบแบบฮิวริสติก (heuristic evaluation) ด้วยแบบประเมินฮิวริสติก ประกอบด้วยเนื้อหาหลักในการประเมินทั้งสิ้น 10 หมวด ได้แก่ ความชัดเจนของระบบ ความสอดคล้องของระบบและการใช้งานจริง การควบคุมและอิสระของผู้ใช้งาน ความสอดคล้องและมาตรฐาน การป้องกันความผิดพลาด การจำข้อมูลมากกว่าการเรียกคืนข้อมูล ความยืดหยุ่นและประสิทธิภาพการใช้งาน การออกแบบที่สวยงาม การช่วยเหลือโดยการกู้คืนระบบจากข้อผิดพลาด และการช่วยเหลือและคู่มือ ประเมินโดยผู้เชี่ยวชาญด้านเทคโนโลยี จำนวน 5 ท่าน ตอนที่ 3 การประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับแก้ด้วยคอมพิวเตอร์แบบไฮบริด ด้วยแบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับแก้ด้วยคอมพิวเตอร์แบบไฮบริด ประกอบด้วยเนื้อหาหลักในการประเมิน 4 องค์ประกอบ ได้แก่ 1) ภาพรวมของระบบ 2) หน้าจอของระบบ 3) การเรียนรู้ และ 4) ความสามารถของระบบ ซึ่งประเมินโดยนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนพิชัยรัตนาคาร จำนวน 74 คนที่เป็นผู้ใช้งานระบบ

### สรุปผลการวิจัย

ผลการวิจัยสรุปเป็น 2 ประเด็นตามระยะการดำเนินการวิจัยได้แก่ 1) ผลการพัฒนา ระบบการพัฒนาระบบการทดสอบปรับแก้ด้วยคอมพิวเตอร์แบบไฮบริด และ 2) ผลการประเมินคุณภาพระบบการพัฒนาระบบการทดสอบปรับแก้ด้วยคอมพิวเตอร์แบบไฮบริด แต่ละหัวข้อมีรายละเอียดดังนี้

### 1. ผลการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ผลการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด แบ่งออกเป็น 3 ตอน คือ 1) การพัฒนาค้างข้อสอบ 2) การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด และ 3) การทดลองใช้การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีรายละเอียดดังนี้

1) การพัฒนาค้างข้อสอบแบบฝึกพัฒนาทักษะภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 640 ข้อ โดยมีการมุ่งวัดความรู้ทางภาษาอังกฤษ ซึ่งประกอบด้วย 4 ด้าน ได้แก่ Conversation, Grammar and Vocabulary, Reading Comprehension และ Miscellaneous มีการทดสอบกับนักเรียนกลุ่มตัวอย่างซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา จำนวน 1,109 คน การตรวจสอบและเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบ พบว่า โมเดล 3 พารามิเตอร์ (3PL) สอดคล้องกลมกลืนกับข้อสอบมากที่สุด และผลการวิเคราะห์คุณภาพของข้อสอบรายข้อตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ มีข้อสอบที่ผ่านเกณฑ์ในการคัดเลือกข้อสอบจำนวน 476 ข้อ โดยมีข้อสอบ ด้าน Conversation ที่ผ่านเกณฑ์ จำนวน 107 ข้อ ข้อสอบด้าน Grammar จำนวน 133 ข้อ ข้อสอบด้าน Reading จำนวน 149 ข้อ และ ข้อสอบด้าน Miscellaneous จำนวน 87 ข้อ การจัดคลังข้อสอบคัดเลือกข้อสอบที่ผ่านเกณฑ์มาจัดเข้าคลังการจัดชุดข้อสอบ จะมีการชุดข้อสอบไว้ล่วงหน้า 1 ชุด ที่มีความยากปานกลาง (ระดับปานกลาง) และจัดชุดข้อสอบ จำนวน 3 ชุด แบ่งออก ชุดข้อสอบระดับง่าย ชุดข้อสอบระดับปานกลาง และชุดข้อสอบระดับยาก และข้อสอบรายข้อที่มีการแบ่งเป็น 3 ระดับ ได้แก่ ระดับง่าย ระดับปานกลาง และระดับยาก

2) ผลการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด การออกแบบและพัฒนาระบบการทดสอบ แบ่งออกเป็น 3 ส่วนดังนี้ 1) รูปแบบของระบบทดสอบ เป็นการทดสอบออนไลน์ มีการออกแบบให้ผู้ใช้งานมีการเข้าถึงข้อมูลที่สะดวก รวดเร็วและทันสมัย มีรายละเอียดของการทดสอบ มีส่วนให้ผู้ใช้งานใส่รหัสผู้เข้าสอบและรหัสผ่าน และจัดทำคู่มือการใช้งานเป็นแนวทางในการนำโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด 2) การทดสอบ สำหรับหน้าจอ การทดสอบจะแสดงข้อสอบพร้อมตัวเลือกสำหรับการทดสอบแต่ละข้อและระยะเวลาสำหรับการทำแบบทดสอบ โดยมีส่วนสำคัญ 5 ส่วน คือ (1) จุดเริ่มต้นการทดสอบ โดยการกำหนดชุดข้อสอบ ที่มีความยากปานกลาง (2) การประมาณค่าความสามารถผู้สอบด้วยวิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (maximum likelihood estimation: MLE) (3) การคัดเลือกข้อสอบด้วยวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum priority index : MPI) (4) การควบคุมการใช้ข้อสอบซ้ำด้วยวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า

(Progressive-restricted standard error : PR-SE) มีอัตราการใช้ข้อสอบซ้ำสูงสุด เท่ากับ 20 เปอร์เซนต์ และ (5) การยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ(SEE) น้อยกว่าหรือเท่ากับ 0.3 และ 3) การรายงานผลการทดสอบ สำหรับ การรายงานการทดสอบประกอบด้วยข้อมูลส่วนตัวของผู้เข้ารับการทดสอบ ได้แก่ ชื่อ - นามสกุล อีเมล (E-mail) และโรงเรียน รายงานการทดสอบแบ่งออกเป็น 2 ส่วน คือ 1) รายงานผลการตอบ ข้อสอบ ร้อยละ ค่ามาตรฐานที่ (T-score) และ 2) กราฟแสดงความสัมพันธ์ของผลการทดสอบในแต่ละด้าน

3) ผลการทดลองใช้ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เป็น นักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนพิชัยรัตนาคาร จังหวัดระนอง จำนวน 74 คน ประกอบด้วย เพศชาย จำนวน 40 คน คิดเป็น 54.05 และเพศหญิง 34 คน คิดเป็นร้อยละ 45.95 นักเรียนผู้ทดลอง ใช้มีความสามารถด้านทักษะภาษาอังกฤษ เฉลี่ยเท่ากับ - 0.17 มีการแจกแจงแบบเบ้ซ้าย และมีความโด่งมากกว่าโค้งปกติ

## 2. ผลการประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

ผลการประเมินคุณภาพระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด แบ่ง ออกเป็น 3 ตอน ได้แก่ 1) ผลการประเมินแบบอิงมาตรฐาน (standard evaluation) 2) การประเมิน ระบบแบบฮิวริสติก (heuristic evaluation) และ 3) การประเมินความพึงพอใจของผู้ใช้งานต่อระบบ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยมีรายละเอียดดังนี้

1) ผลการประเมินแบบอิงมาตรฐาน สำหรับแบบประเมินอิงมาตรฐาน ประกอบด้วยข้อ คำถาม จำนวน 14 ข้อ ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินอิงมาตรฐาน โดย ผู้เชี่ยวชาญด้านการศึกษาคือ ผ่านเกณฑ์การประเมิน 14 ข้อ โดยมีค่าดัชนี IOC อยู่ระหว่าง 0.6 – 1.00 ทั้งนี้ผู้เชี่ยวชาญได้แนะนำให้ดำเนินการปรับแก้แบบประเมินและผู้วิจัยได้ดำเนินการ ปรับแก้ก่อนนำไปใช้งานจริง และผลการประเมินแบบอิงมาตรฐานโดยผู้เชี่ยวชาญด้านเทคโนโลยี พบว่า ภาพรวมระบบการทดสอบมีคุณภาพอยู่ในระดับมากที่สุด ( $M = 4.59$ ,  $SD = 0.35$ ) หาก พิจารณาคุณภาพของระบบการทดสอบในแต่ละด้าน พบว่า ด้านความมีประโยชน์ ( $M = 4.53$ ,  $SD = 0.56$ ) ด้านความเป็นไปได้ ( $M = 4.65$ ,  $SD = 0.42$ ) ด้านความเหมาะสม ( $M = 4.60$ ,  $SD = 0.22$ ) และ ด้านความถูกต้อง ( $M = 4.53$ ,  $SD = 0.65$ ) มีคุณภาพอยู่ในระดับมากที่สุดทั้ง 4 ด้าน

2) ผลการประเมินระบบแบบฮิวริสติก โดยแบบประเมินฮิวริสติก ประกอบด้วยข้อ คำถาม จำนวน 13 ข้อ ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินอิงมาตรฐาน โดย ผู้เชี่ยวชาญด้านการศึกษาคือ ผ่านเกณฑ์การประเมิน 13 ข้อ โดยมีค่าดัชนี IOC อยู่ระหว่าง 0.6 – 1.00 ทั้งนี้ผู้เชี่ยวชาญได้แนะนำให้ดำเนินการปรับแก้แบบประเมินและผู้วิจัยได้ดำเนินการ ปรับแก้ก่อนนำไปใช้งานจริง และผลการประเมินระบบแบบฮิวริสติกโดยผู้เชี่ยวชาญด้านเทคโนโลยี



ประเมิน พบว่า ภาพรวมระบบการทดสอบมีคุณภาพอยู่ในระดับมาก ( $M= 4.42$  ,  $SD= 0.33$ ) ด้านระบบมีความยืดหยุ่นและประสิทธิภาพของการทำงานมีค่าเฉลี่ยมากที่สุด ( $M=5.00$ ,  $SD=0.00$ ) และระบบมีความสามารถในการจดจำข้อมูลมากกว่าการเรียกข้อมูลมีค่าเฉลี่ยน้อยที่สุด ( $M=4.20$ ,  $SD=0.45$ )

3) ผลการศึกษาความพึงพอใจของผู้ใช้ที่มีต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด สำหรับแบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ประกอบด้วยข้อคำถาม จำนวน 10 ข้อ ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินอิงมาตรฐาน โดยผู้เชี่ยวชาญด้านการศึกษาพบว่า ผ่านเกณฑ์การประเมิน 10 ข้อ โดยมีค่าดัชนี IOC อยู่ระหว่าง 0.8 – 1.00 ทั้งนี้ผู้เชี่ยวชาญได้แนะนำให้ดำเนินการปรับแก้แบบประเมินและผู้วิจัยได้ดำเนินการปรับแก้ก่อนนำไปใช้งานจริง และผลการศึกษาความพึงพอใจของผู้ใช้ที่มีต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดโดยนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนพิชัยรัตนาคาร จำนวน 74 คนที่เป็นผู้ใช้งานระบบ พบว่า ภาพรวมนักเรียนมีความพึงพอใจระดับมากต่อระบบการทดสอบ ( $M=3.91$ ,  $SD= 0.50$ ) ประเด็นผู้ใช้งานสามารถเรียนรู้การใช้งานและการทำงานของระบบด้วยตนเองมีค่าเฉลี่ยมากที่สุด ( $M=4.12$ ,  $SD= 0.68$ ) และ การออกแบบการใช้สื่ที่เหมาะสมต่อการใช้งานมีค่าเฉลี่ยน้อยที่สุด ( $M=3.78$  ,  $SD=0.75$ )

### อภิปรายผลการวิจัย

จากผลการวิจัยข้างต้น นำมาซึ่งการอภิปรายผลการวิจัยได้ ประเด็น คือ 1) ผลการพัฒนา ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด และ 2) ผลการประเมินคุณภาพระบบ การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยแต่ละประเด็นมีรายละเอียด ดังนี้

#### 1. ผลการพัฒนาระบบการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

1.1 การพัฒนาคำข้อสอบโดยการรวบรวมแบบฝึกพัฒนาทักษะภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ของสถาบันภาษาอังกฤษ สำนักงานการศึกษาขั้นพื้นฐาน ตั้งแต่ปี พ.ศ. 2556 – 2559 จำนวน 640 ข้อ ที่มีรูปแบบของคำตอบเป็นแบบ 4 ตัวเลือก โดยมีการมุ่งวัดความรู้ทางภาษาอังกฤษ ซึ่งประกอบด้วย 4 ด้าน ได้แก่ Conversation, Grammar and Vocabulary, Reading Comprehension และ Miscellaneous ซึ่งข้อสอบทั้งหมดครอบคลุมตามตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 และสอดคล้องกับคุณภาพของผู้เรียนระดับชั้นมัธยมศึกษาชั้นปีที่ 3 ที่ระบุไว้ว่า ผู้เรียนจะต้องปฏิบัติตามคำขอร้อง คำแนะนำ คำชี้แจง และคำอธิบายที่ฟังและอ่าน

อ่านออกเสียงข้อความ ขำ และโฆษณา สั้นๆ ถูกต้องตามหลักการอ่าน ระบุ/เขียนสื่อที่ไม่ใช่ความเรียงรูปแบบต่างๆ สัมพันธ์กับประโยคและข้อความที่ฟังหรืออ่าน เลือก/ระบุหัวข้อเรื่อง ใจความสำคัญ รายละเอียดสนับสนุน และแสดงความคิดเห็นเกี่ยวกับเรื่องที่ฟังและอ่านจากสื่อประเภทต่างๆ พร้อมทั้งให้เหตุผลและยกตัวอย่างประกอบ และสนทนาและเขียนโต้ตอบข้อมูลเกี่ยวกับตนเองและเรื่องต่างๆ ใกล้ตัว สถานการณ์ ขำ เรื่องที่อยู่ในความสนใจของสังคมและสื่อสารอย่างต่อเนื่องและเหมาะสม ใช้คำขอร้อง คำชี้แจง และคำอธิบาย ให้คำแนะนำอย่างเหมาะสม พูดและเขียนแสดงความต้องการ เสนอและให้ความช่วยเหลือ ตอบรับและปฏิเสธ การให้ความช่วยเหลือ พูดและเขียนเพื่อขอและให้ข้อมูลบรรยาย อธิบาย เปรียบเทียบ และแสดงความคิดเห็นเกี่ยวกับเรื่องที่ฟังหรืออ่านอย่างเหมาะสม พูดและเขียนบรรยายความรู้สึกและความคิดเห็นของตนเองเกี่ยวกับเรื่องต่างๆ กิจกรรม ประสบการณ์ และข่าว/เหตุการณ์ พร้อมทั้งให้เหตุผลประกอบอย่างเหมาะสม (กระทรวงศึกษาธิการ, 2552)

1.2 การตรวจสอบคุณภาพข้อสอบเพื่อประมาณค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ ได้แก่ ค่าความยากของข้อสอบ (b) ค่าอำนาจจำแนกของข้อสอบ (a) และค่าการเดาของข้อสอบ (c) โดยการนำแบบฝึกทักษะภาษาอังกฤษ ไปทดสอบกับนักเรียนกลุ่มตัวอย่างซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา (สพม.) จำนวน 1,109 คน ซึ่งขนาดกลุ่มตัวอย่างที่เหมาะสมและเพียงพอต่อการวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ (Foley, 2010) และผู้วิจัยนำผลจากการทดสอบกับกลุ่มตัวอย่างตรวจสอบและเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการตอบสนองข้อสอบ ผลการวิจัยแสดงให้เห็นว่าโมเดล 3 พารามิเตอร์ (3PL) สอดคล้องกลมกลืนกับข้อสอบมากที่สุด โมเดล 3 พารามิเตอร์ (3PL) มีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า และยังเป็นตัวบ่งชี้ถึงโมเดล 3 พารามิเตอร์ (3PL) ที่ค่าตัววัดที่สุดจากการพิจารณาจากความแตกต่างระหว่างโมเดล (เมฆา นวลศรี, 2559; ชัยวิจิต เจียรชนะ, 2552)

และผลการวิเคราะห์คุณภาพของข้อสอบรายข้อตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์เพื่อคัดเลือกข้อสอบเข้าสู่คลังข้อสอบ จากเดิมผู้วิจัยได้กำหนดเกณฑ์ในการประมาณค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ คือ ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -3 ถึง 3 ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง 2.00 ค่าการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.30 ทำให้ได้ข้อสอบที่ผ่านเกณฑ์จำนวน 320 ข้อ จึงทำให้ต้องขยายเกณฑ์เกณฑ์ในการคัดเลือกข้อสอบตามค่าพารามิเตอร์ของข้อสอบ ได้แก่ 1) ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง  $\infty$  2) ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -4 ถึง 4 และ 3) ค่าการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.50 (Reckase, 2003) จาก

การวิเคราะห์ข้อสอบผ่านเกณฑ์จำนวน 476 ข้อ ดังนั้นในคลังข้อสอบของงานวิจัยครั้งนี้มีจำนวน 476 ข้อ ซึ่งมากกว่าเกณฑ์ขั้นต่ำ 2.4 เท่า และข้อสอบที่ผ่านเกณฑ์ในการคัดเลือก จำนวน 476 ข้อ พบว่าข้อสอบด้าน Grammar เป็นด้านที่ข้อสอบไม่ผ่านการคัดเลือกมากที่สุด จำนวน 73 ข้อ (คิดเป็นร้อยละ 35.44) อาจจะเนื่องมาจากการตรวจสอบคุณภาพข้อสอบที่ผ่านการทดลองโดยกลุ่มตัวอย่างมีระดับความรู้พื้นฐาน ด้าน Grammar หรือไวยากรณ์ภาษาอังกฤษอยู่ในระดับน้อย ซึ่งสอดคล้องกับงานวิจัยของ กาโสม หมดเต็น และ นิสากร จารุมณี (2560) พบว่า นักเรียนชั้นมัธยมศึกษาปีที่ 3 ในสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 16 (จังหวัดสงขลาและสตูล) มีความสามารถด้านไวยากรณ์ภาษาอังกฤษในระดับน้อย และสอดคล้องกับผลการวิจัยของ ชญาพร วจนะคมگیر (2554) ได้ศึกษาระดับความรู้พื้นฐานภาษาอังกฤษของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 โรงเรียนปรีณทสรอยแยลวิทยาลัย จังหวัดเชียงใหม่ พบว่า นักเรียนส่วนใหญ่มีความรู้พื้นฐานไวยากรณ์ภาษาอังกฤษต่ำกว่าเกณฑ์มาตรฐาน เมื่อผู้เรียนมีพื้นฐานไวยากรณ์ต่ำ ทำให้ความสามารถในการเข้าใจภาษาอังกฤษน้อยลง เช่น นักเรียนไม่สามารถอ่านสื่อการเรียนการสอนที่มีคำอธิบายเป็นภาษาอังกฤษได้ เนื่องจากไม่เข้าใจความหมาย เป็นต้น

1.3 การออกแบบและพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้หลักการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เริ่มต้นการทดสอบด้วยการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ด้วยชุดข้อสอบที่กำหนดขึ้น (MST ขั้นที่ 1) ที่มีความยากระดับปานกลาง ระบบจะทำการวิเคราะห์ความสามารถของผู้สอบ ( $\theta$ ) ใช้วิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (MLE) เพื่อให้ค่าความสามารถที่ได้มีความถูกต้องและแม่นยำ (Hulin, Drasgow and Parson, 1983 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) จากนั้นระบบจะคัดเลือกชุดข้อสอบถัดไปที่มีความยากใกล้เคียงกับความสามารถของผู้สอบ (MST ขั้นที่ 2) โมดูลข้อสอบออกเป็น 3 ระดับ ได้แก่ ระดับง่าย ระดับปานกลาง และระดับยาก การออกแบบชุดข้อสอบและข้อสอบ ออกเป็น 3 ระดับ คือ ระดับง่าย ปานกลาง และระดับยากสอดคล้องกับ Lord (1980) ที่ได้เสนอแนะว่าควรใช้ 2 – 4 ระดับในแต่ละขั้นตอนก็เพียงพอ การออกแบบระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยมีขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จำนวน 2 ขั้นตอน (two-stage) ซึ่งเป็นไปตามขั้นตอนพื้นฐานของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) (Sari, Sari, & Huggins Manley, 2016)

การคัดเลือกข้อสอบด้วยวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) ซึ่งเป็นวิธีการที่ใช้ค่าฟังก์ชันสารสนเทศสูงสุด ทำให้การคัดเลือกข้อสอบมีความแม่นยำมากขึ้น (Y. Cheng & Chang, 2009) และควบคุมการใช้ข้อสอบซ้ำด้วยวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-restricted standard error: PR-SE) เป็นวิธีการที่ควบคุม

การการใช้ข้อสอบซ้ำสำหรับคลังข้อสอบขนาดใหญ่ และทำให้มีจำนวนการซ้อนทับข้อสอบกันน้อยที่สุด (Leroux et al., 2013) และมีการควบคุมอัตราการใช้ข้อสอบซ้ำสูงสุดเท่ากับ 20 เปอร์เซ็นต์ เมื่อผู้สอบทำชุดข้อสอบตามระดับความสามารถแล้ว ระบบจะคัดเลือกข้อสอบถัดไปเป็นทีละข้อ ซึ่งเป็นส่วนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ จนการทดสอบมีความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) น้อยกว่าหรือเท่ากับ 0.3 (ริงสรณ์ มณีเล็ก, 2540) แล้วยุติการทดสอบ และการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดในครั้งนี้ เป็นการพัฒนาระบบการทดสอบแบบออนไลน์ (online) โดยการถอดขั้นตอนการดำเนินงาน (algorithm) ต่าง ๆ ของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดเพื่อนำขั้นตอนดังกล่าวไปใช้ในการเขียนโปรแกรมให้รองรับระบบการใช้งานออนไลน์ซึ่งเขียนด้วยภาษา PHP

## 2. ผลการประเมินคุณภาพระบบการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

การตรวจสอบคุณภาพของระบบการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ได้ประเมินคุณภาพระบบ 3 แบบ คือ การประเมินแบบอิงมาตรฐาน (standard evaluation) การประเมินระบบแบบฮิวริสติก (heuristic evaluation) และการประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยการประเมินแบบอิงมาตรฐาน (standard evaluation) การประเมินระบบแบบฮิวริสติก (heuristic evaluation) ใช้สำหรับการประเมินคุณภาพก่อนนำระบบไปใช้งานจริง ซึ่งการประเมินแบบอิงมาตรฐาน (standard evaluation) ตามแนวคิดของ Stufflebeam และคณะ ประกอบด้วยเนื้อหาหลักในการประเมินทั้งสิ้น 4 หมวด ได้แก่ ความมีประโยชน์ (utility) ความเป็นไปได้ (feasibility) ความเหมาะสม (propriety) และความถูกต้อง (accuracy) (ศิริชัย กาญจนวาสี, 2554) และการประเมินแบบฮิวริสติก (heuristic evaluation) ประกอบด้วยเนื้อหาหลักในการประเมินทั้งสิ้น 10 หมวด ได้แก่ ความชัดเจนของระบบ ความสอดคล้องของระบบและการใช้งานจริง การควบคุมและอิสระของผู้ใช้งาน ความสอดคล้องและมาตรฐาน การป้องกันความผิดพลาดการจำข้อมูลมากกว่าการเรียกคืนข้อมูล ความยืดหยุ่นและประสิทธิภาพการใช้งาน การออกแบบที่สวยงาม การช่วยเหลือโดยการกู้คืนระบบจากข้อผิดพลาด และการช่วยเหลือและคู่มือ ซึ่งสอดคล้องกับแนวทางของณภัทร ชัยมงคล (2558) ที่มีการตรวจสอบคุณภาพของระบบการทดสอบปรับเหมาะหลายขั้นตอนด้วยวิธีออนเดอะฟลายที่มีการสะท้อนข้อมูลย้อนกลับในการทดสอบมาตรฐานวิชาชีพของบุคลากรสาขาไอที และโซติกา ภาชีผล (2561) ซึ่งมีการประเมินระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่

21 ขอนิสิตนักศึกษาปริญญาตรี โดยดำเนินการประเมินรูปแบบที่พัฒนาขึ้นตามแนวคิดของการประเมินแบบอิงมาตรฐาน

สำหรับการประเมินความพึงพอใจของผู้ใช้ที่มีต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ประกอบด้วยเนื้อหาหลักในการประเมิน 4 องค์ประกอบ ได้แก่ 1) ภาพรวมของระบบ 2) หน้าจอของระบบ 3) การเรียนรู้ และ 4) ความสามารถของระบบ ซึ่งประเมินจากผู้เข้ามาใช้ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เพื่อต้องการทราบว่าระบบการทดสอบที่พัฒนาขึ้นสามารถตอบสนองการใช้งานของผู้ทดลองใช้ระบบได้หรือไม่ เนื่องจากระบบที่ได้รับการพัฒนาขึ้นเป็นระบบที่ต้องใช้งานผ่านระบบออนไลน์ ดังนั้นจึงต้องประเมินผ่านระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ซึ่งสอดคล้องกับแนวทางของ ญานิศรา มุนินทร์สาคร (2558) ที่มีการประเมินความคิดเห็นของผู้ทดลองใช้โปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการทดสอบ O-NET ระดับชั้นประถมศึกษาปีที่ 6 โดยการประเมินผ่านทางเว็บไซต์ (website)

## ข้อเสนอแนะ

### ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. สถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ซึ่งเป็นผู้ดูแลแบบฝึกทักษะภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 สามารถนำผลการตรวจสอบคุณภาพข้อสอบในการวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ ได้แก่ ค่าความยากของข้อสอบ (b) ค่าอำนาจจำแนกของข้อสอบ (a) และค่าการเดาของข้อสอบ (c) มาใช้ในการคัดเลือกข้อสอบเข้าสู่คลังข้อสอบเพื่อใช้ในการจัดทำคลังข้อสอบของแบบฝึกทักษะภาษาอังกฤษ

2. ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด รายวิชาภาษาอังกฤษ ที่พัฒนาขึ้นในการวิจัยครั้งนี้เหมาะสำหรับนำไปประยุกต์ใช้กับนักเรียนที่กำลังศึกษาในชั้นมัธยมศึกษาตอนต้น ที่มีความสนใจจะพัฒนาภาษาอังกฤษสามารถเข้ามาใช้งานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด เพื่อประเมินศักยภาพทางภาษาอังกฤษของตนเองว่าอยู่ในระดับใดควรพัฒนาศักยภาพทางภาษาอังกฤษในด้านใดบ้าง และเป็นเครื่องมือชี้แนะแนวทางในการพัฒนาทักษะภาษาอังกฤษที่มีความเหมาะสมกับตนเอง และสถานศึกษาสามารถนำไปเครื่องมือในการประเมินความรู้ความสามารถทางด้านทักษะภาษาอังกฤษ ผลการทดสอบจะช่วยช่วยให้สถานศึกษานำไปพัฒนาผู้เรียนและจัดการเรียนการสอนให้สอดคล้องกับการตัวชี้วัดตามตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

3. ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด มีการประเมินคุณภาพระดับมากขึ้นไปในทุกด้าน และนักเรียนที่ได้ทดลองใช้ระบบมีความพึงพอใจต่อระบบการทดสอบทุกด้าน จึงสามารถนำระบบไปใช้จริงได้ ดังนั้นอาจารย์ในสถานศึกษาที่ต้องการพัฒนาศักยภาพทางด้านภาษาอังกฤษ สามารถนำระบบที่พัฒนาไปใช้หรือเป็นแนวทางในการพัฒนาต่อยอดโดยศึกษาจากการใช้คู่มือของระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

4. ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ที่พัฒนาขึ้นเป็นแนวทางสำหรับผู้สนใจพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์สามารถนำไปใช้ในการพัฒนาการทดสอบประเภทอื่นๆ เช่น การพัฒนาทักษะในกลุ่มสาระการเรียนรู้อื่น ๆ ตามตัวชี้วัดและสาระการเรียนรู้แกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 หรือ การทดสอบวัดความรู้ระดับชาติ เพื่อลดปริมาณการใช้กระดาษและสามารถรายงานผลการทดสอบได้ทันที และใช้ในการพัฒนาการจัดทำคลังข้อสอบด้วยการวิเคราะห์ข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบที่มีการตรวจสอบโมเดลที่เหมาะสมเพื่อเพิ่มประสิทธิภาพในการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ

### ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. การวิจัยครั้งนี้เป็นการพัฒนาคลังข้อสอบเพื่อพัฒนาทักษะภาษาอังกฤษ ซึ่งคลังข้อสอบมีขนาดปานกลาง จำนวน 476 ข้อ ดังนั้นในการวิจัยครั้งต่อไปควรพัฒนาคลังข้อสอบให้มีขนาดใหญ่เพื่อทดสอบคุณภาพของระบบในการคัดเลือกข้อสอบว่าจะสามารถคัดเลือกข้อสอบได้อย่างมีประสิทธิภาพและประสิทธิภาพสูงกว่าการทดสอบที่มีขนาดปานกลางหรือไม่

2. การวิจัยครั้งนี้เป็นการพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยใช้แบบฝึกทักษะภาษาอังกฤษ สำหรับนักเรียนชั้นมัธยมศึกษา ปีที่ 3 ที่มุ่งวัดความรู้ภาษาอังกฤษเฉพาะด้าน ซึ่งนักเรียนในระดับชั้นมัธยมศึกษาตอนปลาย หรือในระดับที่สูงขึ้นมีการพัฒนาทักษะภาษาอังกฤษในหลายด้านร่วมกัน ดังนั้นควรมีการศึกษาหรือพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดที่มีการวัดคุณลักษณะแฝงของแต่ละด้านโดยใช้ข้อสอบชุดเดียวกัน

3. ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดในงานวิจัยนี้เป็นการประเมินผลสรุป (summative assessment) เพื่อประเมินศักยภาพทางภาษาอังกฤษของนักเรียนว่าอยู่ในระดับใด ควรมีการพัฒนาโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดในรูปแบบของการประเมินระหว่างเรียน (formative assessment) โดยการออกแบบระบบให้มีการเพิ่มการสะท้อนข้อมูลย้อนกลับให้นักเรียน เพื่อให้นักเรียนสามารถนำผลการประเมินไปใช้ในการพัฒนาปรับปรุงสมรรถนะของตนเองให้ดียิ่งขึ้นต่อไป

4. ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดในครั้งนี้มีการมุ่งวัดความรู้ทางภาษาอังกฤษ ซึ่งประกอบด้วย 4 ด้าน ได้แก่ ด้าน Conversation, ด้าน Grammar and Vocabulary, ด้าน Reading Comprehension และ ด้าน Miscellaneous ไม่มีการวัดเนื้อหาทางด้าน writing ซึ่งเป็นทักษะการสื่อสารของภาษาอังกฤษ ทั้งนี้เป็นเพราะว่าข้อสอบวัดเนื้อหาทางด้าน writing ควรเป็นข้อสอบความเรียงไม่จำกัดคำตอบ แต่ด้วยข้อจำกัดของงานวิจัยที่พัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดต้องมีคำตอบถูกที่แน่นอนและมีการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบไว้ล่วงหน้าในการคัดเลือกข้อสอบให้ผู้สอบทำ จึงไม่สามารถที่จะใช้ข้อสอบความเรียงแบบไม่จำกัดคำตอบเพื่อวัดด้าน writing ได้ ดังนั้นในการวิจัยครั้งต่อไปควรมีการศึกษาเพิ่มเติมโดยใช้ข้อสอบในรูปแบบที่หลากหลายและต้องประยุกต์ใช้ในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ภายใต้ทฤษฎีการตอบสนองข้อสอบที่มีการให้คะแนนหลายค่า (polytomous scoring) เพื่อให้วัดทักษะภาษาอังกฤษได้อย่างครอบคลุมมากขึ้น





จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**



ภาคผนวก ก  
รายนามผู้เชี่ยวชาญ

รายนามผู้เชี่ยวชาญในการตรวจสอบคุณภาพเครื่องมือวิจัย

- |                                                |                                                                                |
|------------------------------------------------|--------------------------------------------------------------------------------|
| 1. ผู้ช่วยศาสตราจารย์ ดร.จิตต์วิมล คล้ายสุบรรณ | อาจารย์ประจำคณะครุศาสตร์<br>มหาวิทยาลัยสวนดุสิต                                |
| 2. ผู้ช่วยศาสตราจารย์ ดร.นภาพร บุญเส็ง         | อาจารย์ประจำคณะศึกษาศาสตร์<br>มหาวิทยาลัยการกีฬาแห่งชาติ<br>วิทยาเขตสุพรรณบุรี |
| 3. อาจารย์ ดร.จิณัฐตา สอนสังข์                 | อาจารย์ประจำคณะครุศาสตร์<br>มหาวิทยาลัยราชภัฏสุราษฎร์ธานี                      |
| 4. อาจารย์ ดร.จุฑาภรณ์ มาสันเทียะ              | อาจารย์ประจำคณะศึกษาศาสตร์<br>มหาวิทยาลัยรามคำแหง                              |
| 5. อาจารย์ ดร.จุฑาทิพย์ สว่างสุวรรณ            | อาจารย์ประจำคณะศึกษาศาสตร์<br>มหาวิทยาลัยการกีฬาแห่งชาติ<br>วิทยาเขตสมุทรสาคร  |

### รายนามผู้เชี่ยวชาญในการตรวจสอบคุณภาพระบบ

- |                                          |                                                                                            |
|------------------------------------------|--------------------------------------------------------------------------------------------|
| 1. ผู้ช่วยศาสตราจารย์ ดร.ดริณภพ เพียรจัด | อาจารย์ประจำคณะครุศาสตร์<br>มหาวิทยาลัยราชภัฏราชชนครินทร์                                  |
| 2. ผู้ช่วยศาสตราจารย์ ดร.ชนิตา ไกรเพชร   | อาจารย์ประจำคณะศึกษาศาสตร์<br>มหาวิทยาลัยการกีฬาแห่งชาติ<br>วิทยาเขตสมุทรสาคร              |
| 3. อาจารย์ ดร.ณภัทร ชัยมงคล              | อาจารย์ประจำคณะครุศาสตร์<br>จุฬาลงกรณ์มหาวิทยาลัย                                          |
| 4. อาจารย์ ดร.อนุสรณ์ เกิดศรี            | อาจารย์ประจำศูนย์วิชาการประเมินผล<br>สำนักทะเบียนและวัดผล<br>มหาวิทยาลัยสุโขทัยธรรมมาธิราช |
| 5. อาจารย์ ดร.ชุตีวัฒน์ สุวดีพิงศ์       | อาจารย์ประจำสำนักเทคโนโลยี<br>การศึกษา<br>มหาวิทยาลัยสุโขทัยธรรมมาธิราช                    |

ภาคผนวก ข

ผลการวิเคราะห์คุณภาพของข้อสอบรายข้อ (IRT) ตามทฤษฎีการตอบสนองข้อสอบ

ตาราง 1 ผลการวิเคราะห์คุณภาพของข้อสอบรายข้อ (IRT) ตามทฤษฎีการตอบสนองข้อสอบ (item response theory)

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
1	7.379	1.718	0.367		ยาก	Conversation
2	6.213	2.904	0.362		ยาก	Conversation
3	8.142	1.692	0.297		ยาก	Conversation
4	8.742	1.974	0.316		ยาก	Conversation
5	8.497	4.602	0.380	คัดออก		Conversation
6	0.019	-1.793	0.091		ง่าย	Conversation
7	7.869	1.875	0.284		ยาก	Conversation
8	9.462	3.153	0.359		ยาก	Conversation
9	11.499	4.834	0.406	คัดออก		Conversation
10	4.306	-0.276	0.309		ปานกลาง	Conversation
11	7.912	-0.273	0.275		ปานกลาง	Conversation
12	8.322	1.715	0.272		ยาก	Conversation
13	10.411	3.488	0.358		ยาก	Conversation
14	-0.490	-1.853	0.001	คัดออก		Conversation
15	-0.023	-1.546	0.099	คัดออก		Conversation
16	7.124	2.216	0.366		ยาก	Conversation
17	3.854	0.492	0.340		ปานกลาง	Conversation
18	9.363	3.257	0.315		ยาก	Conversation
19	7.596	0.880	0.292		ปานกลาง	Conversation
20	9.501	-0.619	0.257		ปานกลาง	Conversation
21	5.086	-0.394	0.244		ปานกลาง	Conversation
22	5.119	0.849	0.224		ปานกลาง	Conversation
23	6.037	1.229	0.321		ปานกลาง	Conversation
24	6.444	1.303	0.326		ปานกลาง	Conversation
25	2.591	-0.118	0.280		ปานกลาง	Conversation
26	7.637	2.712	0.322		ยาก	Conversation
27	8.032	3.284	0.284		ยาก	Conversation
28	6.649	0.916	0.220		ปานกลาง	Conversation

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
29	11.246	4.139	0.281	คัดออก		Conversation
30	0.141	-0.961	0.045		ปานกลาง	Conversation
31	5.205	1.268	0.268		ปานกลาง	Conversation
32	6.992	2.297	0.217		ยาก	Conversation
33	1.413	0.054	0.268		ปานกลาง	Conversation
34	7.969	2.462	0.317		ยาก	Conversation
35	8.229	-1.442	0.232		ง่าย	Conversation
36	-0.028	-1.964	0.144	คัดออก		Conversation
37	6.169	1.607	0.203		ยาก	Conversation
38	7.681	2.918	0.285		ยาก	Conversation
39	3.327	-0.387	0.213		ปานกลาง	Conversation
40	-1.690	-5.422	0.150	คัดออก		Conversation
41	6.332	0.217	0.231		ปานกลาง	Grammar
42	4.804	1.395	0.279		ยาก	Grammar
43	1.140	-2.320	0.000		ง่าย	Grammar
44	6.015	1.324	0.212		ปานกลาง	Grammar
45	3.875	-1.477	0.272		ง่าย	Grammar
46	5.779	1.046	0.307		ปานกลาง	Grammar
47	-0.411	-1.634	0.000	คัดออก		Grammar
48	-1.969	-7.525	0.210	คัดออก		Grammar
49	4.342	1.218	0.210		ปานกลาง	Grammar
50	3.299	-0.183	0.233		ปานกลาง	Grammar
51	2.655	-0.189	0.257		ปานกลาง	Grammar
52	4.143	0.578	0.233		ปานกลาง	Grammar
53	3.452	-0.619	0.270		ปานกลาง	Grammar
54	-2.021	-7.121	0.195	คัดออก		Grammar
55	5.777	0.251	0.273		ปานกลาง	Grammar
56	0.189	-0.874	0.026		ปานกลาง	Grammar
57	4.011	-0.252	0.184		ปานกลาง	Grammar
58	19.707	-2.900	0.227		ง่าย	Grammar
59	-0.517	-2.361	0.001	คัดออก		Grammar
60	-1.573	-7.036	0.185	คัดออก		Grammar
61	6.518	1.057	0.329		ปานกลาง	Grammar

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
62	6.438	1.542	0.329		ยาก	Grammar
63	0.429	-1.867	0.167		ง่าย	Grammar
64	-0.860	-3.761	0.215	คัดออก		Grammar
65	0.037	-1.698	0.178		ง่าย	Grammar
66	3.827	0.851	0.242		ปานกลาง	Grammar
67	6.480	-0.318	0.287		ปานกลาง	Grammar
68	0.500	-2.090	0.000		ง่าย	Grammar
69	5.504	-3.651	0.270		ง่าย	Grammar
70	4.636	0.932	0.333		ปานกลาง	Grammar
71	6.356	1.203	0.390		ปานกลาง	Grammar
72	0.668	-1.526	0.272		ง่าย	Grammar
73	8.338	2.434	0.289		ยาก	Grammar
74	7.186	-1.658	0.486		ง่าย	Grammar
75	7.097	-3.590	0.352		ง่าย	Grammar
76	6.531	0.479	0.319		ปานกลาง	Grammar
77	6.009	1.978	0.280		ยาก	Grammar
78	6.016	0.774	0.146		ปานกลาง	Grammar
79	14.766	-0.951	0.249		ปานกลาง	Grammar
80	-0.291	-1.634	0.003	คัดออก		Grammar
81	2.813	-9.392	0.333	คัดออก		Grammar
82	7.873	0.693	0.277		ปานกลาง	Grammar
83	8.122	0.850	0.285		ปานกลาง	Grammar
84	4.432	0.318	0.186		ปานกลาง	Grammar
85	0.792	-0.975	0.149		ปานกลาง	Grammar
86	4.176	-0.331	0.150		ปานกลาง	Grammar
87	2.527	-1.636	0.209		ง่าย	Grammar
88	5.090	0.643	0.179		ปานกลาง	Grammar
89	11.238	2.737	0.324		ยาก	Grammar
90	3.706	-0.342	0.220		ปานกลาง	Grammar
91	5.303	-0.367	0.242		ปานกลาง	Grammar
92	-1.792	-6.871	0.333	คัดออก		Grammar
93	3.637	-0.713	0.169		ปานกลาง	Grammar
94	0.985	-2.992	0.219		ง่าย	Grammar

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
95	5.287	1.055	0.218		ปานกลาง	Grammar
96	12.080	3.225	0.211		ยาก	Grammar
97	4.534	0.213	0.248		ปานกลาง	Grammar
98	9.838	2.157	0.218		ยาก	Grammar
99	6.964	2.223	0.133		ยาก	Grammar
100	2.958	0.137	0.254		ปานกลาง	Grammar
101	11.976	4.101	0.262	คัดออก		Reading
102	2.196	-0.265	0.150		ปานกลาง	Reading
103	7.549	1.693	0.309		ยาก	Reading
104	2.716	0.118	0.181		ปานกลาง	Reading
105	8.658	2.009	0.244		ยาก	Reading
106	5.215	-0.530	0.303		ปานกลาง	Reading
107	10.057	3.235	0.313		ยาก	Reading
108	7.369	-0.104	0.243		ปานกลาง	Reading
109	3.841	-0.005	0.197		ปานกลาง	Reading
110	7.525	2.261	0.306		ยาก	Reading
111	7.710	0.697	0.329		ปานกลาง	Reading
112	11.069	2.781	0.293		ยาก	Reading
113	10.838	2.920	0.178		ยาก	Reading
114	4.689	0.865	0.272		ปานกลาง	Reading
115	5.616	-0.308	0.296		ปานกลาง	Reading
116	6.480	0.127	0.343		ปานกลาง	Reading
117	5.757	-0.470	0.202		ปานกลาง	Reading
118	7.156	0.508	0.271		ปานกลาง	Reading
119	-5.311	- 15.927	0.280	คัดออก		Reading
120	12.949	2.656	0.274		ยาก	Reading
121	5.423	-0.632	0.247		ปานกลาง	Reading
122	10.609	3.053	0.237		ยาก	Reading
123	12.966	4.009	0.345		ยาก	Reading
124	2.185	-0.306	0.201		ปานกลาง	Reading
125	12.514	2.693	0.250		ยาก	Reading
126	6.383	0.181	0.230		ปานกลาง	Reading

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
127	12.197	2.326	0.281		ยาก	Reading
128	5.142	0.335	0.275		ปานกลาง	Reading
129	7.131	0.897	0.286		ปานกลาง	Reading
130	11.440	-0.948	0.183		ปานกลาง	Reading
131	13.284	0.693	0.321		ปานกลาง	Reading
132	5.555	0.432	0.166		ปานกลาง	Reading
133	7.145	0.088	0.185		ปานกลาง	Reading
134	16.208	0.615	0.216		ปานกลาง	Reading
135	7.991	1.612	0.274		ยาก	Reading
136	10.641	1.955	0.254		ยาก	Reading
137	-0.867	-3.394	0.198	คัดออก		Reading
138	7.674	-0.009	0.177		ปานกลาง	Reading
139	10.594	-0.829	0.176		ปานกลาง	Reading
140	6.910	0.280	0.204		ปานกลาง	Reading
141	6.444	0.992	0.190		ปานกลาง	Reading
142	2.271	-1.981	0.242		ง่าย	Reading
143	7.405	-3.650	0.294		ง่าย	Reading
144	3.436	0.103	0.284		ปานกลาง	Reading
145	16.709	-2.788	0.173		ง่าย	Reading
146	6.819	-1.667	0.315		ง่าย	Reading
147	1.567	-1.512	0.177		ง่าย	Reading
148	11.267	3.839	0.184		ยาก	Reading
149	2.272	-0.509	0.177		ปานกลาง	Reading
150	7.143	2.588	0.215		ยาก	Reading
151	6.683	1.409	0.196		ยาก	Reading
152	12.345	0.724	0.313		ปานกลาง	Reading
153	7.333	2.309	0.247		ยาก	Reading
154	5.654	1.850	0.307		ยาก	Reading
155	9.927	-0.821	0.183		ปานกลาง	Reading
156	0.012	-2.025	0.154		ง่าย	Reading
157	1.685	-2.721	0.306		ปานกลาง	Reading
158	7.396	-0.104	0.228		ปานกลาง	Reading
159	6.421	0.188	0.213		ปานกลาง	Reading

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
160	0.600	-3.860	0.000		ง่าย	Reading
161	7.314	2.306	0.297		ยาก	Miscellaneous
162	6.359	2.715	0.258		ยาก	Miscellaneous
163	4.060	1.459	0.247		ยาก	Miscellaneous
164	9.171	2.178	0.254		ยาก	Miscellaneous
165	6.989	0.616	0.167		ปานกลาง	Miscellaneous
166	8.681	0.509	0.244		ปานกลาง	Miscellaneous
167	9.735	2.110	0.277		ยาก	Miscellaneous
168	8.648	2.511	0.315		ยาก	Miscellaneous
169	6.839	1.205	0.230		ปานกลาง	Miscellaneous
170	11.860	2.565	0.252		ยาก	Miscellaneous
171	10.159	2.355	0.243		ยาก	Miscellaneous
172	33.168	-1.383	0.315		ง่าย	Miscellaneous
173	5.797	1.009	0.202		ปานกลาง	Miscellaneous
174	10.389	2.826	0.228		ยาก	Miscellaneous
175	11.708	3.148	0.256		ยาก	Miscellaneous
176	11.312	2.373	0.273		ยาก	Miscellaneous
177	10.500	2.443	0.214		ยาก	Miscellaneous
178	6.405	0.886	0.243		ปานกลาง	Miscellaneous
179	10.815	2.697	0.167		ยาก	Miscellaneous
180	9.375	2.401	0.225		ยาก	Miscellaneous
181	10.477	-1.639	0.221		ง่าย	Miscellaneous
182	4.366	0.296	0.193		ปานกลาง	Miscellaneous
183	13.737	0.946	0.188		ปานกลาง	Miscellaneous
184	6.148	0.696	0.229		ปานกลาง	Miscellaneous
185	16.887	-1.435	0.225		ง่าย	Miscellaneous
186	2.299	-0.424	0.189		ปานกลาง	Miscellaneous
187	27.496	-5.584	0.222	คัดออก		Miscellaneous
188	27.476	-0.873	0.275		ปานกลาง	Miscellaneous
189	5.435	-2.894	0.184		ง่าย	Miscellaneous
190	9.301	-0.588	0.298		ปานกลาง	Miscellaneous
191	29.061	-0.500	0.173		ปานกลาง	Miscellaneous
192	0.538	-0.584	0.001		ปานกลาง	Miscellaneous



ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
193	3.848	0.069	0.209		ปานกลาง	Miscellaneous
194	5.267	-0.865	0.313		ปานกลาง	Miscellaneous
195	2.173	-0.851	0.204		ปานกลาง	Miscellaneous
196	30.749	-5.594	0.167	คัดออก		Miscellaneous
197	9.799	0.141	0.284		ปานกลาง	Miscellaneous
198	12.494	-2.475	0.178		ง่าย	Miscellaneous
199	4.127	-2.143	0.245		ง่าย	Miscellaneous
200	7.076	-0.214	0.248		ปานกลาง	Miscellaneous
201	-0.631	-1.946	0.254	คัดออก		Conversation
202	0.092	0.185	0.047		ปานกลาง	Conversation
203	-0.103	0.140	0.018	คัดออก		Conversation
204	1.963	-2.218	0.406		ง่าย	Conversation
205	15.924	-1.908	0.129		ง่าย	Conversation
206	0.152	-0.301	0.007		ปานกลาง	Conversation
207	0.067	-2.835	0.192		ง่าย	Conversation
208	-0.263	-0.319	0.008	คัดออก		Conversation
209	0.034	0.112	0.067		ปานกลาง	Conversation
210	0.053	-0.652	0.114		ปานกลาง	Conversation
211	0.227	-0.194	0.002		ปานกลาง	Conversation
212	0.603	-0.808	0.001		ปานกลาง	Conversation
213	-0.082	-0.295	0.037	คัดออก		Conversation
214	0.448	-2.028	0.321		ง่าย	Conversation
215	-0.046	-0.419	0.133	คัดออก		Conversation
216	0.260	-0.296	0.020		ปานกลาง	Conversation
217	0.289	-0.009	0.034		ปานกลาง	Conversation
218	0.290	-0.582	0.003		ปานกลาง	Conversation
219	-0.095	-0.151	0.022	คัดออก		Conversation
220	-0.178	0.120	0.008	คัดออก		Conversation
221	2.917	-1.978	0.335		ง่าย	Conversation
222	0.261	-1.526	0.038		ง่าย	Conversation
223	4.244	-3.557	0.363		ง่าย	Conversation
224	1.165	-0.664	0.046		ปานกลาง	Conversation
225	-	-	0.232	คัดออก		Conversation

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
	11.836	34.683				
226	3.954	-1.949	0.440		ง่าย	Conversation
227	-0.150	-1.136	0.249	คัดออก		Conversation
228	0.048	-1.868	0.133		ง่าย	Conversation
229	-0.070	-1.878	0.219	คัดออก		Conversation
230	-0.095	-0.727	0.059	คัดออก		Conversation
231	0.377	-1.472	0.132		ง่าย	Conversation
232	0.334	-0.943	0.289		ปานกลาง	Conversation
233	-0.027	-0.765	0.209	คัดออก		Conversation
234	-2.615	-8.260	0.554	คัดออก		Conversation
235	0.116	-0.463	0.041		ปานกลาง	Conversation
236	4.281	-2.340	0.383		ง่าย	Conversation
237	-8.482	-24.456	0.376	คัดออก		Reading
238	-0.509	-1.550	0.315	คัดออก		Reading
239	-0.263	-0.816	0.014	คัดออก		Reading
240	0.104	-0.686	0.043		ปานกลาง	Reading
241	0.785	-2.650	0.261		ง่าย	Reading
242	0.203	-0.386	0.017		ปานกลาง	Reading
243	-0.284	-1.299	0.008	คัดออก		Reading
244	0.456	-1.677	0.279		ง่าย	Reading
245	0.934	-0.853	0.192		ปานกลาง	Reading
246	4.397	-2.124	0.228		ง่าย	Reading
247	2.285	-2.078	0.199		ง่าย	Reading
248	-0.064	-0.179	0.080	คัดออก		Reading
249	13.996	-4.121	0.468	คัดออก		Reading
250	0.064	-1.786	0.037		ง่าย	Reading
251	-0.581	-1.472	0.346	คัดออก		Reading
252	-0.416	-1.751	0.155	คัดออก		Reading
253	0.013	-0.623	0.173		ปานกลาง	Reading
254	-5.240	-15.941	0.253	คัดออก		Reading
255	-8.097	-23.707	0.414	คัดออก		Reading
256	2.202	-1.596	0.192		ง่าย	Reading

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
257	2.329	-6.673	0.396	คัดออก		Reading
258	-0.172	-0.245	0.052	คัดออก		Reading
259	-0.141	-0.489	0.101	คัดออก		Reading
260	-0.071	-3.245	0.114	คัดออก		Reading
261	-2.826	-9.298	0.256	คัดออก		Reading
262	-0.838	-1.584	0.165	คัดออก		Reading
263	0.294	-1.782	0.005		ง่าย	Reading
264	0.490	-2.017	0.280		ง่าย	Reading
265	-0.429	-1.396	0.001	คัดออก		Reading
266	-0.118	0.070	0.035	คัดออก		Reading
267	0.669	-1.582	0.209		ง่าย	Reading
268	-0.412	-1.658	0.001	คัดออก		Reading
269	0.684	-1.489	0.340		ง่าย	Reading
270	-0.206	-0.164	0.047	คัดออก		Reading
271	- 23.437	- 53.043	0.566	คัดออก		Reading
272	0.185	-0.913	0.004		ปานกลาง	Reading
273	-0.135	-1.414	0.002	คัดออก		Reading
274	2.965	-2.992	0.441		ง่าย	Reading
275	0.237	-0.359	0.001		ปานกลาง	Reading
276	-0.040	-1.083	0.063	คัดออก		Reading
277	0.051	-0.533	0.035		ปานกลาง	Reading
278	-1.323	-6.742	0.261	คัดออก		Reading
279	0.095	-0.690	0.006		ปานกลาง	Reading
280	4.875	-7.140	0.421	คัดออก		Reading
281	- 24.251	- 60.519	0.299	คัดออก		Reading
282	- 17.509	- 42.639	0.285	คัดออก		Reading
283	- 13.902	- 33.014	0.422	คัดออก		Reading
284	-0.055	-1.434	0.017	คัดออก		Reading
285	0.476	-0.486	0.000		ปานกลาง	Reading
286	-0.073	-1.066	0.017	คัดออก		Reading
287	0.117	-0.554	0.005		ปานกลาง	Reading

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
288	3.460	-6.759	0.405	คัดออก		Reading
289	0.138	-0.564	0.007		ปานกลาง	Reading
290	- 23.715	- 56.237	0.377	คัดออก		Reading
291	0.089	-0.520	0.009		ปานกลาง	Reading
292	0.023	-0.566	0.121		ปานกลาง	Reading
293	-1.504	-3.399	0.397	คัดออก		Reading
294	- 17.388	- 40.037	0.397	คัดออก		Reading
295	0.555	-1.384	0.000		ง่าย	Reading
296	- 16.038	- 36.775	0.494	คัดออก		Reading
297	0.104	-0.540	0.008		ปานกลาง	Reading
298	- 25.359	- 58.047	0.493	คัดออก		Reading
299	-5.839	- 14.356	0.503	คัดออก		Reading
300	1.567	-1.512	0.177		ง่าย	Reading
301	1.112	-1.919	0.145		ง่าย	Reading
302	0.345	-1.319	0.001		ปานกลาง	Reading
303	0.646	-1.173	0.000	คัดออก		Reading
304	0.868	-1.053	0.000	คัดออก		Grammar
305	- 28.304	- 71.482	0.236	คัดออก		Grammar
306	3.028	-2.103	0.212		ง่าย	Grammar
307	0.342	-1.556	0.000		ง่าย	Grammar
308	- 33.560	- 84.019	0.235	คัดออก		Grammar
309	- 26.319	- 64.640	0.306	คัดออก		Grammar
310	0.183	-2.093	0.004		ง่าย	Grammar
311	11.220	-2.969	0.416		ง่าย	Grammar
312	0.089	-1.363	0.010		ง่าย	Grammar
313	0.110	-1.807	0.005		ง่าย	Grammar
314	-4.268	- 10.488	0.413	คัดออก		Grammar
315	5.914	-2.969	0.251		ง่าย	Grammar
316	0.146	-0.311	0.013		ปานกลาง	Grammar

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
317	0.574	-0.442	0.001		ปานกลาง	Grammar
318	-3.935	-9.353	0.376	คัดออก		Grammar
319	-0.183	-1.466	0.002	คัดออก		Grammar
320	-0.065	-1.067	0.016	คัดออก		Grammar
321	- 17.087	- 43.962	0.303	คัดออก		Grammar
322	0.183	-0.754	0.003		ปานกลาง	Grammar
323	5.153	-2.950	0.266		ง่าย	Grammar
324	0.385	-1.434	0.000		ง่าย	Grammar
325	3.856	-1.970	0.211		ง่าย	Grammar
326	0.232	-1.444	0.029		ง่าย	Grammar
327	-0.080	-1.296	0.010	คัดออก		Grammar
328	- 25.738	- 64.242	0.245	คัดออก		Grammar
329	0.629	-1.388	0.000		ง่าย	Grammar
330	0.156	-1.719	0.002		ง่าย	Grammar
331	3.142	-2.233	0.231		ง่าย	Grammar
332	1.650	-9.973	0.334	คัดออก		Grammar
333	5.435	-2.894	0.184		ง่าย	Grammar
334	0.125	-2.006	0.002		ง่าย	Grammar
335	-0.321	-1.533	0.001	คัดออก		Grammar
336	0.532	-0.471	0.000		ปานกลาง	Grammar
337	0.471	-0.278	0.000		ปานกลาง	Grammar
338	4.397	-2.124	0.228		ง่าย	Grammar
339	-3.036	-8.513	0.242	คัดออก		Grammar
340	1.474	-9.532	0.353	คัดออก		Grammar
341	1.054	-0.940	0.215		ปานกลาง	Grammar
342	0.359	-0.514	0.003		ปานกลาง	Grammar
343	-0.080	-0.031	0.032	คัดออก		Grammar
344	-0.086	-0.047	0.031	คัดออก		Grammar
345	5.147	-1.940	0.310		ง่าย	Grammar
346	7.097	-3.590	0.352		ง่าย	Grammar
347	-0.162	-0.291	0.036	คัดออก		Grammar
348	-0.433	-0.940	0.002	คัดออก		Grammar

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
349	-0.428	-0.522	0.001	คัดออก		Grammar
350	0.040	-1.537	0.044		ง่าย	Grammar
351	2.285	-2.078	0.199		ง่าย	Grammar
352	-0.112	-1.808	0.010	คัดออก		Grammar
353	-0.529	-1.089	0.001	คัดออก		Grammar
354	2.202	-1.596	0.192		ง่าย	Grammar
355	0.191	-1.613	0.011		ง่าย	Grammar
356	0.310	-0.028	0.006		ปานกลาง	Grammar
357	0.757	-0.415	0.145		ปานกลาง	Grammar
358	0.814	-0.547	0.160		ปานกลาง	Grammar
359	0.285	-0.264	0.011		ปานกลาง	Grammar
360	1.567	-1.512	0.177		ง่าย	Miscellaneous
361	0.457	-0.231	0.004		ปานกลาง	Miscellaneous
362	0.356	-0.499	0.001		ปานกลาง	Miscellaneous
363	-0.514	-1.049	0.001	คัดออก		Miscellaneous
364	5.417	-2.517	0.220		ง่าย	Miscellaneous
365	0.246	-1.091	0.002		ปานกลาง	Miscellaneous
366	0.833	-3.069	0.167		ง่าย	Miscellaneous
367	0.985	-0.699	0.162		ปานกลาง	Miscellaneous
368	-0.083	-0.370	0.039	คัดออก		Miscellaneous
369	2.353	-2.698	0.164		ง่าย	Miscellaneous
370	3.688	-2.988	0.355		ง่าย	Miscellaneous
371	0.330	-0.196	0.001		ปานกลาง	Miscellaneous
372	1.390	-0.794	0.150		ปานกลาง	Miscellaneous
373	0.261	-0.054	0.008		ปานกลาง	Miscellaneous
374	-0.069	-0.438	0.090	คัดออก		Miscellaneous
375	1.554	-0.883	0.115		ปานกลาง	Miscellaneous
376	0.014	-2.371	0.147		ง่าย	Miscellaneous
377	-0.184	-1.269	0.008	คัดออก		Miscellaneous
378	1.356	-0.748	0.202		ปานกลาง	Miscellaneous
379	0.727	-0.691	0.155		ปานกลาง	Miscellaneous
380	-0.141	-1.882	0.078	คัดออก		Miscellaneous
381	-1.518	-6.590	0.246	คัดออก		Miscellaneous

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
382	0.410	-0.524	0.002		ปานกลาง	Miscellaneous
383	0.370	-0.252	0.003		ปานกลาง	Miscellaneous
384	1.461	-2.291	0.139		ง่าย	Miscellaneous
385	-1.755	-6.802	0.095	คัดออก		Miscellaneous
386	1.661	-9.703	0.325	คัดออก		Miscellaneous
387	0.945	-1.165	0.190		ปานกลาง	Miscellaneous
388	2.275	-3.036	0.136		ง่าย	Miscellaneous
389	-0.866	-5.499	0.194	คัดออก		Miscellaneous
390	0.884	-1.881	0.170		ง่าย	Miscellaneous
391	-0.157	-0.290	0.048	คัดออก		Miscellaneous
392	0.346	-0.103	0.021		ปานกลาง	Miscellaneous
393	0.529	-0.384	0.027		ปานกลาง	Miscellaneous
394	-0.529	-2.980	0.164	คัดออก		Miscellaneous
395	0.256	-0.738	0.006		ปานกลาง	Miscellaneous
396	-0.278	-1.129	0.011	คัดออก		Miscellaneous
397	-0.188	-1.226	0.003	คัดออก		Miscellaneous
398	-2.834	-10.461	0.227	คัดออก		Miscellaneous
399	0.771	-2.476	0.085		ง่าย	Miscellaneous
400	3.283	-2.138	0.140		ง่าย	Miscellaneous
401	1.364	3.686	0.020		ยาก	Conversation
402	1.283	3.296	0.002		ยาก	Conversation
403	3.459	4.810	0.396	คัดออก		Conversation
404	0.868	2.140	0.001		ยาก	Conversation
405	1.051	2.506	0.001		ยาก	Conversation
406	1.465	3.352	0.001		ยาก	Conversation
407	2.274	3.279	0.079		ยาก	Conversation
408	1.681	2.763	0.051		ยาก	Conversation
409	1.189	2.443	0.006		ยาก	Conversation
410	1.348	2.869	0.000		ยาก	Conversation
411	2.501	2.599	0.164		ยาก	Conversation
412	2.821	5.128	0.153	คัดออก		Conversation
413	1.056	2.368	0.001		ยาก	Conversation

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
414	1.329	2.559	0.000		ยาก	Conversation
415	0.607	0.700	0.000		ปานกลาง	Conversation
416	0.553	0.431	0.000		ปานกลาง	Conversation
417	1.234	2.015	0.049		ยาก	Conversation
418	1.887	3.908	0.002		ยาก	Conversation
419	1.979	3.753	0.178		ยาก	Conversation
420	0.828	0.879	0.000		ปานกลาง	Conversation
421	0.872	1.723	0.000		ยาก	Conversation
422	1.264	2.368	0.000		ยาก	Conversation
423	5.147	-1.940	0.310		ง่าย	Conversation
424	1.651	2.961	0.140		ยาก	Conversation
425	1.218	2.382	0.000		ยาก	Conversation
426	2.326	3.936	0.106		ยาก	Conversation
427	1.096	1.803	0.000		ยาก	Conversation
428	0.650	0.635	0.000		ปานกลาง	Conversation
429	2.288	4.441	0.087	คัดออก		Conversation
430	2.271	4.432	0.032	คัดออก		Conversation
431	1.345	1.805	0.077		ยาก	Conversation
432	1.384	2.607	0.000		ยาก	Conversation
433	0.865	1.433	0.001		ยาก	Conversation
434	0.498	-0.387	0.000		ปานกลาง	Conversation
435	0.745	1.024	0.000		ปานกลาง	Conversation
436	0.357	0.470	0.001		ปานกลาง	Conversation
437	1.498	2.333	0.002		ยาก	Conversation
438	0.673	0.712	0.001		ปานกลาง	Conversation
439	1.763	2.293	0.110		ยาก	Conversation
440	-0.287	-1.665	0.001	คัดออก		Conversation
441	0.584	0.535	0.001		ปานกลาง	Conversation
442	-0.589	-2.735	0.000	คัดออก		Conversation
443	3.241	4.129	0.153	คัดออก		Conversation
444	1.038	0.964	0.003		ปานกลาง	Conversation
445	2.163	2.712	0.138		ยาก	Conversation
446	1.679	2.504	0.090		ยาก	Conversation



ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
447	0.814	0.932	0.001		ปานกลาง	Conversation
448	0.379	0.045	0.002		ปานกลาง	Conversation
449	-0.393	-2.079	0.000	คัดออก		Conversation
450	0.146	-0.650	0.004		ปานกลาง	Conversation
451	0.262	0.050	0.004		ปานกลาง	Conversation
452	1.168	2.037	0.000		ยาก	Conversation
453	1.340	1.131	0.128		ปานกลาง	Conversation
454	1.563	1.242	0.091		ปานกลาง	Conversation
455	1.799	1.746	0.115		ยาก	Conversation
456	3.720	4.697	0.165	คัดออก		Conversation
457	2.113	2.226	0.162		ยาก	Conversation
458	2.186	1.843	0.197		ยาก	Conversation
459	2.445	2.549	0.224		ยาก	Conversation
460	0.252	-0.430	0.004		ปานกลาง	Conversation
461	0.590	-1.828	0.180		ง่าย	Grammar
462	1.435	1.189	0.235		ปานกลาง	Grammar
463	0.886	-0.649	0.264		ปานกลาง	Grammar
464	0.228	-0.003	0.016		ปานกลาง	Grammar
465	0.244	-0.399	0.002		ปานกลาง	Grammar
466	0.892	1.341	0.000		ยาก	Grammar
467	8.649	-2.466	0.351		ง่าย	Grammar
468	0.693	1.013	0.000		ปานกลาง	Grammar
469	-0.220	-1.410	0.001	คัดออก		Grammar
470	0.829	0.484	0.000		ปานกลาง	Grammar
471	0.440	-1.630	0.171		ง่าย	Grammar
472	1.223	0.176	0.179		ปานกลาง	Grammar
473	-0.074	-1.825	0.053	คัดออก		Grammar
474	-0.235	-1.429	0.001	คัดออก		Grammar
475	-4.523	- 13.675	0.165	คัดออก		Grammar
476	0.995	0.359	0.233		ปานกลาง	Grammar
477	0.563	-1.625	0.142		ง่าย	Grammar
478	0.205	-0.960	0.157		ปานกลาง	Grammar

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
479	-2.461	- 10.040	0.251	คัดออก		Grammar
480	-0.065	-0.891	0.018	คัดออก		Grammar
481	5.490	2.347	0.518		ยาก	Grammar
482	0.242	-0.670	0.005		ปานกลาง	Grammar
483	-0.026	-1.250	0.149	คัดออก		Grammar
484	-0.136	-1.354	0.004	คัดออก		Grammar
485	-0.542	-2.817	0.000	คัดออก		Grammar
486	-0.268	-1.861	0.002	คัดออก		Grammar
487	- 12.142	- 35.300	0.207	คัดออก		Grammar
488	6.606	-3.234	0.349		ง่าย	Grammar
489	0.358	0.218	0.004		ปานกลาง	Grammar
490	8.468	-1.677	0.377		ง่าย	Grammar
491	2.467	1.098	0.301		ปานกลาง	Grammar
492	5.591	1.899	0.368		ยาก	Grammar
493	-0.179	-1.186	0.003	คัดออก		Grammar
494	-0.658	-2.471	0.000	คัดออก		Grammar
495	3.756	0.568	0.348		ปานกลาง	Grammar
496	-0.010	-1.952	0.166	คัดออก		Grammar
497	-0.085	-0.979	0.107	คัดออก		Grammar
498	0.154	-0.863	0.007		ปานกลาง	Grammar
499	3.034	0.877	0.216		ปานกลาง	Grammar
500	3.835	2.665	0.308		ยาก	Grammar
501	10.691	-2.064	0.335		ง่าย	Grammar
502	-0.430	-1.254	0.002	คัดออก		Grammar
503	1.116	-0.949	0.079		ปานกลาง	Grammar
504	0.606	-0.858	0.299		ปานกลาง	Grammar
505	0.200	0.066	0.013		ปานกลาง	Grammar
506	-0.301	-1.043	0.001	คัดออก		Grammar
507	0.058	-2.843	0.178		ง่าย	Grammar
508	0.060	-0.942	0.054		ปานกลาง	Grammar
509	4.482	-2.471	0.272		ง่าย	Grammar
510	0.726	0.031	0.107		ปานกลาง	Grammar

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
511	-7.830	- 24.210	0.221	คัดออก		Grammar
512	0.122	-1.013	0.007		ปานกลาง	Grammar
513	0.208	0.215	0.034		ปานกลาง	Grammar
514	11.681	-1.565	0.387		ง่าย	Grammar
515	1.863	-0.554	0.296		ปานกลาง	Grammar
516	1.094	0.030	0.246		ปานกลาง	Grammar
517	2.960	1.439	0.147		ยาก	Grammar
518	0.726	-0.426	0.195		ปานกลาง	Grammar
519	0.060	-0.912	0.019		ปานกลาง	Grammar
520	1.260	0.276	0.195		ปานกลาง	Grammar
521	-0.088	-1.699	0.025	คัดออก		Grammar
522	-0.575	-1.826	0.001	คัดออก		Grammar
523	7.211	-0.419	0.322		ปานกลาง	Grammar
524	-0.015	-0.120	0.023	คัดออก		Grammar
525	2.546	0.949	0.405		ปานกลาง	Grammar
526	0.645	0.464	0.015		ปานกลาง	Grammar
527	3.993	-2.614	0.292		ง่าย	Grammar
528	0.200	-2.038	0.211		ง่าย	Grammar
529	0.032	-2.033	0.044		ง่าย	Grammar
530	-0.073	-0.980	0.027	คัดออก		Grammar
531	0.670	-0.316	0.338		ปานกลาง	Grammar
532	0.775	-0.415	0.116		ปานกลาง	Grammar
533	-0.406	-2.196	0.001	คัดออก		Grammar
534	-0.143	-0.887	0.016	คัดออก		Grammar
535	1.664	0.237	0.171		ปานกลาง	Grammar
536	1.918	0.284	0.231		ปานกลาง	Grammar
537	2.396	1.272	0.444		ปานกลาง	Grammar
538	3.161	2.071	0.365		ยาก	Grammar
539	2.087	1.213	0.313		ปานกลาง	Grammar
540	1.695	0.305	0.364		ปานกลาง	Grammar
541	2.070	1.059	0.288		ปานกลาง	Grammar
542	0.652	-0.316	0.074		ปานกลาง	Grammar

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
543	2.075	0.873	0.253		ปานกลาง	Grammar
544	0.277	-0.206	0.023		ปานกลาง	Grammar
545	7.711	4.904	0.292	คัดออก		Grammar
546	1.341	0.212	0.199		ปานกลาง	Grammar
547	-0.372	-1.752	0.000	คัดออก		Grammar
548	1.523	-0.356	0.296		ปานกลาง	Grammar
549	2.376	0.759	0.232		ปานกลาง	Grammar
550	0.891	-0.321	0.183		ปานกลาง	Grammar
551	1.282	-0.600	0.125		ปานกลาง	Reading
552	1.201	0.229	0.296		ปานกลาง	Reading
553	2.294	0.307	0.360		ปานกลาง	Reading
554	2.461	-0.758	0.235		ปานกลาง	Reading
555	2.842	-0.329	0.210		ปานกลาง	Reading
556	3.381	0.486	0.371		ปานกลาง	Reading
557	0.571	0.771	0.033		ปานกลาง	Reading
558	4.980	-1.421	0.390		ง่าย	Reading
559	5.175	2.585	0.346		ยาก	Reading
560	0.464	-0.466	0.000		ปานกลาง	Reading
561	-0.167	-1.640	0.002	คัดออก		Reading
562	2.783	2.065	0.104		ยาก	Reading
563	1.429	-0.616	0.175		ปานกลาง	Reading
564	3.205	1.744	0.423		ยาก	Reading
565	0.359	-0.233	0.003		ปานกลาง	Reading
566	-0.008	-2.255	0.186	คัดออก		Reading
567	-0.008	-2.255	0.186	คัดออก		Reading
568	3.065	1.675	0.212		ยาก	Reading
569	0.078	-1.035	0.039		ปานกลาง	Reading
570	-0.603	-1.801	0.000	คัดออก		Reading
571	3.936	-0.962	0.265		ปานกลาง	Reading
572	2.467	1.934	0.444		ยาก	Reading
573	0.419	-0.670	0.001		ปานกลาง	Reading
574	0.759	0.375	0.109		ปานกลาง	Reading
575	1.740	0.741	0.328		ปานกลาง	Reading

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
576	0.798	-0.264	0.188		ปานกลาง	Reading
577	0.149	-0.966	0.019		ปานกลาง	Reading
578	-0.023	-3.513	0.354	คัดออก		Reading
579	1.225	0.097	0.228		ปานกลาง	Reading
580	0.506	0.147	0.003		ปานกลาง	Reading
581	2.892	1.721	0.291		ยาก	Reading
582	0.106	-1.043	0.018		ปานกลาง	Reading
583	2.774	0.329	0.245		ปานกลาง	Reading
584	-0.085	-1.635	0.016	คัดออก		Reading
585	2.474	0.520	0.356		ปานกลาง	Reading
586	0.076	-0.647	0.014		ปานกลาง	Reading
587	-0.201	-1.128	0.001	คัดออก		Reading
588	-0.016	-0.213	0.053	คัดออก		Reading
589	-0.451	-1.836	0.000	คัดออก		Reading
590	1.490	0.894	0.315		ปานกลาง	Reading
591	2.981	1.362	0.301		ยาก	Reading
592	-0.150	-1.341	0.004	คัดออก		Reading
593	3.863	2.475	0.376		ยาก	Reading
594	6.052	4.200	0.296	คัดออก		Reading
595	4.648	3.548	0.282		ยาก	Reading
596	1.604	0.614	0.311		ปานกลาง	Reading
597	2.175	0.823	0.338		ปานกลาง	Reading
598	1.318	0.679	0.270		ปานกลาง	Reading
599	3.612	2.238	0.367		ยาก	Reading
600	1.071	0.211	0.208		ปานกลาง	Reading
601	-0.311	-1.919	0.001	คัดออก		Reading
602	2.195	0.164	0.243		ปานกลาง	Reading
603	1.240	-3.572	0.244		ง่าย	Reading
604	1.051	0.362	0.302		ปานกลาง	Reading
605	5.853	3.703	0.176		ยาก	Reading
606	2.002	-0.693	0.191		ปานกลาง	Reading
607	1.596	-1.389	0.352		ง่าย	Reading
608	2.784	1.212	0.393		ยาก	Reading

ข้อที่	a	b	c	ผลประเมิน	กลุ่ม	ด้าน
609	3.664	1.859	0.346		ยาก	Reading
610	1.472	-0.249	0.297		ปานกลาง	Miscellaneous
611	0.286	0.042	0.006		ปานกลาง	Miscellaneous
612	3.280	1.853	0.334		ยาก	Miscellaneous
613	1.848	0.328	0.294		ปานกลาง	Miscellaneous
614	1.853	-0.161	0.228		ปานกลาง	Miscellaneous
615	-0.035	-2.446	0.276	คัดออก		Miscellaneous
616	0.051	-1.067	0.020		ปานกลาง	Miscellaneous
617	-0.287	-1.632	0.001	คัดออก		Miscellaneous
618	1.038	0.098	0.289		ปานกลาง	Miscellaneous
619	-0.497	-1.748	0.000	คัดออก		Miscellaneous
620	3.266	2.139	0.334		ยาก	Miscellaneous
621	1.500	-0.005	0.245		ปานกลาง	Miscellaneous
622	0.355	0.461	0.007		ปานกลาง	Miscellaneous
623	2.072	-1.067	0.245		ปานกลาง	Miscellaneous
624	0.343	-0.758	0.003		ปานกลาง	Miscellaneous
625	2.177	1.299	0.195		ปานกลาง	Miscellaneous
626	0.048	-1.083	0.040		ปานกลาง	Miscellaneous
627	0.937	-0.198	0.145		ปานกลาง	Miscellaneous
628	1.340	0.070	0.261		ปานกลาง	Miscellaneous
629	2.656	1.687	0.259		ยาก	Miscellaneous
630	3.269	-1.969	0.218		ง่าย	Miscellaneous
631	-0.133	-1.276	0.006	คัดออก		Miscellaneous
632	-0.102	-1.085	0.010	คัดออก		Miscellaneous
633	-0.055	-1.549	0.046	คัดออก		Miscellaneous
634	-5.179	-16.497	0.210	คัดออก		Miscellaneous
635	1.480	-0.117	0.195		ปานกลาง	Miscellaneous
636	0.173	-0.848	0.002		ปานกลาง	Miscellaneous
637	4.178	-0.178	0.253		ปานกลาง	Miscellaneous
638	-1.711	-5.746	0.203	คัดออก		Miscellaneous
639	-0.039	-2.973	0.177	คัดออก		Miscellaneous
640	1.340	-1.099	0.274		ปานกลาง	Miscellaneous

ภาคผนวก ค  
เครื่องมือที่ใช้ในการวิจัย

**แบบประเมินอิงมาตรฐานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด**

**คำชี้แจง:** ให้ท่านพิจารณาข้อความต่อไปนี้แล้วทำเครื่องหมาย ✓ ลงในช่องที่ตรงกับความคิดเห็นของท่านมากที่สุด โดยมียกระดับความคิดเห็น 5 ระดับ ดังนี้

1 หมายถึง น้อยที่สุด, 2 หมายถึง น้อย, 3 หมายถึง ปานกลาง, 4 หมายถึง มาก, 5 หมายถึง มากที่สุด

ข้อคำถาม	ระดับความคิดเห็น				
	1 น้อยที่สุด	2 น้อย	3 ปานกลาง	4 มาก	5 มากที่สุด
<b>ความมีประโยชน์ (utility)</b>					
1. ระบบตอบสนองความต้องการของผู้ใช้งาน					
2. การรายงานผลมีความชัดเจน เข้าใจง่าย					
3. ระบบการทดสอบสามารถนำไปใช้ประโยชน์ในการพัฒนาทางด้านทักษะภาษาอังกฤษของผู้สอบได้					
<b>ความเป็นไปได้ (feasibility)</b>					
1. กระบวนการและขั้นตอนของระบบสามารถนำไปปฏิบัติได้จริง					
2. ระบบมีการเรียกใช้งานง่าย สะดวก และรวดเร็ว					
3. ผลที่ได้จากการทดสอบมีความคุ้มค่า					
4. คำถามควรเปลี่ยนเป็นระบบมีความสอดคล้องตามสถานการณ์ปัจจุบัน					
<b>ความเหมาะสม (propriety)</b>					
1. ระบบการทดสอบมีคู่มือแนะนำการใช้งานที่เหมาะสมเข้าใจง่าย					
2. ขอบเขตของเนื้อหาในการทดสอบมีความเหมาะสม					
3. ระยะเวลาในการทดสอบมีความเหมาะสม					
4. การรายงานผลการทดสอบมีความครบถ้วน ครอบคลุม					
<b>ความถูกต้อง (accuracy)</b>					
1. ระบบจัดการทดสอบได้ตรงตามวัตถุประสงค์					
2. กระบวนการทดสอบของระบบมีความถูกต้อง					
3. การรายงานผลคะแนนมีความถูกต้องตรงตามคุณลักษณะที่แท้จริง					

### แบบประเมินอิทธิพลระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

**คำชี้แจง:** ให้ท่านพิจารณาข้อความต่อไปนี้แล้วทำเครื่องหมาย ✓ ลงในช่องที่ตรงกับความคิดเห็นของท่านมากที่สุด โดยมีระดับความคิดเห็น 5 ระดับ ดังนี้

- 1 หมายถึง น้อยที่สุด
- 2 หมายถึง น้อย
- 3 หมายถึง ปานกลาง
- 4 หมายถึง มาก
- 5 หมายถึง มากที่สุด

ข้อคำถาม	ระดับความคิดเห็น				
	1 น้อยที่สุด	2 น้อย	3 ปานกลาง	4 มาก	5 มากที่สุด
1. มีการแสดงสถานะของระบบอย่างชัดเจน					
2. ระบบมีความสอดคล้องกับโลกของความเป็นจริง					
3. การควบคุมระบบมีประสิทธิภาพและผู้ใช้งานมีความสะดวกในการเข้าถึง					
4. การทำงานของระบบมีความเสถียร					
5. ระบบมีการป้องกันความผิดพลาด					
6. ระบบมีความสามารถในการจดจำข้อมูลมากกว่าการเรียกข้อมูล					
7. ระบบมีความยืดหยุ่นและประสิทธิภาพของการใช้งาน					
8. ระบบมีรูปแบบสวยงามและเรียบง่าย					
9. ระบบมีการช่วยเหลือผู้ใช้งานระบุตัวตน วินิจฉัยและกู้คืนจากข้อผิดพลาด					
10. ระบบมีการช่วยเหลือและคู่มือการใช้งาน					
11. ระบบมีการสนับสนุนและเพิ่มทักษะของผู้ใช้งาน					
12. ระบบมีการตรวจสอบความพึงพอใจและการยอมรับปฏิสัมพันธ์กับผู้ใช้งาน					
13. ระบบมีการป้องกันข้อมูลส่วนบุคคล					



**แบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบ  
ไฮบริด**

**คำชี้แจง:** ให้ท่านพิจารณาข้อความต่อไปนี้แล้วทำเครื่องหมาย ✓ ลงในช่องที่ตรงกับระดับความพึงพอใจของท่านมากที่สุด โดยมีระดับความพึงพอใจ 5 ระดับ ดังนี้

1 หมายถึง น้อยที่สุด, 2 หมายถึง น้อย, 3 หมายถึง ปานกลาง, 4 หมายถึง มาก, 5 หมายถึง มากที่สุด

คำถาม	ระดับความพึงพอใจ				
	1 น้อยที่สุด	2 น้อย	3 ปานกลาง	4 มาก	5 มากที่สุด
1. การทำงานของระบบสะดวก ง่าย ต่อการใช้งาน และไม่ซับซ้อน					
2. ผู้ใช้งานสามารถเรียนรู้การใช้งานและการทำงานของระบบด้วยตนเอง					
3. กระบวนการทำงานของระบบมีความรวดเร็วในการเรียกใช้บริการ					
4. รูปแบบของข้อความ ส่วนประกอบต่าง ๆ บนหน้าจอมีความเหมาะสม					
5. ลำดับการแสดงผลภาพและส่วนประกอบต่างๆ บนหน้าจอเป็นลำดับขั้นตอน					
6. การออกแบบการใช้สีที่เหมาะสมต่อการใช้งาน					
7. ความรวดเร็วในการรายงานผลการทดสอบ					
8. การรายงานผลการทดสอบมีความชัดเจน เข้าใจง่าย					
9. ผลการทดสอบนำไปใช้ในการเตรียมความพร้อมก่อน5การทดสอบภาษาอังกฤษได้					
10. ความรู้สึกโดยรวมที่มีต่อระบบ					



ภาคผนวก ง

คู่มือการใช้งานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY



คู่มือการใช้งานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด  
(HYBRID COMPUTERIZED ADAPTIVE TESTING SYSTEM)

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

## คำนำ

คู่มือการใช้งานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด จัดทำขึ้นเพื่อวัดทักษะภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3 มีองค์ประกอบ 2 ส่วน ได้แก่ ส่วนที่ 1 คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป (user) และ ส่วนที่ 2 คู่มือการใช้งานสำหรับผู้จัดการระบบ (admin) ผู้วิจัยหวังเป็นอย่างยิ่งว่าคู่มือเล่มนี้จะมีประโยชน์อย่างมากต่อผู้ใช้ระบบดังกล่าวข้างต้น



## สารบัญ

	หน้า
คำนำ	ก
สารบัญ	๗
สารบัญภาพ	ค
คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป (user)	1
- การลงทะเบียน	1
- การทดสอบ	7
- การรายงานผลการทดสอบ	9
คู่มือการใช้งานสำหรับผู้จัดการระบบ (admin)	11
- การจัดการข้อมูล	12
- การรายงานผลการทดสอบ	18

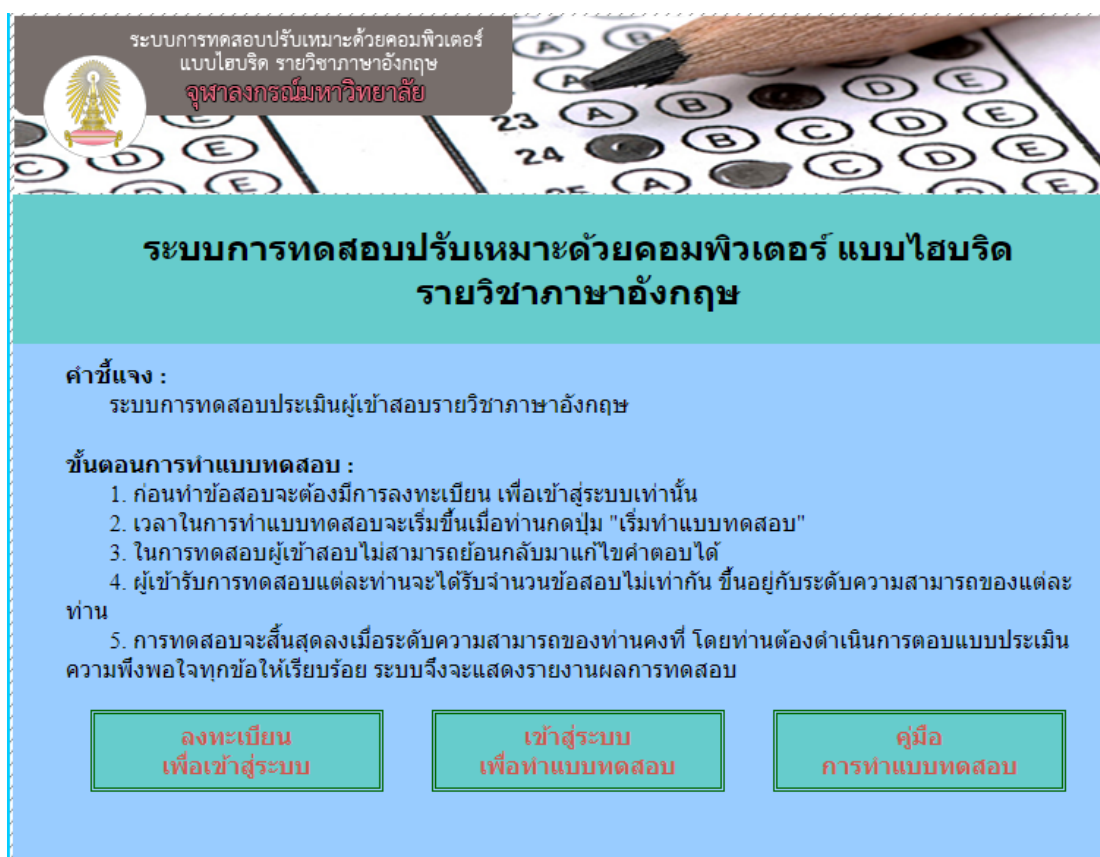
## สารบัญภาพ

### หน้า

ภาพ 1 หน้าจอหลัก	1
ภาพ 2 หน้าจอหลักระบบการทดสอบ	2
ภาพ 3 การลงทะเบียนผู้เข้าสอบ	3
ภาพ 4 หน้าจอคู่มือการใช้งานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด	3
ภาพ 5 การเข้าระบบการทดสอบ	4
ภาพ 6 หน้าจอแสดงรายละเอียดเกี่ยวกับผู้เข้ารับการทดสอบ	5
ภาพ 7 หน้าจอแก้ไขข้อมูลส่วนตัว	6
ภาพ 8 หน้าจอแสดงประวัติการทำข้อสอบ	7
ภาพ 9 หน้าจอแสดงรายละเอียดคำชี้แจง	8
ภาพ 10 หน้าจอข้อสอบ	8
ภาพ 11 หน้าจอรายงานผลการทดสอบ	9
ภาพ 12 แบบประเมินความพึงพอใจต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด	10
ภาพ 13 การเข้าระบบของผู้จัดการระบบ	11
ภาพ 14 หน้าจอหลักของผู้จัดการระบบ	12
ภาพ 15 การจัดการข้อมูลโรงเรียน	13
ภาพ 16 ตัวอย่างชุดข้อสอบสำหรับการทดสอบขั้นแรกและขั้นที่สอง	14
ภาพ 17 การเพิ่มข้อสอบ	15
ภาพ 18 ตัวอย่างข้อสอบที่จะแก้ไข	16
ภาพ 19 ตัวอย่างการแก้ไขข้อสอบ	17
ภาพ 20 ตัวอย่างการแก้ไขข้อมูลของสมาชิก	18

### คู่มือการใช้งานสำหรับผู้ใช้งานทั่วไป (user)

ผู้ใช้งานสามารถเข้าสู่ระบบการทดสอบโดยพิมพ์ Url: <http://www.hybrid-cat.com> ซึ่งจะเจอหน้าจอหลัก ดังภาพ 1



ภาพ 1 หน้าจอหลัก

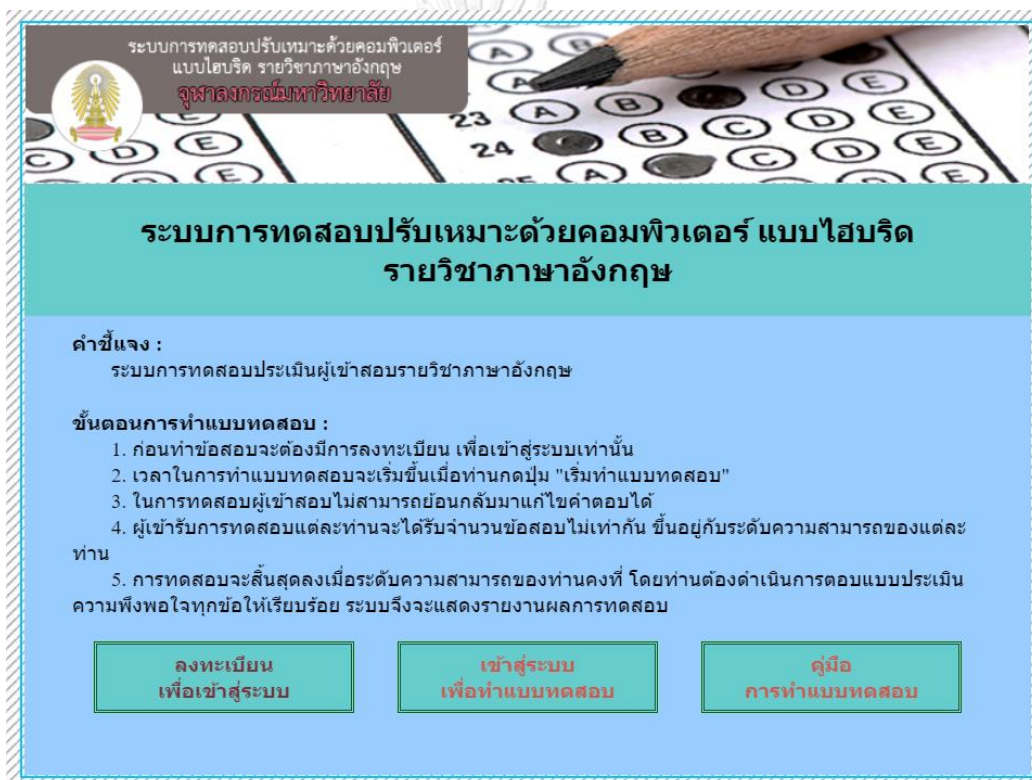
หน้าจอหลักของระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด สำหรับผู้ใช้งานทั่วไป ประกอบด้วย

1. คำชี้แจง ระบบการทดสอบประเมินผู้เข้าสอบรายวิชาภาษาอังกฤษ
2. ขั้นตอนการทำแบบทดสอบ มีรายละเอียดดังนี้
  - 1) ก่อนทำข้อสอบจะต้องมีการลงทะเบียน เพื่อเข้าสู่ระบบเท่านั้น
  - 2) เวลาในการทำแบบทดสอบจะเริ่มขึ้นเมื่อท่านกดปุ่ม "เริ่มทำแบบทดสอบ"
  - 3) ในการทดสอบผู้เข้าสอบไม่สามารถย้อนกลับมาแก้ไขคำตอบได้
  - 4) ผู้เข้ารับการทดสอบแต่ละท่านจะได้รับจำนวนข้อสอบไม่เท่ากัน ขึ้นอยู่กับระดับความสามารถของแต่ละท่าน
  - 5) การทดสอบจะสิ้นสุดลงเมื่อระดับความสามารถของท่านคงที่ โดยท่านต้องดำเนินการตอบแบบประเมินความพึงพอใจทุกข้อให้เรียบร้อย ระบบจึงจะแสดงรายงานผลการทดสอบ

3. ลงทะเบียนเพื่อเข้าสู่ระบบ
4. เข้าสู่ระบบเพื่อทำแบบทดสอบ

#### การลงทะเบียน

ในหน้าจอหลัก ผู้ใช้งานคลิก\_ลงทะเบียนเพื่อเข้าสู่ระบบ เพื่อลงทะเบียน ดังภาพ 2 โดยเมื่อกดปุ่มลงทะเบียนเพื่อเข้าสู่ระบบ จะให้ผู้ใช้งานกรอกข้อมูลส่วนบุคคล ได้แก่ E-mail – รหัสผู้ใช้งาน รหัสผ่าน ยืนยันรหัสผ่าน ชื่อ นามสกุล เบอร์โทรศัพท์ และโรงเรียนของผู้ใช้งาน จะปรากฏหน้าจอ ดังภาพ 3



ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

### ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ แบบไฮบริด รายวิชาภาษาอังกฤษ

**คำชี้แจง :**  
ระบบการทดสอบประเมินผู้เข้าสอบรายวิชาภาษาอังกฤษ

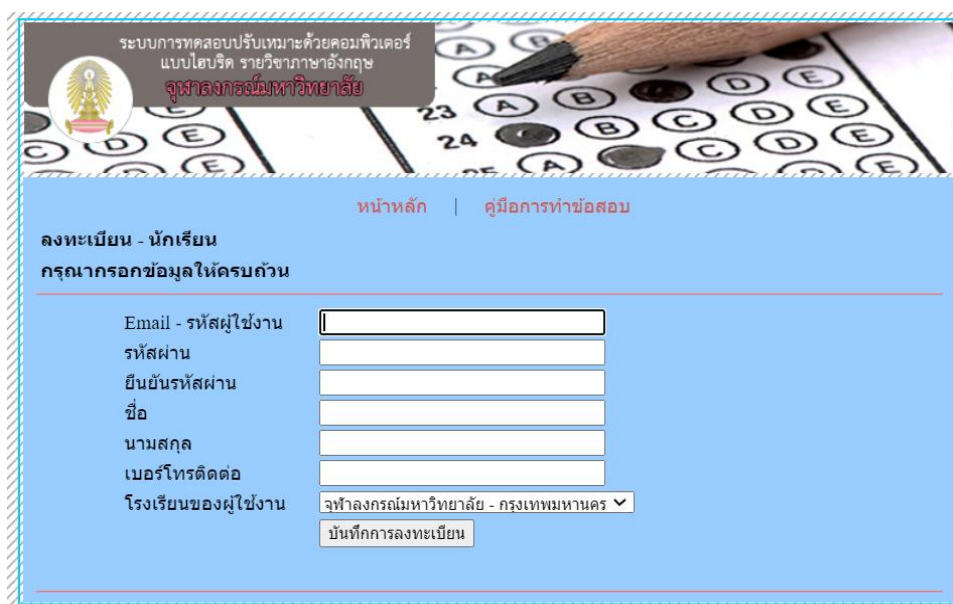
**ขั้นตอนการทำแบบทดสอบ :**

1. ก่อนทำข้อสอบจะต้องมีการลงทะเบียน เพื่อเข้าสู่ระบบเท่านั้น
2. เวลาในการทำแบบทดสอบจะเริ่มขึ้นเมื่อท่านกดปุ่ม "เริ่มทำแบบทดสอบ"
3. ในการทดสอบผู้เข้าสอบไม่สามารถย้อนกลับมาแก้ไขคำตอบได้
4. ผู้เข้ารับการทดสอบแต่ละท่านจะได้รับจำนวนข้อสอบไม่เท่ากัน ขึ้นอยู่กับระดับความสามารถของแต่ละท่าน
5. การทดสอบจะสิ้นสุดลงเมื่อระดับความสามารถของท่านคงที่ โดยท่านต้องดำเนินการตอบแบบประเมินความพึงพอใจทุกข้อให้เรียบร้อย ระบบจึงจะแสดงรายงานผลการทดสอบ

**ลงทะเบียนเพื่อเข้าสู่ระบบ**      **เข้าสู่ระบบเพื่อทำแบบทดสอบ**      **คู่มือการทำแบบทดสอบ**

ภาพ 2 หน้าจอหลักระบบการทดสอบ





ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

หน้าหลัก | คู่มือการทำข้อสอบ

ลงทะเบียน - นักเรียน  
กรณณากรอกข้อมูลให้ครบถ้วน

Email - รหัสผู้ใช้งาน

รหัสผ่าน

ยืนยันรหัสผ่าน

ชื่อ

นามสกุล

เบอร์โทรศัพท์

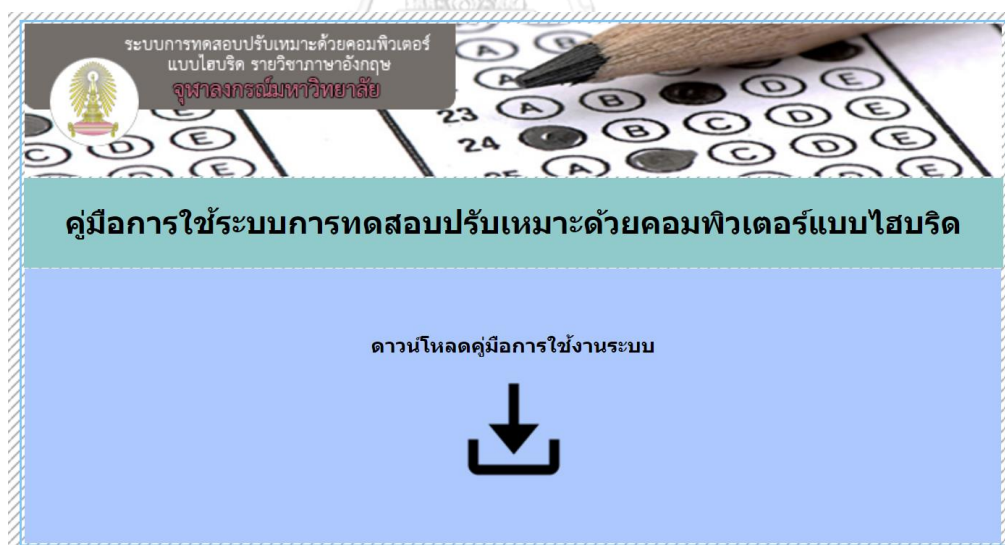
โรงเรียนของผู้ใช้งาน

จุฬาลงกรณ์มหาวิทยาลัย - กรุงเทพมหานคร

บันทึกการลงทะเบียน

ภาพ 3 การลงทะเบียนผู้เข้าสอบ

ในหน้าจอหลัก หากผู้ใช้งานเลือก “คู่มือการทำข้อสอบ” และคลิกที่ดาวน์โหลดคู่มือการใช้งานระบบ จะมีรายละเอียดคู่มือการใช้งานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด หน้าจอ ดังภาพ 4




ภาพ 4 หน้าจอคู่มือการใช้งานระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด

เมื่อผู้ใช้งานลงทะเบียนสมัครสมาชิกสมบูรณ์แล้ว จะดำเนินการเข้าสู่ระบบการทดสอบในหน้าแรก โดยผู้ใช้งานดำเนินการกรอกรหัสผู้ใช้งาน และรหัสผ่านให้ถูกต้อง หลังจากนั้นคลิกปุ่ม ตกลง เพื่อเข้าสู่ระบบ จะปรากฏหน้าจอ ดังภาพ 5

ภาพ 5 การเข้าระบบการทดสอบ

จากภาพ 5 เมื่อผู้ใช้งานเข้าสู่ระบบแล้ว จะปรากฏหน้าจอที่แสดงรายละเอียดเกี่ยวกับผู้เข้ารับการทดสอบ ประกอบด้วยเมนู แก้ไขข้อมูลส่วนตัว ประวัติการทำแบบทดสอบ คู่มือการทำแบบทดสอบ ดังภาพ 6



ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

แก้ไขข้อมูลส่วนตัว | ประวัติการทำแบบทดสอบ | คู่มือการทำแบบทดสอบ | ออกจากระบบ

**คำชี้แจง :**  
ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ แบบไฮบริด รายวิชาภาษาอังกฤษ

**ขั้นตอนการทำแบบทดสอบ :**

1. ก่อนทำข้อสอบจะต้องมีการลงทะเบียน เพื่อเข้าสู่ระบบเท่านั้น
2. เวลาในการทำแบบทดสอบจะเริ่มขึ้นเมื่อท่านกดปุ่ม "เริ่มทำแบบทดสอบ"
3. ในการทดสอบผู้เข้าสอบไม่สามารถย้อนกลับมาแก้ไขคำตอบได้
4. ผู้เข้ารับการทดสอบแต่ละท่านจะได้รับจำนวนข้อสอบไม่เท่ากัน ขึ้นอยู่กับระดับความสามารถของแต่ละท่าน
5. การทดสอบจะสิ้นสุดลงเมื่อระดับความสามารถของท่านคงที่ โดยท่านต้องดำเนินการตอบแบบประเมินความพึงพอใจทุกข้อให้เรียบร้อย ระบบจึงจะแสดงรายงานผลการทดสอบ

[เริ่มทำแบบทดสอบ](#)

ภาพ 6 หน้าจอแสดงรายละเอียดเกี่ยวกับผู้เข้ารับการทดสอบ

จากภาพ 6 เมื่อใช้งานต้องการแก้ไขข้อมูลส่วนตัว สามารถกดที่เมนู แก้ไขข้อมูลส่วนตัว จะปรากฏหน้าจอดังภาพ 7 หลังจากนั้นผู้ใช้งานสามารถแก้ไขข้อมูล เมื่อดำเนินการแก้ไขเรียบร้อยแล้ว คลิก “บันทึกการแก้ไขข้อมูล”

ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

แก้ไขข้อมูลส่วนตัว | ประวัติการทำข้อสอบ | คู่มือการทำข้อสอบ | ออกจากระบบ

**แก้ไขข้อมูลการลงทะเบียน - นักเรียน**  
กรุณากรอกข้อมูลให้ครบถ้วน

ลำดับที่	107
Email - รหัสผู้ใช้งาน	rungsimaporn.n@gmail.com
รหัสผ่าน	<input type="password"/>
ยืนยันรหัสผ่าน	<input type="password"/>
ชื่อ	Rungsimaporn
นามสกุล	Noonoi
เบอร์โทรศัพท์ติดต่อ	-
โรงเรียนของผู้ใช้งาน	จุฬาลงกรณ์มหาวิทยาลัย - กรุงเทพมหานคร ▼

บันทึกการแก้ไขข้อมูล

ภาพ 7 หน้าจอแก้ไขข้อมูลส่วนตัว

จากภาพ 6 เมื่อผู้ใช้งานต้องการทราบข้อมูลเกี่ยวกับประวัติการทำข้อสอบ สามารถกดที่เมนู “ประวัติการทำแบบทดสอบ” จะแสดงผลการทดสอบของผู้ใช้งาน จะปรากฏหน้าจอ ดังภาพ 8

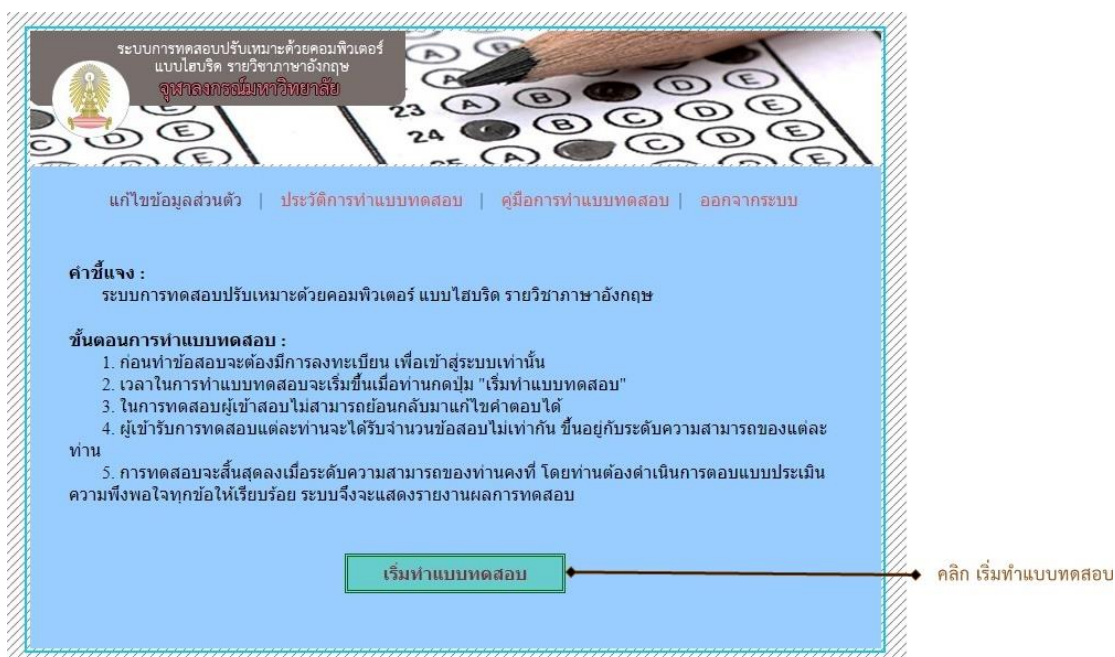
ครั้งที่ 2 วันที่ 04 / 06 / 2020						
ด้าน	ผลการทำแบบทดสอบ (จำนวนข้อ)			ร้อยละของ ระดับความสามารถ	T-Score	ผล
	ทั้งหมด	ถูก	ผิด			
Conversation	3	3	0	100	57.8	ดีมาก
Reading	5	5	0	100	57.8	ดีมาก
Grammar	2	1	1	50	36.9	พอใช้
Miscellaneous	4	3	1	75	47.4	ดี
รวม	14	12	2	85.71		
ระดับความสามารถจากการทดสอบครั้งนี้ <b>ดีมาก</b>  ท่านได้คะแนนสอบภาษาอังกฤษ ด้าน <b>Conversation</b> มากที่สุด คิดเป็นร้อยละ 100 ของข้อที่ทำได้ ด้าน <b>Reading</b> มากที่สุด คิดเป็นร้อยละ 100 ของข้อที่ทำได้ ด้าน <b>Grammar</b> น้อยที่สุด คิดเป็นร้อยละ 50 ของข้อที่ทำได้						
ครั้งที่ 1 วันที่ 04 / 06 / 2020						
ด้าน	ผลการทำแบบทดสอบ (จำนวนข้อ)			ร้อยละของ ระดับความสามารถ	T-Score	ผล
	ทั้งหมด	ถูก	ผิด			
Conversation	4	4	0	100	57.4	ดีมาก
Reading	5	4	1	80	48.9	ดีมาก
Grammar	2	1	1	50	36.2	พอใช้
Miscellaneous	3	3	0	100	57.4	ดีมาก
รวม	14	12	2	85.71		
ระดับความสามารถจากการทดสอบครั้งนี้ <b>ดีมาก</b>  ท่านได้คะแนนสอบภาษาอังกฤษ ด้าน <b>Conversation</b> มากที่สุด คิดเป็นร้อยละ 100 ของข้อที่ทำได้ ด้าน <b>Miscellaneous</b> มากที่สุด คิดเป็นร้อยละ 100 ของข้อที่ทำได้ ด้าน <b>Grammar</b> น้อยที่สุด คิดเป็นร้อยละ 50 ของข้อที่ทำได้						

จุฬาลงกรณ์มหาวิทยาลัย  
ภาพ 8 หน้าจอแสดงประวัติการทำข้อสอบ  
CHULALONGKORN UNIVERSITY

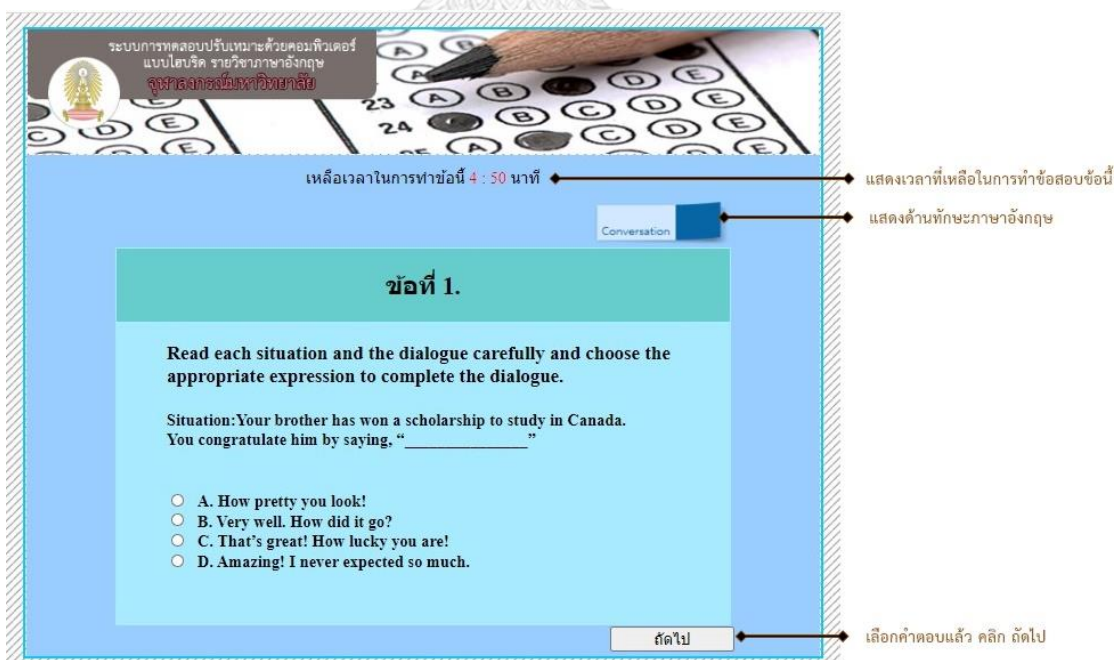
### การทดสอบ

จากภาพ 6 เมื่อผู้ใช้งานดำเนินการเข้าระบบเรียบร้อยแล้ว จะพบหน้าต่างชี้แจง และขั้นตอนการทำแบบทดสอบ เมื่อคลิก “เริ่มทำแบบทดสอบ” ดังภาพ 9 จะเข้าสู่ห้องสอบ จะปรากฏหน้าจอ ดังภาพ 10 โดยแสดงข้อสอบทีละข้อ จะมีการแสดงเวลาที่เหลือในการทำข้อสอบ และแสดงเนื้อหาทักษะภาษาอังกฤษของข้อสอบ เมื่อผู้เข้ารับการทดสอบเลือกคำตอบแล้ว คลิกที่ปุ่ม “ถัดไป” เพื่อไปทำในข้อถัดไปเรื่อย ๆ จนกระทั่งถึงเกณฑ์การยุติการทดสอบ





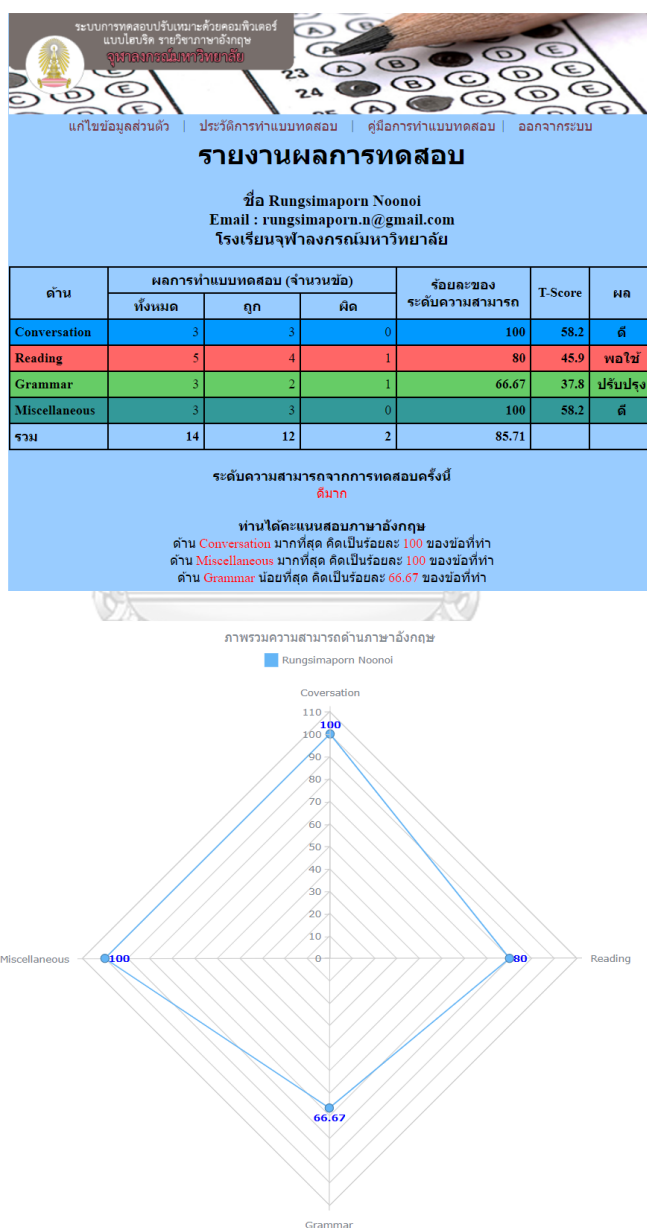
ภาพ 9 หน้าจอแสดงรายละเอียดคำชี้แจง



ภาพ 10 หน้าจอข้อสอบ

### การรายงานผลการทดสอบ

เมื่อผู้เข้ารับการทดสอบทำแบบทดสอบครบถ้วนจนถึงการยุติการทดสอบตามเกณฑ์ที่ระบบกำหนดไว้ จะขึ้นหน้าจอแสดงผลการทดสอบ โดยจะมีการรายงานด้านเนื้อหาในการทำการทดสอบ ผลการทำแบบทดสอบ ประกอบไปด้วย ข้อสอบทั้งหมด จำนวนข้อถูก จำนวนข้อผิด ร้อยละของระดับความสามารถของผู้สอบ ค่า T-score และผลการประเมิน และรวมทั้งนำเสนอผลการประเมินในรูปของกราฟ ดังแสดงในภาพ 11



ภาพ 11 หน้าจอรายงานผลการทดสอบ

จากภาพ 10 เมื่อทำการทดสอบเสร็จเรียบร้อยแล้ว ผู้เข้ารับการทดสอบสามารถประเมินความพึงพอใจที่มีต่อการใช้ระบบ โดยทำ “แบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด” ดังภาพ 12

ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

**แบบประเมินความพึงพอใจของผู้ใช้งานต่อระบบ  
การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด**

ขอให้นักเรียนเลือกตอบตรงความพึงพอใจของท่าน

ข้อ ที่	คำถาม	ความพึงพอใจ				
		น้อยที่สุด	น้อย	ปานกลาง	มาก	มากที่สุด
1	การทำงานของระบบสะดวก ง่าย ต่อการใช้งาน และ ไม่ซับซ้อน	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	ผู้ใช้งานสามารถเรียนรู้การใช้งานและ การทำงาน ของระบบด้วยตนเอง	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	กระบวนการทำงานของระบบมีความรวดเร็วในการ เรียกใช้บริการ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	รูปแบบของข้อความ ส่วนประกอบต่างๆ บนหน้าจอมี ความเหมาะสม	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	ลำดับการแสดงผลภาพและส่วนประกอบต่างๆ บนหน้า จอ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	การออกแบบการใช้สีที่เหมาะสมต่อการใช้งาน	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	ความรวดเร็วในการรายงานผลการทดสอบ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8	การรายงานผลการทดสอบมีความชัดเจน เข้าใจง่าย	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9	ผลการทดสอบนำไปใช้ในการเตรียมความพร้อม ก่อนการทดสอบภาษาอังกฤษได้	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10	ความรู้สึกรวมที่มีต่อระบบ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

บันทึกข้อมูล

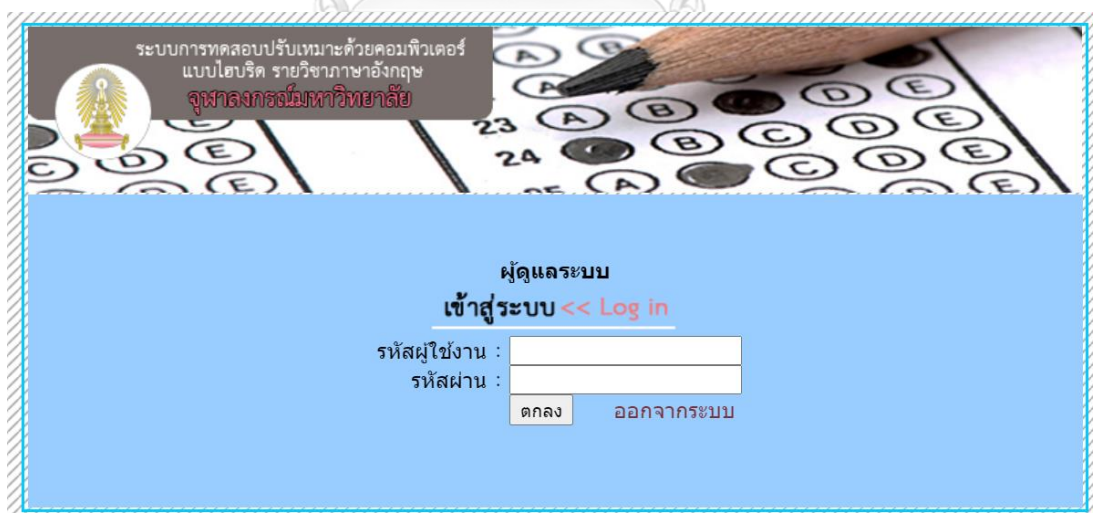
ภาพ 12 แบบประเมินความพึงพอใจต่อระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด



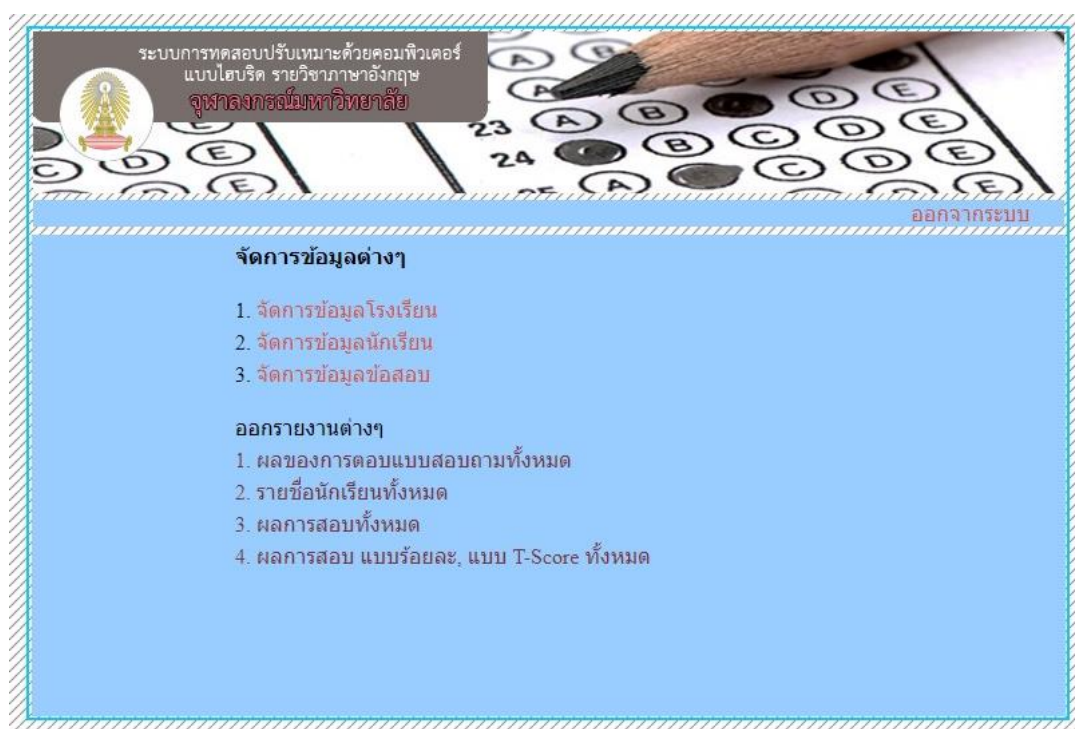
### คู่มือการใช้งานสำหรับผู้จัดการระบบ (admin)

คู่มือการใช้งานสำหรับผู้จัดการระบบ (admin) เพื่อใช้สำหรับจัดการคลังข้อสอบและจัดการประวัติการทำแบบทดสอบ การเข้าสู่ระบบจะต้องใช้รหัสผู้ใช้งานและรหัสผ่านที่ได้รับจากผู้พัฒนาโปรแกรม ดังภาพ 13 เมื่อเข้าสู่ระบบเรียบร้อยแล้ว หน้าจอหลักของผู้จัดการระบบ ประกอบด้วยข้อมูล การจัดการข้อมูลต่าง ๆ ออกรายงานต่าง ๆ ปรากฏรายละเอียดดังภาพ 14 โดยมีรายละเอียดดังนี้

1. การจัดการข้อมูลต่างๆ มีรายละเอียดดังนี้
  - จัดการข้อมูลโรงเรียน : เพิ่ม/แก้ไขข้อมูลโรงเรียน
  - จัดการข้อมูลนักเรียน : จัดการข้อมูลของผู้เข้ารับการทดสอบ
  - จัดการข้อมูลข้อสอบ : เพิ่ม/แก้ไข ข้อสอบในระบบ
2. ออกรายงานต่าง ๆ มีรายละเอียดดังนี้
  - ผลการตอบแบบประเมินทั้งหมด
  - รายชื่อนักเรียนทั้งหมด
  - ผลการสอบทั้งหมด
  - ผลการสอบแบบร้อยละ, T-score ทั้งหมด



ภาพ 13 การเข้าสู่ระบบของผู้จัดการระบบ



ภาพ 14 หน้าจอหลักของผู้จัดการระบบ

#### การจัดการข้อมูล

##### 1. การจัดการข้อมูลโรงเรียน

ในส่วนหน้าจอจัดการข้อมูลต่าง ๆ เมื่อเลือกในส่วนจัดการข้อมูลโรงเรียน จะปรากฏหน้าจอดังภาพ 15 ถ้าผู้จัดการระบบต้องการจะเพิ่มโรงเรียน ให้คลิก “เพิ่มข้อมูลใหม่” ถ้าผู้จัดการระบบจะลบโรงเรียน ให้เลือกโรงเรียน แล้ว คลิก “ลบ” โรงเรียนที่ต้องการจะลบ

ระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

จัดการข้อมูลต่างๆ | ออกรายงานต่างๆ | ออกจากระบบ

ข้อมูลโรงเรียน เพิ่มข้อมูลใหม่

ลำดับที่	ชื่อโรงเรียน	จังหวัด	ลบ
2	จุฬาลงกรณ์มหาวิทยาลัย	กรุงเทพมหานคร	ลบ
1	พิชัยรัตนาคาร	ระนอง	ลบ

ภาพ 15 การจัดการข้อมูลโรงเรียน

## 2. การจัดการข้อสอบ

ส่วนของการจัดการข้อสอบ ผู้จัดการระบบสามารถ เพิ่ม ซึ่งมีรายละเอียด ดังนี้ ประเภทข้อสอบ ระดับข้อสอบ ค่าพารามิเตอร์อำนาจจำแนก (a) ค่าพารามิเตอร์ความยาก (b) ค่าพารามิเตอร์การเดา (c) คำสั่งการทดสอบ การเพิ่มรูปภาพ รายละเอียดประกอบสำหรับทักษะ Reading คำถามของข้อสอบ ตัวเลือก คำตอบ และคำตอบที่ถูกต้อง ได้ตามความต้องการ แล้วคลิก “บันทึกข้อสอบ” ปรากฏดังภาพ 17

จัดการข้อมูลต่างๆ		ออกรายงานต่างๆ		ออกจากระบบ
ข้อมูล - ข้อสอบ				
ข้อสอบ				
ชุดที่ 1				
1	2	3	4	5
ชุดที่ 2				
Low Level				
1	2	3	4	5
Medium Level				
1	2	3	4	5
High Level				
1	2	3	4	5


ภาพ 16 ตัวอย่างชุดข้อสอบสำหรับการทดสอบขั้นแรกและขั้นที่สอง

**ข้อที่ 109**

เลือกประเภทข้อสอบ	ประเภทข้อสอบ	<input checked="" type="radio"/> C=Conversation <input type="radio"/> R=Reading <input type="radio"/> G=Grammar <input type="radio"/> M=Miscellaneous
เลือกระดับของข้อสอบ	ระดับข้อสอบ	<input type="radio"/> L=Low <input type="radio"/> M=Medium <input checked="" type="radio"/> H=High
กรอกค่าพารามิเตอร์อำนาจจำแนก	ค่า a	<input type="text" value="0"/>
กรอกค่าพารามิเตอร์ความยาก	ค่า b	<input type="text" value="0"/>
กรอกค่าพารามิเตอร์การเดา	ค่า c	<input type="text" value="0"/>
กรอกค่าสิ่งของข้อสอบ	ค่าสิ่ง	<input type="text"/>
กรอกรายละเอียดประกอบสำหรับทักษะ Reading	รูปโจทย์ Reading	<input type="text"/>
กรอกคำถามของข้อสอบ	คำถาม	<input type="text"/>
	ตัวเลือก A	<input type="text"/>
กรอกตัวเลือกคำตอบ	ตัวเลือก B	<input type="text"/>
	ตัวเลือก C	<input type="text"/>
	ตัวเลือก D	<input type="text"/>
เลือกคำตอบที่ถูกต้อง	คำตอบ	<input checked="" type="radio"/> A <input type="radio"/> B <input type="radio"/> C <input type="radio"/> D
อัปโหลดรูปภาพ	โจทย์ที่มีรูปประกอบ Upload รูปภาพ (* .jpg เท่านั้น)	<input type="button" value="เลือกไฟล์"/> <input type="button" value="ไม่ได้เลือกไฟล์ใด"/> <input type="button" value="บันทึกรูปภาพ"/>

จุฬาลงกรณ์มหาวิทยาลัย  
**ภาพ 17 การเพิ่มข้อสอบ**  
 CHULALONGKORN UNIVERSITY

ผู้จัดการระบบสามารถแก้ไขข้อสอบจากหน้าจอหลักผู้จัดการระบบ โดย คลิก “จัดการข้อมูลข้อสอบ” ตามภาพ 14 จะปรากฏภาพ 18 เลือกข้อสอบที่จะแก้ไข แล้วคลิก “แก้ไขข้อสอบ” ผู้จัดการระบบสามารถแก้ไขข้อสอบตามต้องการ ดังภาพ 19 แล้วคลิก “บันทึกข้อสอบ”



ระบบการทดสอบปรับคะแนนด้วยคอมพิวเตอร์  
แบบไฮบริด รายวิชาภาษาอังกฤษ  
จุฬาลงกรณ์มหาวิทยาลัย

จัดการข้อมูลต่างๆ | ออกรายงานต่างๆ | ออกจากระบบ

ข้อสอบ

## ชุดที่ 1

### ข้อที่ 1

ประเภทข้อสอบ : C=Conversation  
 ระดับข้อสอบ : M=Medium  
 คำ a : 1.413  
 คำ b : 0.054  
 คำ c : 0.268

---

1. Read each situation and the dialogue carefully and choose the appropriate expression to complete the dialogue.

Situation: Your brother has won a scholarship to study in Canada.  
 You congratulate him by saying, "\_\_\_\_\_"

A. How pretty you look!  
 B. Very well. How did it go?  
 C. That's great! How lucky you are!  
 D. Amazing! I never expected so much.

Answer : C

แก้ไขข้อสอบ

คลิก เพื่อแก้ไขข้อสอบ

**ข้อที่ 1**

ประเภทข้อสอบ ☒ C=Conversation  
☐ R=Reading  
☐ G=Grammar  
☐ M=Miscellaneous

ระดับข้อสอบ ☐ L=Low ☒ M=Medium ☐ H=High

ค่า a 1.413  
 ค่า b 0.054  
 ค่า c 0.258

คำสั่ง Read each situation and the dialogue carefully and choose the appropriate expression to complete the dialogue.

รูปโจทย์ Reading

คำถาม Situation: Your brother has won a scholarship to study in Canada.<br>
 You congratulate him by saying, "\_\_\_\_\_"

ตัวเลือก A How pretty you look!

ตัวเลือก B Very well. How did it go?

ตัวเลือก C That's great! How lucky you are!

ตัวเลือก D Amazing! I never expected so much.

คำตอบ ☐ A ☐ B ☒ C ☐ D

บันทึกข้อสอบ

โจทย์ที่มีรูปประกอบ

Upload รูปภาพ   (\*jpg เท่านั้น)

แก้ไขข้อมูลตามต้องการ

กดบันทึกข้อสอบเมื่อแก้ไขข้อมูลเสร็จเรียบร้อยแล้ว

ภาพ 19 ตัวอย่างการแก้ไขข้อสอบ

### 3. การจัดการสมาชิก

จากหน้าจอหลักภาพ 12 ผู้จัดการระบบต้องการ เพิ่ม ลบ หรือแก้ไขข้อมูลของผู้เข้ารับ การทดสอบ โดย คลิก “จัดการข้อมูลนักเรียน” เลือกสมาชิกที่ต้องการแก้ไข ระบบจะแสดงข้อมูลของสมาชิก ผู้จัดการระบบสามารถแก้ไขข้อมูลสมาชิก เมื่อแก้ไขข้อมูลสมาชิกเสร็จแล้ว คลิก “บันทึกการแก้ไขข้อมูล” ดังภาพ 20



ภาพ 20 ตัวอย่างการแก้ไขข้อมูลของสมาชิก

#### การรายงานผลการทดสอบ

ผู้จัดการระบบสามารถดูรายงานผลการทดสอบของสมาชิกได้ โดยจากหน้าจอหลัก ภาพ 12 เลือกออกรายงานต่างๆ ซึ่งมีรายละเอียด ดังนี้

- ผลการตอบแบบสอบถามทั้งหมด
- รายชื่อนักเรียนทั้งหมด
- ผลการสอบทั้งหมด
- ผลการสอบแบบร้อยละ, T-score ทั้งหมด

ผู้จัดการระบบสามารถเลือกการออกรายงานตามต้องการ โดยการคลิกรายงานที่ต้องการ ข้อมูลจะเป็นไฟล์ Excel



## บรรณานุกรม

- กระทรวงศึกษาธิการ. (2552). หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551. กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.
- กระทรวงศึกษาธิการ. (2553). แผนยุทธศาสตร์ปฏิบัติการเรียนการสอนภาษาอังกฤษ เพื่อเพิ่มขีดความสามารถในการแข่งขันของประเทศ (พ.ศ.2549 - 2553).
- กาโสม หมดเด้น และ นิสากร จารุมณี. (2560). ความสามารถด้านไวยากรณ์ภาษาอังกฤษและความสัมพันธ์ระหว่างความสามารถด้านไวยากรณ์กับทักษะการอ่านและการเขียน. วารสารมนุษยศาสตร์ มหาวิทยาลัยนเรศวร, 14(2).
- ชฎาพร วณะคัมภีร์. (2554). ความสามารถด้านไวยากรณ์ภาษาอังกฤษและความสัมพันธ์ระหว่างความสามารถด้านไวยากรณ์กับทักษะการอ่านและการเขียน. วารสารมนุษยศาสตร์ มหาวิทยาลัยนเรศวร, 14(2).
- ชัยวิชิต เขียรชนะ. (2552). การวิเคราะห์พหุมิติ (Multidimensional Analysis). วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น, 32(4).
- โชติกา ภาชีพล. (2561). ระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทักษะด้านเทคโนโลยีสารสนเทศและการสื่อสารแห่งศตวรรษที่ 21 ของนิสิตนักศึกษาปริญญาตรี : รายงานผลการวิจัย. กรุงเทพฯ: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ญาณิกา ลุนราศรี. (2557). ผลกระทบของการทดสอบทางการศึกษาระดับชาติด้านขั้นพื้นฐานที่มีต่อการเรียนการสอนภาษาอังกฤษในชั้นมัธยมศึกษาปีที่ 3. (ปริญญามหาบัณฑิต), จุฬาลงกรณ์มหาวิทยาลัย, CHULALONGKORN UNIVERSITY
- ญานิศรา มุนินทร์สาคร. (2558). การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการจัดสอบ O-NET ระดับชั้นประถมศึกษาปีที่ 6. (ปริญญามหาบัณฑิต), วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา,
- ณภัทร ชัยมงคล และศิริชัย กาญจนวาสี. (2563). การประยุกต์ใช้โปรแกรมอาร์สำหรับการทดสอบทางการศึกษา. วารสารวิธีวิทยาการวิจัย, 33(1).
- ณภัทร ชัยมงคล. (2558). การพัฒนาระบบการทดสอบแบบปรับเหมาะหลายขั้นตอนด้วยวิธีออนเดอะฟลายที่มีการสะท้อนข้อมูลย้อนกลับในการทดสอบมาตรฐานวิชาชีพของบุคลากรสาขาไอที. (ปริญญามหาบัณฑิต), จุฬาลงกรณ์มหาวิทยาลัย,

- ทัศนศิรินทร์ สว่างบุญ. (2554). การเปรียบเทียบวิธีการคัดเลือกข้อสอบขั้นแรกและลำดับข้อสอบที่มีต่อคุณภาพของการทดสอบปรับเหมาะแบบพหุมิติด้วยคอมพิวเตอร์. (ปริญญาดุขฎฐิบัณทิต), จุฬาลงกรณัณมหาวิทยาลัย,
- ทัศนศิรินทร์ สว่างบุญ ศิริเดช สุชีวะ ศิริชัย กาญจนวาสี และ Muraki, E. (2555). ประสิทธิภาพการวัดของการทดสอบแบบปรับเหมาะแบบพหุมิติด้วยคอมพิวเตอร์. วารสารวิธีวิทยาการวิจัย, 25(1), 27-45.
- พงษ์พิชิต พรหมสิทธิ์ ไพศาล สุวรรณน้อย และ รังสรรค์ มณีเล็ก. (2555). การศึกษาวิธีการประมาณความสามารถ วิธีการคัดเลือกข้อสอบและเกณฑ์ยุติการทดสอบ ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ. วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น, 35(1), 38-47.
- พิมพ์สิริ เจริญนเรศธู. (2549). การเปรียบเทียบค่าความสามารถ ลักษณะของการเปลี่ยนคำตอบ และเวลาที่ใช้ในการทำข้อสอบของผู้สอบ ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์เมื่อเงื่อนไขการทดสอบและระดับความสามารถของผู้สอบแตกต่างกัน. (ปริญญามหาบัณทิต), จุฬาลงกรณัณมหาวิทยาลัย,
- เมษา นวลศรี. (2559). การพัฒนาและตรวจสอบโครงสร้างพหุมิติของความเป็นพลเมืองที่มี ความรับผิดชอบของนักเรียนชั้นมัธยมศึกษาตอนต้น : การประยุกต์ใช้แนวคิดการสร้างแผนที่โครงสร้าง. (ปริญญาดุขฎฐิบัณทิต), จุฬาลงกรณัณมหาวิทยาลัย,
- รังสรรค์ มณีเล็ก. (2540). ผลของตัวแปรบางตัวต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์. (ปริญญาการศึกษาดุขฎฐิบัณทิต), มหาวิทยาลัยศรีนครินทรวิโรฒ,
- ศิริชัย กาญจนวาสี. (2554). ทฤษฎีการประเมิน (พิมพ์ครั้งที่ 7). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณัณมหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2555). ทฤษฎีการทดสอบแนวใหม่ (พิมพ์ครั้งที่ 4). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณัณมหาวิทยาลัย.
- สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน). (2562). สรุปผลการทดสอบทางการศึกษาระดับชาตึ้นพื้นฐาน ระดับมัธยมศึกษาปีที่ 2 ปีการศึกษา 2562. จาก [http://www.newonetestresult.niets.or.th/AnnouncementWeb/PDF/SummaryONETM3\\_2562.pdf](http://www.newonetestresult.niets.or.th/AnnouncementWeb/PDF/SummaryONETM3_2562.pdf)

- สิริลักษณ์ เกษรปทุมานันท์. (2549). การเปรียบเทียบความตรงตามสภาพในการประมาณค่าความสามารถของผู้สอบจากการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ ที่ใช้เกณฑ์การคัดเลือกข้อสอบขั้นแรกอัตราการใช้ข้อสอบซ้ำ และเกณฑ์ยุติการทดสอบที่ต่างกัน. (ปริญญามหาบัณฑิต), จุฬาลงกรณ์มหาวิทยาลัย,
- สุดาร์ตน์ หวลมุกดา. (2550). ประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกภายใต้เงื่อนไขต่างกันโดยใช้วิธีการจำลองข้อมูล. (ปริญญานิพนธ์การศึกษาดุษฎีบัณฑิต), มหาวิทยาลัยศรีนครินทรวิโรฒ,
- อนุสรณ์ เกิดศรี. (2557). ประสิทธิภาพของวิธีการคัดเลือกสองวิธีในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับโมเดลการตอบสนองข้อสอบที่ใช้แบบทดสอบย่อย: การเปรียบเทียบระหว่างวิธีมอนติ คาร์โล ซีเอที และวิธีแบ่งกลุ่มค่าอำนาจจำแนกแบบถ่วงน้ำหนักที่มีการบังคับ. (ปริญญาดุษฎีบัณฑิต), จุฬาลงกรณ์มหาวิทยาลัย,
- Armstrong, R., & Little, J. (2003). *The assembly of multiple form structures*. Paper presented at the 2003 annual meeting of National Council of Measurement in Education, (NCME), Chicago, IL.
- Brossman, B. G., & Guille, R. A. (2014). A Comparison of Multi-Stage and Linear Test Designs for Medium-Size Licensure and Certification Examinations. *Journal of Computerized Adaptive Testing*, 2(2).
- Chang, H. H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20(3), 213-229.
- Chang, H. H., & Ying, Z. (1999). a-Stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, 23(3), 211-222.
- Chang, H. H., & Zhang, J. (2002). Hypergeometric family and item overlap rates in computerized adaptive testing. *Psychometrika*, 67(3), 387-398.  
doi:10.1007/bf02294991
- Chen, S.-Y., Ankenmann, R. D., & Spray, J. A. (2003). The Relationship Between Item Exposure and Test Overlap in Computerized Adaptive Testing. *Journal of Educational Measurement*, 40(2), 129-145. doi:10.1111/j.1745-3984.2003.tb01100

- Chen, S.-Y., & Doong, S. H. (2008). Predicting item exposure parameters in computerized adaptive testing. *Br J Math Stat Psychol*, 61(Pt 1), 75-91.  
doi:10.1348/000711006X129553
- Chen, S. Y., & Macredie, R. D. (2005). The assessment of usability of electronic shopping: A heuristic evaluation. *International Journal of Information Management*, 25, 516-532.
- Cheng, P. E., & Liou, M. (2003). Computerized adaptive testing using the nearest-neighbors criterion. *Applied Psychological Measurement*, 27(3), 204-216.
- Cheng, Y.-Y., Wang, W.-C., & Ho, Y.-H. (2008). Multidimensional rasch analysis of a psychological test with multiple subtests. *Educational and Psychological Measurement*, Sep, 1-20.
- Cheng, Y., Chang, H., & Yi, Q. (2007). Two-phase item selection procedure for flexible content balancing in CAT. *Applied Psychological Measurement*, 31, 467-482.
- Cheng, Y., & Chang, H. H. (2009). The maximum priority index method for severely constrained item selection in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, 62, 369-383.
- Cheng, Y., Patton, J. M., & Shao, C. (2014). a-Stratified Computerized Adaptive Testing in the Presence of Calibration Error. *Educational and Psychological Measurement*, 75(2), 260-283.
- Choi, S. W., Grad, M. W., & Dodd, B. G. (2011). A New Stopping Rule for Computerized Adaptive Testing. *Educational and Psychological Measurement*, 71(1), 37-53.
- Čisar, S. M., Radosav, D., Markoski, B., Pinter, R., & Čisar, P. (2010). Computer Adaptive Testing of Student Knowledge. *Acta Polytechnica Hungarica*, 7(4), 139-152.
- Daniel, L. M., Barbara, G. D., & Brandon, K. V. (2010). A Comparison of Item Selection Techniques for Testlets. *Applied Psychological Measurement*, 34(6), 424-437.
- Davis, L. L., & Dodd, B. G. (2005). *Strategies for controlling item exposure in computerized adaptive testing with partialcredit model*. Retrieved from Austin, TX:
- Deng, H., Ansley, T., & Chang, H.-H. (2010). Stratified and Maximum Information Item Selection Procedures in Computer Adaptive Testing. *Journal of Educational Measurement*, 2(47), 202-226.

- Drasgow, F., Luecht, R. M., & Bennett, R. (2006). *Technology and testing*. Washington, DC: American Council on Education.
- Eggen, T. J. H. M. (2001). *Overexposure and underexposure of item in computerized adaptive testing*. Retrieved from Arnhem:
- Eggen, T. J. H. M., & Straetmans, G. J. J. M. (2000). Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement*, 60(5), 713-734.
- Elissavet, G., Evangelos, T., & Economides, A. A. (2007). Review of Item Exposure Control Strategies for Computerized Adaptive Testing Developed from 1983 to 2005. *The Journal of Technology, Learning, and Assessment*, 5(8).
- Flaugher, R. (2000). Item Pools. In H. Wainer (Ed.), *Computerized adaptive testing: A primer* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Foley, B. P. (2010). *Improving IRT parameter estimates with small sample sizes: Evaluating the efficacy of a new data augmentation technique*. (PhD Dissertation), University of Nebraska, USA.
- Gómez, R. Y., Caballero, D. C., & Sevillano, J.-L. (2014). Heuristic Evaluation on Mobile Interfaces: A New Checklist. *The Scientific World Journal*, 2014, 1-19.
- Guille, R. A., Becker, K. A., Zhu, R. X., Zhang, Y., Song, H., & Sun, L. (2011). *Comparison of asymmetric early termination MST with linear testing*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Hambleton, R. K., & Xing, D. (2006). Optimal and nonoptimal computer-based test designs for making pass-fail decisions. *Applied Measurement in Education*, 19, 221-239.
- He, W., Diao, Q., & Hauser, C. (2014). A Comparison of Four Item-Selection Methods for Severely Constrained CATs. *Educational and Psychological Measurement*, 74(4), 677-696.
- He, W., & Reckase, M. D. (2014). Item Pool Design for an Operational Variable-Length Computerized Adaptive Test. *Educational and Psychological Measurement*, 74(3), 473-494.

- Jian-quan, T., Dan-min, M., Xia, Z., & Jing-jing, G. (2007). An introduction to the computerized adaptive testing. *US-China Education Review*, 4(1), 72-81.
- Jodoin, M. G., Zenisky, A., & Hambleton, R. K. (2006). Comparison of the psychometric properties of several computer-based test designs for credentialing exams with multiple purposes. *Applied Measurement in Education*, 19, 203-220.
- Latu, E., & Chapman, E. (2002). Computerised adaptive testing. *British Journal of Educational Technology*, 33(5), 619-622.
- Lee, H., & Dodd, B. G. (2011). Comparison of Exposure Controls, Item Pool Characteristics, and Population Distributions for CAT Using the Partial Credit Model. *Educational and Psychological Measurement*.
- Leroux, A. J., Lopez, M., Hembry, I., & Dodd, B. G. (2013). A Comparison of Exposure Control Procedures in CATs Using the 3PL Model. *Educational and Psychological Measurement*, 73(5), 857-874.
- Leung, C.-K., Chang, H.-H., & Hau, K.-T. (2002). Item selection in computerized adaptive testing: Improving the a-stratified design with the Sympton-Hetter algorithm. *Applied Psychological Measurement*, 26(4), 376-392.
- Lord, F. M. (1980). *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Luecht, R. L., & Sireci. (2011). *A review of models for computer-based testing*. New York: The College Board.
- McClarty, K. L., Sperling, R. A., & Dodd, B. G. (2006). *A variant of the progressive-restricted item exposure control procedure in computerized adaptive testing systems based on the 3PL and partial credit models*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Molich, R., & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM* 33, 3(March), 338-348.
- Nielsen, J. (1992). Finding usability problems through heuristic evaluation. *Proceedings ACMCHI'92 Conference*(Monterey, CA, May 3-7), 373-380.

- Park, R., Kim, J., Chung, H., & Dodd, B. G. (2014). Enhancing Pool Utilization in Constructing the Multistage Test Using Mixed-Format Tests. *Applied Psychological Measurement*, 38, 268-280.
- Patsula, L. N. (1999). *A comparison of computerized-adaptive testing and multi-stage testing*. (Unpublished doctoral dissertation), University of Massachusetts, Amherst.
- Patton, J. M., Cheng, Y., Yuan, K., & Diao, Q. (2013). The Influence of Item Calibration Error on Variable-Length Computerized Adaptive Testing. *Applied Psychological Measurement*, 37(1), 24-40.
- Reckase, M. D. (2003). *Item pool design for computerized adaptive tests*. Paper presented at the National Council on Measurement in Education, Chicago, IL.
- Reckase, M. D. (2010). Designing item pools to optimize the functioning of a computerized adaptive test. *Psychological Test and Assessment Modeling*, 52(2), 127-141.
- Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 26(2), 144-163.
- Sari, H. I., Sari, H. Y., & Huggins Manley, A. C. (2016). Computer Adaptive Multistage Testing: Practical Issues, Challenges and Principles. *Journal of Measurement and Evaluation in Education and Psychology*, 7(2), 388-406.
- Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61(2), 331-354.
- Stocking, M. L., & Lewis, C. (1998). Controlling Item Exposure Conditional on Ability in Computerized Adaptive Testing. *Journal of Educational and Behavioral Statistics*, 23(1), 57-75.
- Stocking, M. L., & Lewis, C. (2002). Methods of Controlling the Exposure of Items in CAT. In W. J. Linden & G. A. W. Glas (Eds.), *Computerized Adaptive Testing: Theory and Practice* (pp. 163-182). Netherlands: Springer.
- Van der Linden, W. J. (1999). *Adaptive testing with equated number-correct scoring*. (OMD research report; No. 99-02). Enschede: University of Twente, Faculty Educational Science and Technology.

- Wang, S., Lin, H., Chang, H.-H., & Douglas, J. (2016). Hybrid Computerized Adaptive Testing: From Group Sequential Design to Fully Sequential Design. *Journal of Education Measurement*, 53(1), 45-62.
- Way, W. D. (1998). Protecting the integrity of computerized testing item pools. *Educational Measurement: Issues and Practice*, 17(4), 17-27.
- Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 37, 70-84.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21(4), 361-375.
- Zenisky, A. L. (2004). *Evaluating the effects of several multi-stage testing design variables on selected psychometric outcomes for certification and licensure assessment*. University of Massachusetts, Amherst.
- Zheng, Y., & Chang, H.-H. (2015). On-the-Fly Assembled Multistage Adaptive Testing. *Applied Psychological Measurement*, 39(2), 1-15.
- Zheng, Y., Nozawa, Y., Gao, X., & Chang, H.-H. (2012). *Multistage Adaptive Testing for a Large-Scale Classification Test: Design, Heuristic Assembly, and Comparison with Other Testing Modes: ACT*.



## ประวัติผู้เขียน

ชื่อ-สกุล

นางสาวรังสิมาภรณ์ หนูน้อย

วัน เดือน ปี เกิด

21 พฤษภาคม 2529

สถานที่เกิด

ระนอง

วุฒิการศึกษา

สำเร็จการศึกษาระดับปริญญาตรี หลักสูตรศึกษาศาสตรบัณฑิต (เกียรตินิยม อันดับ 2) สาขาวิชาการศึกษาดลอดชีวิต มหาวิทยาลัยศิลปากร ในปีการศึกษา 2550 สำเร็จการศึกษาระดับปริญญาโท หลักสูตรศึกษาศาสตรมหาบัณฑิต สาขาการวิจัยและประเมินผลการศึกษา มหาวิทยาลัยเกษตรศาสตร์ ในปีการศึกษา 2553 และได้เข้าศึกษาต่อระดับปริญญาเอก หลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาจิตวิทยาและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2557



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY