



การทดสอบแบบปรับเหมาะที่มีการสะท้อนข้อมูลย้อนกลับ ในการทดสอบมาตรฐานวิชาชีพไอที

Computerized Adaptive Testing with Reflective Feedback in Information Technology Professional Examination

ณภัทร ชัยมงคล (Nhabhat Chaimongkol)* โชติกา ภาชีผล (Shotiga Pasiphol)*
และ ศิริชัย กาญจนาวาสี (Sirichai Kanjanawasee)*

บทคัดย่อ

การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (Computerized Adaptive Testing: CAT) คือการทดสอบที่ผู้สอบใช้แบบทดสอบที่แตกต่างกัน โดยมีการคัดเลือกข้อสอบที่มีระดับความยากตามความสามารถของผู้สอบ ซึ่งใช้วิธีการทางคอมพิวเตอร์เข้ามาช่วยในการคัดเลือกข้อสอบและประมาณค่าความสามารถของผู้สอบ ในการวิจัยนี้ได้ดำเนินการพัฒนาระบบการทดสอบแบบปรับเหมาะที่มีการสะท้อนข้อมูลย้อนกลับให้กับผู้เข้ารับการทดสอบหลังการทดสอบเสร็จสิ้น โดยข้อสอบที่นำมาใช้พัฒนาระบบเป็นข้อสอบสำหรับการทดสอบมาตรฐานวิชาชีพไอที ระดับ ไอพี (Information Technology Professional Examination: Level IP) ผลการวิจัยพบว่า ระบบที่ได้รับการพัฒนาขึ้นมีคุณภาพตามเกณฑ์การประเมินระบบที่กำหนด และเมื่อนำระบบที่ได้รับการพัฒนาขึ้นไปทดลองใช้กับกลุ่มตัวอย่างในการวิจัยพบว่า ตัวอย่างในการวิจัยมีความพึงพอใจต่อการใช้งานระบบในทุกมิติ

คำสำคัญ: การทดสอบแบบปรับเหมาะ การสะท้อนข้อมูลย้อนกลับ มาตรฐานวิชาชีพไอที

Abstract

Computerize adaptive testing (CAT) is testing which each examinee will get different items. The items are selected up to the level of an examinee's ability. Item selection and ability

estimation are operated by computer. This research aims to develop computerized adaptive testing system with reflective feedback in Information Technology Professional Examination (ITPE) on the IP level. Result from evaluation revealed that the quality of the system was outstanding.

Keywords: Adaptive Testing, Reflection Feedback, Information Technology Professional.

1. บทนำ

การทดสอบโดยทั่วไปจะใช้แบบทดสอบฉบับเดียวกับผู้สอบทุกคน ซึ่งอาจส่งผลกระทบต่อความถูกต้องแม่นยำในการวัด เนื่องจากข้อสอบหนึ่งฉบับมีระดับความยากที่ต่างกันส่งผลให้ผู้สอบบางคนอาจต้องทำข้อสอบที่ยากหรือง่ายเกินไป ดังนั้นการทดสอบในอุดมคติจึงมีความคิดว่า ข้อสอบในแต่ละข้อควรมีความเหมาะสมกับความสามารถในแต่ละบุคคล ซึ่งเรียกว่าการทดสอบแบบปรับเหมาะ (Adaptive Testing or Tailored Testing) แนวคิดดังกล่าวเริ่มนำมาใช้ในการทดสอบเชาว์ปัญญาด้วยแบบสอบของบินेट (Binet) ซึ่งผู้ดำเนินการสอบจะพิจารณาข้อสอบ ข้อแรก หากข้อสอบข้อแรกยากหรือง่ายเกินไป ข้อถัดไปก็จะง่ายลงหรือยากขึ้นเพื่อให้เกิดความเหมาะสมกับระดับเชาว์ปัญญาของผู้สอบ และเมื่อทำข้อสอบไปเรื่อยๆ จะถึงข้อที่ยากมากขึ้นจนผู้สอบไม่สามารถทำข้อสอบได้ถูกเป็นจำนวนหนึ่ง จึงยุติการทดสอบ [1] เมื่อเทคโนโลยีคอมพิวเตอร์เริ่มมีการพัฒนาขึ้น David Weiss

* ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

* Department of Research and Psychology, Faculty of Education, Chulalongkorn University.



จึงได้นำคอมพิวเตอร์เมนเฟรม (Mainframe Computer) มาใช้กับการทดสอบแบบปรับเหมาะ ต่อมาเรียกว่าการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (Computer Adaptive Testing: CAT)

การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (CAT) เป็นวิธีการทดสอบที่รวมทฤษฎีการวัดและเทคโนโลยีเข้าด้วยกันเพื่อให้การวัดมีความแม่นยำ โดยมีองค์ประกอบที่สำคัญ 3 องค์ประกอบ ได้แก่ จุดเริ่มต้นในการทดสอบ การคัดเลือกข้อสอบและการประมาณค่าความสามารถ [2], [3], [4] และเกณฑ์ที่ใช้ในทดสอบดำเนินการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์เสร็จสิ้น จะมีการรายงานผลการทดสอบตามระดับความสามารถของผู้สอบ แต่ยังไม่พบการนำการสะท้อนข้อมูลย้อนกลับ (Reflective Feedback) มาประยุกต์ใช้กับการทดสอบแบบปรับเหมาะ

ด้วยเหตุนี้ผู้วิจัยจึงมีความสนใจพัฒนาระบบการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์สำหรับการทดสอบมาตรฐานวิชาชีพของบุคลากรสาขาไอที (Information Technology Professional Examination: ITPE) ระดับ IT Passport (IP) ซึ่งสามารถสะท้อนข้อมูลย้อนกลับให้กับผู้สอบเพื่อพัฒนาการเรียนรู้ รวมทั้งใช้เป็นแนวทางสำหรับอนาคตในการจัดสอบมาตรฐานวิชาชีพเทคโนโลยีสารสนเทศสำหรับประเทศไทยด้วยการทดสอบแบบปรับเหมาะ และเพื่อให้บุคลากรสาขาไอทีของประเทศมีระดับขีดความสามารถด้านเทคโนโลยีเทียบเท่ามาตรฐานในระดับสากล

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

สำหรับทฤษฎีและงานวิจัยที่เกี่ยวข้องในการวิจัยนี้ประกอบด้วยข้อมูลทั้งสิ้น 4 ตอน ได้แก่ 1) ทฤษฎีการทดสอบแบบปรับเหมาะ 2) การสะท้อนข้อมูลย้อนกลับ 3) การทดสอบมาตรฐานวิชาชีพไอที และ 4) การประเมินคุณภาพระบบโดยมีรายละเอียดดังนี้

2.1 มโนทัศน์เกี่ยวกับการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์

สำหรับการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์เป็นการปรับข้อสอบให้เหมาะกับความสามารถของผู้สอบ โดยแนวทางในการคัดเลือกข้อสอบของแต่ละคนขึ้นอยู่กับผลการตอบข้อสอบข้อที่ผ่านมา ในการทดสอบผู้สอบจะเริ่มทำข้อสอบข้อแรกโดยเป็นข้อที่มีค่าความยากระดับปานกลาง

หลังจากนั้นจะทำการประมาณค่าความสามารถของผู้สอบจากผลการตอบข้อสอบ เพื่อทำการคัดเลือกข้อสอบข้อถัดไปที่มีความยากและอำนาจจำแนกที่เหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งมีหลักคือหากผู้สอบตอบข้อสอบข้อที่ผ่านมาถูกต้อง ข้อถัดไปจะมีค่าระดับความยากที่ยากขึ้น แต่หากผู้สอบตอบข้อสอบผิด ข้อถัดไปจะมีค่าระดับความยากที่ลดลง จนกระทั่งการประมาณค่าความสามารถผู้สอบอยู่ในระดับที่น่าเชื่อถือคืออยู่ในเกณฑ์ของความคลาดเคลื่อนที่กำหนดจึงยุติการทดสอบ [1], [5]

สำหรับการประมาณค่าความสามารถผู้สอบในงานวิจัยนี้ใช้การประมาณค่าร่วมกัน 2 แบบได้แก่ วิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood Estimation: MLE) และวิธีการประมาณค่าด้วยค่าคาดหวังภายหลัง (Expected a Posterior: EAP) ซึ่งวิธีการประมาณค่าด้วยค่าคาดหวังภายหลังจะใช้สำหรับกรณีผู้สอบตอบข้อสอบถูกหรือผิดทุกข้อ หากผู้สอบตอบข้อสอบถูกและผิดผสมกันจะใช้วิธีการประมาณค่าด้วยวิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด

2.2 มโนทัศน์เกี่ยวกับการสะท้อนข้อมูลย้อนกลับ

การให้ข้อมูลย้อนกลับ (Feedback) เป็นมโนทัศน์ของข้อมูลที่ได้รับจากผู้มีส่วนได้ส่วนเสียเช่น ครู เพื่อน หนังสือ ประสบการณ์หรือตนเอง ซึ่งมีความเกี่ยวข้องกับความสามารถหรือความเข้าใจของผู้เรียน นอกจากนี้การให้ข้อมูลย้อนกลับยังเป็นองค์ประกอบหลักของการประเมินความก้าวหน้าระหว่างเรียน (Formative Assessment) และเป็นปัจจัยที่มีความสำคัญในการเรียนรู้สำหรับผู้เรียน วัตถุประสงค์หลักของการให้ข้อมูลย้อนกลับคือการปิดช่องว่างระหว่างความสามารถในปัจจุบันของผู้เรียนกับความสามารถที่ปรารถนา [6], [7]

ดังนั้น การให้ข้อมูลย้อนกลับจึงเป็นองค์ประกอบที่สำคัญในการเรียน และเป็นส่วนหนึ่งในการประเมินความก้าวหน้าระหว่างเรียน (Formative Assessment) เพื่อระบุให้เห็นถึงจุดแข็งและจุดอ่อนของผู้เรียนจนนำไปสู่การทำให้ผู้เรียนสามารถบรรลุผลสัมฤทธิ์ตามศักยภาพของผู้เรียน โดยมีการศึกษานำข้อมูลย้อนกลับมาประยุกต์ใช้ในรูปแบบแฟ้มสะสมงานซึ่งประกอบด้วยข้อมูลย้อนกลับ 2 ส่วน ได้แก่ ข้อมูลย้อนกลับแบบทั่วไป (Feedback) และการให้ข้อมูลย้อนกลับเพื่อการพัฒนา (Feedforward) โดยการให้ข้อมูลย้อนกลับทั่วไป

ได้แก่การให้ข้อความที่สามารถอธิบายผลงานของนักเรียนว่าถูกหรือผิดและดีหรือไม่ดีอย่างไร โดยอิงจากเกณฑ์ในการประเมิน สำหรับการให้ข้อมูลย้อนกลับเพื่อการปรับปรุงได้แก่ การให้ข้อความซึ่งเป็นข้อมูลที่สามารถนำไปใช้เป็นแนวทางในการพัฒนาและปรับปรุงผลงานในอนาคต [8]

2.3 มโนทัศน์เกี่ยวกับการทดสอบมาตรฐานวิชาชีพของบุคลากรสาขาไอที

การทดสอบมาตรฐานวิชาชีพไอทีดำเนินการโดยคณะกรรมการโครงการสอบมาตรฐานวิชาชีพไอที (Information Technology Professionals Examination Council: ITPEC) โดยความร่วมมือของกลุ่มภาคีจำนวน 7 ประเทศ ได้แก่ ญีปุ่น มาเลเซีย ฟิลิปปินส์ เวียดนาม เมียนมา มองโกเลีย และประเทศไทย ซึ่งเป็นการทดสอบวัดระดับความรู้และทักษะด้านเทคโนโลยีสารสนเทศแบบไม่อิงผลิตภัณฑ์ โดยแบ่งออกเป็น 4 ระดับ ได้แก่ ระดับ 1 Information Technology Passport Examination: IP ระดับที่ 2 Fundamental Information Technology Engineer Examination: FE ระดับที่ 3 Applied Information Technology Engineering Examination: AP และระดับที่ 4 Advance Professional Examination ซึ่งเป็นความเชี่ยวชาญเฉพาะ [9] โดยการวิจัยนี้นำข้อสอบในระดับที่ 1 ซึ่งเป็นระดับพื้นฐานมาใช้ในการพัฒนาระบบ โดยเนื้อหาในการทดสอบมาตรฐานวิชาชีพของบุคลากรสาขาไอทีระดับ IT Passport (IP) แบ่งออกเป็น 3 ส่วนได้แก่ ด้านกลยุทธ์ (Strategy) ด้านการจัดการ (Management) และด้านเทคโนโลยี (Technology) [10]

2.4 การประเมินคุณภาพระบบ

ในการพัฒนาระบบสิ่งสำคัญคือการตรวจสอบคุณภาพด้วยการประเมินผลการใช้งานของระบบ ซึ่งมีวิธีการและแนวทางแตกต่างกันออกไป สำหรับการประเมินผลระบบการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์มีการนำแนวคิดการประเมินคุณลักษณะโปรแกรมที่นิยมใช้กันเรียกว่า การประเมินแบบฮิวริสติก (Heuristic Evaluation) และการประเมินจากผู้ใช้ (User Evaluation) มาใช้ร่วมกัน ซึ่งการประเมินจากผู้ใช้ที่นิยมใช้คือการประเมินความพึงพอใจระหว่างอินเตอร์เฟซกับผู้ใช้ (User Satisfaction of the Human-Computer Interface)

การประเมินแบบฮิวริสติกเป็นวิธีการค้นหาปัญหาจากการใช้งานในการออกแบบหน้าจอระบบ (User Interface) โดยมีชุดของนักประเมินจำนวนหนึ่งในการตรวจสอบหน้าจอ

ระบบและตัดสินใจเพื่อยอมรับว่าหน้าจอสามารถใช้งานได้โดยทั่วไป พร้อมให้ความคิดเห็นว่าหน้าจอส่วนใดที่ดีและส่วนใดควรต้องปรับปรุง [11], [12] การประเมินแบบฮิวริสติกเป็นวิธีการที่เร็วที่สุด ประหยัด และเป็นวิธีที่มีประสิทธิภาพในการระบุปัญหาของการใช้งานโดยการประเมินจะใช้กับระบบจริงหรือระบบต้นแบบ (Prototype) ซึ่งควบคุมโดยผู้เชี่ยวชาญ (Expert) ในการประเมินผู้เชี่ยวชาญจะใช้รายการตรวจสอบตามรายการประเมินแบบฮิวริสติก ซึ่งโดยธรรมชาติของการประเมินการใช้งานของหน้าจอระบบไม่สามารถทำได้โดยอัตโนมัติซึ่งยังคงต้องใช้มนุษย์ในการตรวจสอบ [13], [14]

สำหรับรายการประเมินฮิวริสติกแบบดั้งเดิมตามแนวคิดของ Nielsen ประกอบด้วยรายการประเมินทั้งสิ้น 10 รายการ ได้แก่ 1) การมองเห็นสถานะของระบบ (Visibility of System Status) 2) ความสอดคล้องระหว่างระบบกับโลกของความเป็นจริง (Match Between System and the Real World) 3) การควบคุมและความเป็นอิสระของผู้ใช้งาน (User Control and Freedom) 4) ความคงเส้นคงวาและได้มาตรฐาน (Consistency and Standards) 5) การป้องกันความผิดพลาด (Error Prevention) 6) ใช้การจำข้อมูลมากกว่าการเรียกคืนข้อมูล (Recognition Rather Than Recall) 7) ความยืดหยุ่นและประสิทธิภาพของการใช้งาน (Flexibility and Efficiency of Use) 8) รูปแบบสวยงาม และเรียบง่าย (Aesthetic and Minimalist Design) 9) การช่วยเหลือผู้ใช้ระบบอัตโนมัติ วินิจฉัย และกู้คืนจากข้อผิดพลาด (Help Users Recognize, Diagnose, and Recover From Errors) และ 10) การช่วยเหลือและคู่มือการใช้งานระบบ (Help and Documentation) [13], [14]

ในการศึกษาต่อมาของ Gómez [14] ได้เพิ่มเติมองค์ประกอบในการประเมินอีก 3 ประเด็น ได้แก่ สนับสนุนและเพิ่มทักษะของผู้ใช้งาน (Support and Extend the User's Current Skills) ความพึงพอใจและการยอมรับปฏิสัมพันธ์กับผู้ใช้ (Pleasurable and Respectful Interaction With the User) และการป้องกันข้อมูลส่วนบุคคล (Protect the Personal Information) เพื่อให้การประเมินมีคุณภาพมากยิ่งขึ้น ดังนั้นในงานวิจัยครั้งนี้จึงใช้องค์ประกอบในการประเมินทั้งสิ้น 13 องค์ประกอบ นอกจากการประเมินความสามารถของระบบจากผู้เชี่ยวชาญด้วยการประเมินเพื่อช่วยแก้ปัญหาแล้ว การประเมินที่มีความสำคัญอีกประการคือการประเมินความพึงพอใจระหว่างอินเตอร์เฟซกับผู้ใช้ (User Satisfaction of the Human-Computer Interface) เนื่องจากผู้ใช้งานเป็น



กฎเกณฑ์สำคัญในการพัฒนาระบบ หากระบบสามารถตอบสนองการใช้งานของผู้ใช้งานได้เป็นอย่างดีจนทำให้ผู้ใช้งานเกิดความพึงพอใจ จะทำให้ระบบได้รับการขับเคลื่อนจนนำมาซึ่งการใช้งานจริงในองค์กรหรือสังคม แต่หากระบบไม่ทำให้ผู้ใช้งานเกิดความพึงพอใจ อาจทำให้ระบบไม่ได้รับการขับเคลื่อนหรืออาจต้องสูญเสียเวลาเปล่าในการพัฒนา ดังนั้นจึงได้มีผู้พัฒนาเครื่องการประเมินความพึงพอใจระหว่างอินเตอร์เฟซกับผู้ใช้งานขึ้น ซึ่งเป็นเครื่องมือเพื่อประเมินความพึงพอใจของอินเตอร์เฟซกับผู้ใช้งานโดยทั่วไป (Generic User Interface Questionnaire: QUIS) โดยรายละเอียดในการประเมินประกอบด้วยเนื้อหาหลัก 5 ด้านได้แก่

- 1) ภาพรวมปฏิสัมพันธ์ของระบบ (Overall Reaction to the Software)
- 2) หน้าจอ (Screen)
- 3) การใช้คำศัพท์และสารสนเทศระบบ (Terminology and System Information)
- 4) การเรียนรู้ (Learning) และ
- 5) ความสามารถระบบ (System Capabilities) [15], [16]

3. วิธีดำเนินการวิจัย

สำหรับการดำเนินการวิจัยแบ่งออกเป็น 3 ระยะ ได้แก่

- 1) การพัฒนาระบบ
- 2) การทดลองใช้ระบบ และ
- 3) การประเมินคุณภาพระบบ โดยมีรายละเอียดดังนี้

3.1 การพัฒนาระบบ

ศึกษากระบวนการทำงาน (Algorithm) ของการทดสอบแบบ ปรับเหมาะและรูปแบบการรายงานผลที่สะท้อนข้อมูลย้อนกลับ หลังจากนั้นดำเนินการออกแบบและพัฒนาระบบการทดสอบ โดยมีองค์ประกอบได้แก่ หน้าจอการทำงาน ระบบ การจัดการคลังข้อสอบ การจัดชุดข้อสอบ การจัดการผู้เข้ารับการทดสอบและแบบรายงานผลสะท้อนข้อมูลย้อนกลับ

นอกจากนี้ต้องดำเนินการวิเคราะห์ข้อสอบเพื่อจัดทำคลังข้อสอบ โดยนำผลการตอบข้อสอบมาทำการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ ได้แก่ ค่าความยาก (b) อำนาจจำแนก (a) และการเดา (c) ซึ่งในการวิเคราะห์ข้อมูลจะใช้โปรแกรม R ในการวิเคราะห์ เนื่องจากเป็นโปรแกรมที่สามารถวิเคราะห์ข้อสอบที่ให้คะแนนแบบสองค่าได้และเป็นโปรแกรมแบบโอเพ่นซอร์ส (Open Source) ในการคัดเลือกข้อสอบเพื่อจัดทำคลังมีเกณฑ์ในการคัดเลือกข้อสอบตามค่าพารามิเตอร์ของข้อสอบตั้งนี้ ค่าความยากอยู่ในช่วง -4 ถึง 4 ค่าอำนาจจำแนกอยู่ระหว่าง

0 ถึง 0.5 และความน่าจะเป็นในการเดาไม่เกิน 0.50 [1], [17] หากข้อสอบมีค่าพารามิเตอร์ของข้อสอบไม่เป็นไปตามเกณฑ์จะตัดข้อสอบข้อดังกล่าวออกโดยไม่นำมาจัดทำคลังข้อสอบ

สำหรับผลการตอบข้อสอบที่นำมาใช้ในการวิเคราะห์หาค่าพารามิเตอร์ของข้อสอบ ได้รับความอนุเคราะห์จากสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.) ซึ่งเป็นผลการตอบข้อสอบมาตรฐานวิชาชีพไอที (Information Technology Professional Examination: ITPE) ระดับ IT Passport ตั้งแต่ปี 2012-2014 จำนวน 600 ข้อ

3.2 การทดลองใช้ระบบ

กลุ่มตัวอย่างในการทดลองใช้งานระบบมีจำนวน 30 คน โดยผู้วิจัยได้ทำการคัดเลือกผู้เข้าร่วมทดลองใช้ระบบการทดสอบแบบปรับเหมาะที่มีการสะท้อนข้อมูลย้อนกลับในการทดสอบมาตรฐานวิชาชีพของบุคลากรสาขาไอที โดยใช้วิธีการเลือกแบบเฉพาะเจาะจง (Purposive Sampling) ซึ่งแบ่งออกเป็นผู้ที่เคยสอบมาตรฐานวิชาชีพของบุคลากรสาขาไอทีระดับ IP ที่เคยสอบผ่านจำนวน 15 คน และยังไม่ผ่าน จำนวน 15 คน

3.3 การประเมินคุณภาพระบบ

ในการประเมินคุณภาพระบบใช้การประเมิน 2 รูปแบบ ได้แก่ 1) การประเมินระบบด้วยผู้เชี่ยวชาญด้านการทดสอบและด้านเทคโนโลยีสารสนเทศ จำนวน 5 ท่าน ด้วยแบบประเมินระบบแบบฮิวริสติก (Heuristic Evaluation) และการประเมินคุณภาพระบบด้วยความพึงพอใจจากกลุ่มตัวอย่างในการวิจัย โดยใช้แบบประเมินความพึงพอใจผู้ใช้งานของอินเตอร์เฟซระหว่างผู้ใช้งานกับคอมพิวเตอร์ (User Satisfaction of the Human-Computer Interface)

4. ผลการวิจัย

สำหรับผลการวิจัยแบ่งเนื้อหาออกเป็น 3 ตอน ได้แก่

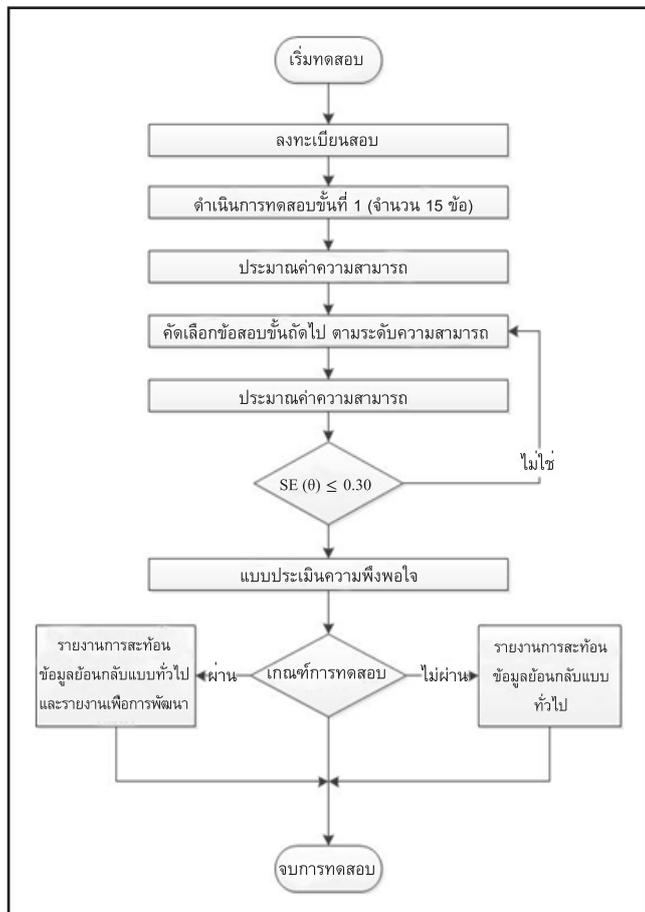
- 1) ผลการพัฒนาระบบ
- 2) ผลการทดลองใช้ระบบ และ
- 3) ผลการประเมินคุณภาพระบบ โดยมีรายละเอียดดังนี้

4.1 ผลการพัฒนาระบบ

สำหรับกระบวนการทำงานของระบบเริ่มจากผู้เข้ารับการทดสอบลงทะเบียนเพื่อทำการทดสอบ หลังจากนั้นดำเนินการทดสอบจำนวน 15 ข้อ ระบบจะทำการประมาณค่าความสามารถผู้สอบ เพื่อนำค่าความสามารถที่ได้ไปใช้ใน

การคัดเลือกข้อสอบอีกจำนวน 15 ข้อ โดยการเลือกข้อสอบในขั้นถัดไปจะพิจารณาจากดัชนีก่อนหน้าสูงสุด (Maximum Priority Index: MPI) ซึ่งมีการควบคุมความครอบคลุมของเนื้อหา (Content Balance) ตามการทดสอบมาตรฐานวิชาชีพไอที เมื่อคัดเลือกข้อสอบได้เรียบร้อยแล้ว ผู้สอบจะดำเนินการทดสอบในระยะที่ 2 อีกจำนวน 15 ข้อ ระบบจะทำการประมาณค่าความสามารถและคำนวณค่าความคลาดเคลื่อนมาตรฐาน (SE(θ)) หากค่าความคลาดเคลื่อนมาตรฐานน้อยกว่าหรือเท่ากับ 0.30 ระบบจะหยุดการทดสอบและนำผู้สอบไปยังหน้าประเมินระบบ แต่หากค่าความคลาดเคลื่อนมาตรฐานมากกว่า 0.30 ระบบจะดำเนินการคัดเลือกข้อสอบมาให้ผู้เข้ารับการทดสอบทำอีก 15 ข้อ ไปเรื่อยๆ แต่หากค่าความคลาดเคลื่อนมาตรฐานยังไม่น้อยกว่า 0.30 เมื่อทำข้อสอบไปแล้ว 7 ชั้น ระบบจะหยุดการทดสอบและนำผู้สอบไปยังหน้าประเมินระบบและรายงานผลการทดสอบ

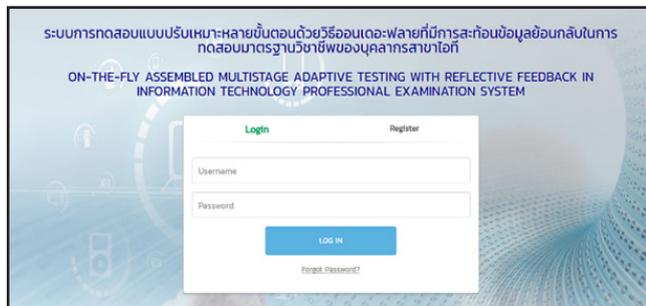
เมื่อผู้เข้ารับการทดสอบมีค่าความคลาดเคลื่อนมาตรฐานน้อยกว่าหรือเท่ากับ 0.30 ระบบจะนำผู้สอบไปยังหน้าประเมินระบบ หลังจากนั้นจะดำเนินการตรวจสอบเงื่อนไข



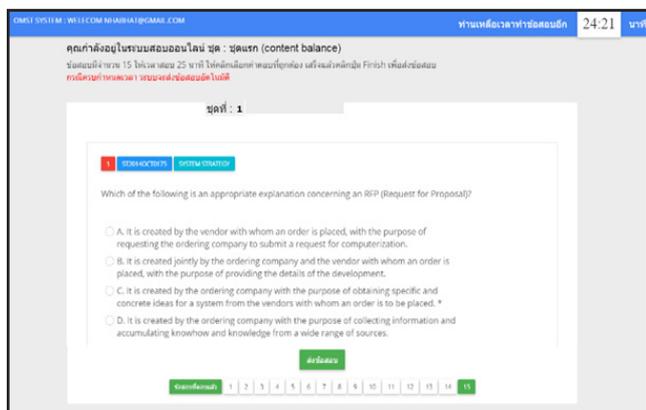
ภาพที่ 1 ผังกระบวนการทดสอบ

ตามเกณฑ์การทดสอบคือ คะแนนสอบระดับด้านของเนื้อหา ต้องมากกว่าหรือเท่ากับร้อยละ 30 และคะแนนรวม 3 ด้าน (กลยุทธ์ การจัดการ และเทคโนโลยี) ต้องมากกว่าหรือเท่ากับ ร้อยละ 55 จึงจะผ่านการทดสอบ หากผู้ทดสอบไม่ผ่านเกณฑ์ รายงานจะแสดงเฉพาะข้อมูลย้อนกลับแบบทั่วไป (Feedback) แต่หากผู้ทดสอบผ่านเกณฑ์ รายงานจะแสดงข้อมูลย้อนกลับแบบทั่วไปพร้อมกับรายงานข้อมูลย้อนกลับเพื่อการปรับปรุง (Feedforward) จึงถือว่าการทดสอบเสร็จสิ้น โดยผังกระบวนการทดสอบแสดงดัง ภาพที่ 1

สำหรับตัวอย่างหน้าจอการทำงานระบบและตัวอย่าง รายงานการสะท้อนข้อมูลย้อนกลับแบบทั่วไป (Feedback) และข้อมูลย้อนกลับเพื่อการพัฒนา (Feedforward) แสดงดังภาพที่ 2-5



ภาพที่ 2 หน้าแรกระบบการทดสอบ



ภาพที่ 3 หน้าการทดสอบ

STAGE	FIELD	CATEGORY	QUANTITY (ITEM)	CORRECT	SCALED SCORES
4	STRATEGY	CORPORATE AND LEGAL AFFAIR	8	6	75.00%
		BUSINESS STRATEGY	8	3	37.50%
		SYSTEM STRATEGY	4	2	50.00%
		TOTAL STRATEGY	20	11	55.00%
MANAGEMENT	DEVELOPMENT TECHNOLOGY	8	4	50.00%	
		PROJECT MANAGEMENT	4	1	25.00%
		SERVICE MANAGEMENT	4	4	100.00%
		TOTAL MANAGEMENT	16	9	56.25%
TECHNOLOGY	BASIC THEORY	8	4	50.00%	
		COMPUTER SYSTEM	8	6	75.00%
		TECHNICAL ELEMENT	8	7	87.50%
		TOTAL TECHNOLOGY	24	17	70.83%
TOTAL			60	37	61.67%

ภาพที่ 4 รายงานผลการสะท้อนข้อมูลย้อนกลับแบบทั่วไป (Feedback)



ภาพที่ 5 รายงานผลการสะท้อนข้อมูลย้อนกลับเพื่อการพัฒนา (Feedforward)

4.2 ผลการทดลองใช้งานระบบ

สำหรับการนำระบบไปทดลองใช้ กับกลุ่มตัวอย่างซึ่งเป็นผู้ที่เคยสอบมาตรฐานวิชาชีพไอที จำนวน 30 ท่าน ผลการทดลองใช้พบว่า ผู้เข้ารับการทดสอบที่เคยสอบผ่านมาตรฐานวิชาชีพไอทีใช้จำนวนขั้นเฉลี่ย (Stage) 5.13 ขั้น ($SD = 1.33$) โดยมีค่าความสามารถ (θ) เฉลี่ยเท่ากับ -0.16 ($SD = 1.26$) ค่าความคลาดเคลื่อนมาตรฐาน (Standard Error) เฉลี่ยเท่ากับ 0.28 ($SD = 0.03$) ซึ่งค่าที่ได้เป็นค่าที่เหมาะสมสำหรับการพัฒนาระบบการทดสอบ โดยรายละเอียดแสดงดังตารางที่ 1

ตารางที่ 1 ผลการทดสอบจากการนำระบบไปทดลองใช้

จำนวน (คน)	Stage		Theta (θ)		SE	
	M	SD	M	SD	M	SD
30	5.13	1.33	-0.16	1.26	0.28	0.03

4.3 ผลการประเมินคุณภาพระบบ

ผลการประเมินระบบจากผู้เชี่ยวชาญด้วยแบบประเมินระบบแบบฮิวริสติก (Heuristic Evaluation) โดยมีประเด็นในการสอบถามทั้งสิ้น 13 ประเด็น โดยใช้มาตรวัด 5 ระดับ (1-5) ผลการประเมินพบว่า ความพึงพอใจและการยอมรับปฏิสัมพันธ์กับผู้ใช้ (Pleasurable and Respectful Interaction With the User) มีค่ามากที่สุด ($M = 4.80$, $SD = 0.40$) และประเด็นที่มีค่าเฉลี่ยน้อยที่สุด คือการใช้การจำข้อมูลมากกว่าการเรียกคืนข้อมูล (Recognition Rather Than Recall) ($M = 4.20$, $SD = 0.75$)

สำหรับการตรวจสอบคุณภาพระบบด้วยแบบประเมินความพึงพอใจระหว่างอินเตอร์เฟซกับผู้ใช้ (User Satisfaction of the Human-Computer Interface) มีประเด็นในการสอบถามทั้งสิ้น 5 ประเด็น ได้แก่ 1) ภาพรวมปฏิสัมพันธ์ของระบบ (Overall) 2) หน้าจอ (Screen) 3) การใช้คำศัพท์และสารสนเทศระบบ (Terminology and System Information) 4) การเรียนรู้

(Learning) และ 5) ความสามารถระบบ (System Capabilities) โดยใช้มาตรวัด 10 ระดับ (0-9) ผลการประเมินพบว่า ด้านภาพรวมปฏิสัมพันธ์ของระบบ ดีมาก ($M = 7.59$, $SD = 0.94$) ด้านหน้าจอประเด็นการใช้งานของเครื่องมือ (Tools) ต่างๆ บนหน้าจอมีค่าสูงสุด หมายความว่า การใช้งานค่อนข้างง่าย ($M = 7.81$, $SD = 0.93$) ด้านการใช้คำศัพท์และสารสนเทศ ประเด็นระบบข้อความบนหน้าจอที่แสดงแก่ผู้ใช้งานเพื่อป้อนข้อมูลมีค่าสูงสุด หมายความว่า ระบบข้อความบนหน้าจอมีความชัดเจน ($M = 7.50$, $SD = 1.02$) ด้านการเรียนรู้ประเด็นผู้ใช้งานสามารถเรียนรู้การใช้งานและการทำงานของระบบมีค่าสูงสุด หมายความว่า ผู้ใช้งานสามารถเรียนรู้และเข้าใจระบบได้อย่างง่าย ($M = 7.33$, $SD = 1.05$) ความสามารถระบบประเด็นความเร็วระบบมีค่าสูงสุด หมายความว่า การทำงานของระบบมีความรวดเร็ว ($M = 7.53$, $SD = 0.99$)

5. สรุปผลการวิจัย

สำหรับการพัฒนาระบบการทดสอบแบบปรับเหมาะที่มีการสะท้อนข้อมูลย้อนกลับในการทดสอบมาตรฐานวิชาชีพไอทีที่ประกอบด้วยขั้นตอนทั้งสิ้น 6 ขั้นตอน ได้แก่ 1) การลงทะเบียนสำหรับผู้เข้ารับการทดสอบ 2) การเข้าสู่ระบบการทดสอบ 3) การดำเนินการทดสอบ 4) การยุติการทดสอบ 5) การประเมินคุณภาพระบบการทดสอบ และ 6) การรายงานผลการทดสอบ และผลการประเมินคุณภาพระบบจากผู้เชี่ยวชาญด้วยแบบประเมินระบบแบบฮิวริสติก (Heuristic Evaluation) พบว่าระบบมีคุณภาพและมีความเหมาะสมในทุกด้าน และการตรวจสอบคุณภาพระบบด้วยแบบประเมินความพึงพอใจระหว่างอินเตอร์เฟซกับผู้ใช้ (User Satisfaction of the Human-Computer Interface) พบว่า ผู้เข้ารับการทดสอบมีความพึงพอใจกับระบบการทดสอบในระดับสูงทุกด้าน ผู้วิจัยหวังเป็นอย่างยิ่งว่าระบบที่พัฒนาขึ้นในครั้งนี้จะ เป็นต้นแบบสำหรับการพัฒนาระบบการทดสอบแบบปรับเหมาะที่มีคุณภาพต่อไปในอนาคต

6. เอกสารอ้างอิง

[1] S. Kanjanawasee. *Modern Test Theory (Ed.4)*. Bangkok: Chulalongkorn University Press, 2012.
 [2] H.-H. Chang. "Understanding Computerized Adaptive Testing: From Robbins-Monro to Lord and Beyond."



- In D. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, pp. 117, California: Sage Publication, Inc, 2004.
- [3] H.-H. Chang. "Psychometrics behind Computerized Adaptive Testing." *Psychometrika*, Vol. 80, No. 1, pp. 1-20, 2015.
- [4] M. Phankokkruad. "Association rules for data mining in item classification algorithm: Web service approach." *Paper presented at the Digital Information and Communication Technology and it's Applications (DICTAP)*, 2012 Second International Conference on, Bangkok, 16-18 May 2012.
- [5] S. M. Čisar, D. Radosav, B. Markoski, R. Pinter, and P. Čisar. "Computer Adaptive Testing of Student Knowledge." *Acta Polytechnica Hungarica*, Vol. 7, No. 4, pp. 139-152, 2010.
- [6] J. Hattie and H. Timperley. "The Power of Feedback." *Review of Educational Research*, Vol. 77, No. 1, pp. 81-112, 2007.
- [7] K. Ludvigsen, R. Krumsvik, and B. Furnes. "Creating formative feedback spaces in large lectures." *Computers and Education*, pp. 88, 2015.
- [8] S. Pasiphol, P. Koraneekij, and P. Sotthayakom. "Development of Electronic Portfolio Model with Reflective Feedback through Assessment of Student Learning Development under The Office of The Basic Education Commission." *Journal of Research Methodology*, Vol. 28, No. 1, pp. 1-25, 2558.
- [9] National Science and Technology Development Agency. *Information Technology Professional Examination*. Available Online at <http://www.nstdaacademy.com/webnsa/index.php/sss>. accessed on 20 April, 2015.
- [10] Information-Technology Promotion Agency Japan. *IT Passport Exam Preparation Book*. Tokyo: Fujitsu Fom Limited, 2010.
- [11] J. Nielsen. "Finding Usability Problems through Heuristic Evaluation." *Paper presented at the ACM CHI'92 Monterey, CA*, 3-7 May 1992.
- [12] J. Nielsen and R. Molich. "Heuristic evaluation of user interfaces." *Paper presented at the ACM CHI'90*, Seattle, WA, 1-5 April, 1990.
- [13] S. Y. Chen and R. D. Macredie. "The assessment of usability of electronic shopping: A heuristic evaluation." *International Journal of Information Management*, Vol. 25, pp. 516-532, 2005.
- [14] R. Y. Gómez, D. C. Caballero, and J.-L. Sevillano. "Heuristic Evaluation on Mobile Interfaces: A New Checklist." *The Scientific World Journal*, pp. 1-19, 2014.
- [15] J. P. Chin, V. A. Diehl, and K. L. Norman. "Development of An Instrument Measuring User Satisfaction of the Human-Computer Interface." *Paper presented at the SIGCHI Conference on Human Factors in Computing Systems*, 1988.
- [16] H. Navas, A. L. Osornio, A. Baum, A. Gomez, D. Luna, and F. G. B. d. Quiros. "Creation and Evaluation of a Terminology Server for the Interactive Coding of Discharge Summaries." *Paper presented at the MEDINFO 2007*, Brisbane, Australia, 2007.
- [17] M. D. Reckase. "Item pool design for computerized adaptive tests." *Paper presented at the The annual meeting of the National Council on Measurement in Education*, Chicago, IL, 2003.