

การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์
สำหรับการจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชน
ตามแนวคิดการเรียนรู้ของเครื่องจักร*

DEVELOPMENT OF COMPUTERIZED ADAPTIVE TESTING PROGRAM
FOR MASSIVE OPEN ONLINE COURSES BASED ON
MACHINE LEARNING APPROACH

วรรณธิดา ยลวิลาศ

Wannatida Yonwilad

ปริญญา เรืองทิพย์

Parinya Ruangtip

กฤษณะ ชินสาร

Krisana Chinnasarn

มหาวิทยาลัยบูรพา

Burapha University, Thailand

E-mail: jscmo555@gmail.com

บทคัดย่อ

บทความวิจัยนี้มีวัตถุประสงค์ 1) เพื่อจัดทำคลังข้อสอบรูปแบบการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์สำหรับการจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชน ตามแนวคิดการเรียนรู้ของเครื่องจักร 2) เพื่อพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะ โดยใช้คอมพิวเตอร์สำหรับการจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชนตามแนวคิดการเรียนรู้ของเครื่องจักร ระเบียบวิธีวิจัยเป็นการวิจัยเชิงสำรวจ การดำเนินการวิจัยแบ่งออกเป็น 3 ขั้นตอน ดังนี้ 1) จัดทำคลังข้อสอบ 2) พัฒนาโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ และ 3) ประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ ด้วยวิธี Black Box Testing กลุ่มตัวอย่างเป็นนักศึกษาระดับปริญญาตรี จำนวน 400 คน ได้มาจากการเลือกตัวอย่างแบบเจาะจง เครื่องมือที่ใช้ในการวิจัย ได้แก่ 1) แบบตรวจสอบความตรงเชิงเนื้อหา 2) โปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ 3) แบบประเมินความเหมาะสมของโปรแกรม สถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ผลการวิจัยปรากฏว่า 1) คลังข้อสอบเป็น

* Received 2 April 2021; Revised 6 September 2021; Accepted 9 September 2021



ข้อสอบแบบเลือกตอบและการคัดเลือกข้อสอบผ่านการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ ได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และค่าการเดาของข้อสอบ (c) ได้ข้อสอบที่ผ่านการคัดเลือกจำนวน 300 ข้อ และมีค่าความเที่ยงทั้งฉบับเท่ากับ .99 2) โปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ มีความเหมาะสมของโปรแกรมอยู่ในระดับมากเป็นที่ยอมรับของผู้เชี่ยวชาญและกลุ่มตัวอย่างที่ทดลองใช้โปรแกรม ซึ่งแสดงให้เห็นว่าโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์สามารถนำไปใช้ในการจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชนได้

คำสำคัญ: โปรแกรมการทดสอบแบบปรับเหมาะ, การจัดการสอนออนไลน์แบบเปิดกว้าง, การเรียนรู้ของเครื่องจักร

Abstract

The objectives of this research article were 1) to construct the item bank in the form of computerized adaptive testing, 2) to develop the computerized adaptive testing program for massive open online courses based on a machine learning approach. The survey research was employed in this study. Research method comprised of three phases such as the creation of an item bank, the development of a computerized adaptive testing program, and the evaluation of the program using the Black Box Testing Method. The sample consisted of 400 undergraduate students using purposive sampling. The research tool used in this study included 1) a Content validity checking form, 2) a Computerized adaptive testing program, and 3) the Suitability assessment form of the program. The mean and standard deviation were employed in statistical data analysis. The results found that: 1) Item Bank is a multiple-choice exam. The selected items passed qualifying analysis based on Item Response Theory (IRT) which showed into three parameters as Discriminant values (a), Difficulty values (b), and guessing values. The qualified exam test contained 300 items with reliability values of 0.99. 2) The computerized adaptive testing program had a high level of suitability. It was accepted by the experts and sample group who tried with the program. Therefore, the computerized adaptive testing program can be applied to massive open online courses effectively.

Keywords: Adaptive Testing Program, Massive Open Online Courses, Machine Learning



บทนำ

การเรียนการสอนแบบออนไลน์ (Online Learning) เข้ามามีบทบาทต่อการเรียนการสอนในระดับอุดมศึกษาที่สอดคล้องกับกระแสโลกาภิวัตน์ และการศึกษาแบบ ไร้พรมแดน รวมทั้งการขยายโอกาสให้กับผู้เรียนเลือกเรียนรู้อะไรก็ได้ทุกที่ทุกเวลา (สำนักงานคณะกรรมการการอุดมศึกษา, 2561) ในสถานการณ์การแพร่ระบาดของโรคติดเชื้อไวรัส โควิดา 2019 (COVID - 19) มีการเรียนรูปแบบออนไลน์ ซึ่งเป็นรูปแบบการเรียนรู้ที่สะดวก และส่งผลทางบวกต่อผู้เรียนที่ลงทะเบียนเรียน (Atiaja, L. & Proenza, R., 2016) หลังจากสถานการณ์การแพร่ระบาดของโรคติดเชื้อไวรัสโควิดา 2019 (COVID - 19) ความต้องการในการเรียนออนไลน์และจำนวนผู้เรียนรูปแบบออนไลน์ได้เพิ่มขึ้น อย่างรวดเร็ว (Chen, T. et al, 2020) ดังนั้นนวัตกรรมใหม่ทางการศึกษาจึงได้มีการจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชน (Massive Open Online Course: MOOCs) ซึ่งเป็นการออกแบบหลักสูตรออนไลน์ที่สามารถเรียนได้ทุกที่ทุกเวลาอย่าง ไม่จำกัด ไม่ต้องเสียค่าใช้จ่าย (Open Courseware) (Purkayastha, N. & Sinha, M. K., 2016) อย่างไรก็ตาม ในการใช้ระบบ MOOCs ยังไม่ตอบวัตถุประสงค์ของการเรียนและผลสัมฤทธิ์ทางการเรียนเท่าที่ควร ปัญหาหลักจาก MOOCs คือ มีการหยุดเรียน (drop-out) ในอัตราที่สูง ผลพบว่า มีอัตราของผู้เรียนที่เรียนสำเร็จหลักสูตรอยู่ในอัตราที่ต่ำอยู่ระหว่าง 5 - 15% (Atiaja, L. & Proenza, R., 2016)

การวิธีการสอนออนไลน์ด้วยรูปแบบ MOOC มีประเด็นที่สำคัญที่ต้องมีการพัฒนาต้องมีปฏิสัมพันธ์กันระหว่างผู้สอนกับผู้เรียน การออกแบบเนื้อหาในการเรียนที่มีความสำคัญต่อการนำไปใช้ในชีวิตประจำวัน (วนิดา พึ่งชมภู และคณะ, 2560) แต่ปัญหาสำคัญอย่างหนึ่งคือ การวัดผลและการประเมินผลการเรียนรู้ที่จำเป็นต้องมีการพัฒนาเพื่อใช้เป็นข้อมูลในการพัฒนาการเรียนรูและการตัดสินใจผลการเรียนผ่านระบบ MOOCs แต่ใช้ระบบข้อสอบชุดเดียวกันและมีจำนวนข้อสอบเท่ากันทุกคน ทำให้ไม่เหมาะสมกับระดับความสามารถของผู้สอบเพราะผู้สอบที่มีความสามารถสูงจะรู้สึกว่าการสอบเสียเวลานานกว่าจะทำข้อสอบเสร็จในขณะที่ผู้สอบที่มีความสามารถต่ำจะรู้สึกว่าการสอบยากเกินไปทำให้ไม่อยากทำข้อสอบต่อไป จึงเป็นปัญหาสำคัญอย่างหนึ่งที่ทำให้การเรียนผ่านระบบ MOOCs ไม่ประสบความสำเร็จเท่าที่ควร ดังนั้นเพื่อให้ผู้เข้าสอบทำแบบทดสอบที่เหมาะสมกับระดับความสามารถ ซึ่งแต่ละคนอาจจะทำข้อสอบไม่เหมือนกันและจำนวนข้อไม่เท่ากัน มีการวัดอย่างเฉลียวฉลาด (Intelligent Measurement: IM) (Bunderson, V. et. al., 1989) เป็นการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (Computerized Adaptive Testing: CAT)

การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ โดยทั่วไปที่ใช้ในการคัดเลือกข้อสอบข้อถัดไปจะใช้ค่าสารสนเทศของฟิชเชอร์สูงสุด (Maximum Fisher Information: MFI) (Delgado - Gómez, D. et. al., 2019) แต่การวัดสารสนเทศนี้ไม่ได้ขึ้นอยู่กับการประมาณค่าพารามิเตอร์ความสามารถของผู้เข้าสอบ (Ability Parameter) การประมาณ



ค่าความสามารถของผู้เข้าสอบ (Ability Estimation) ยังไม่มีความเสถียรมาก การเลือกข้อสอบมีความลำเอียง ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้เข้าสอบมากเป็นต้น (Lu, P. et. al., 2012) เมื่อการประมาณค่าไม่ใช่ค่าที่ถูกปิดหรือ ค่าที่ถูกต้องสำหรับการใช้ค่าสารสนเทศของพีชเซอร์สูงสุดเป็นเกณฑ์ในการวัด การเลือกข้อสอบข้อถัดไปอาจส่งผลให้การคัดเลือกที่ไม่มีประสิทธิภาพ ดังนั้นวิธีการคัดเลือกข้อสอบข้อถัดไปเป็นขั้นตอนที่มีความสำคัญขั้นตอนหนึ่ง เนื่องจากข้อสอบ ที่เลือกมาเป็นข้อสอบข้อถัดไปนั้นส่งผลต่อการประมาณค่าความสามารถของผู้เข้าสอบ (Nurakhmetov, D., 2019) ซึ่งวิธีการเลือกข้อสอบข้อถัดไปอย่างเฉลียวฉลาดผ่านการเรียนรู้ของเครื่องจักร (Machine Learning) โดยใช้ทฤษฎีการตัดสินใจ (Decision Theory) ที่มาประยุกต์การคัดเลือกข้อสอบข้อถัดไปได้ (Ueno, M. & Songmuang, P., 2010)

ทฤษฎีการตัดสินใจ (Decision theory) เป็นทฤษฎีทางเลือกหนึ่งของการนำมาประยุกต์ใช้ในการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (Delgado - Gómez, D. et. al., 2019) โดยวิธีการคัดเลือกข้อสอบข้อถัดไปในการทดสอบแบบปรับให้เหมาะโดยใช้คอมพิวเตอร์ตัดสินใจนี้เป็นการนำวิธีการตัดสินใจความเสี่ยง (Decision making under risk) มาประยุกต์ข้อสอบข้อถัดไป โดยการนำค่าประมาณความสามารถของผู้สอบก่อนที่ตอบข้อสอบที่มีอยู่ในคลังข้อสอบ กรณีที่คาดว่าผู้สอบจะตอบข้อสอบข้อนั้นถูกและกรณีที่คาดว่าจะตอบข้อสอบข้อนั้นผิด จะพิจารณาคัดเลือกข้อสอบข้อถัดไปร่วมกัน โดยไม่นำข้อสอบที่ใช้ไปแล้วมาพิจารณาซ้ำอีก (โสฬส สุขานนท์สวัสดิ์ และคณะ, 2556)

จากทฤษฎีการตัดสินใจข้างต้น สามารถประยุกต์กับการจัดการทดสอบเพื่อนำไปจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชนได้ ผู้วิจัยจึงสนใจพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์สำหรับการจัดการสอนออนไลน์ แบบเปิดกว้างสำหรับมวลชนตามแนวคิดการเรียนรู้ของเครื่องจักรเพื่อนำไปทดสอบความรู้ด้านสถิติ และวิจัยตามความเหมาะสมกับความสามารถของผู้สอบและนำไปประยุกต์ในการจัดการสอนออนไลน์ต่อไปในอนาคต

วัตถุประสงค์ของการวิจัย

1. เพื่อจัดทำคลังข้อสอบรูปแบบการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์สำหรับการจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชนตามแนวคิดการเรียนรู้ของเครื่องจักร
2. เพื่อพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์สำหรับการจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชนตามแนวคิดการเรียนรู้ของเครื่องจักร



วิธีดำเนินการวิจัย

ผู้วิจัยได้ดำเนินการวิจัย ดังนี้

ตอนที่ 1 การจัดทำคลังข้อสอบวิชาโปรแกรมประยุกต์ด้านสถิติและวิจัย มีขั้นตอนการดำเนินการ ดังนี้

1. ศึกษาเอกสารที่เกี่ยวข้องกับหลักการสร้างข้อสอบ หลักการวัดและการประเมินผลการเรียนการสอนรายวิชาโปรแกรมประยุกต์ด้านสถิติและวิจัย สร้างตารางวิเคราะห์จุดประสงค์การเรียนรู้ ตามคำอธิบายรายวิชาโปรแกรมประยุกต์ด้านสถิติและวิจัย

2. สร้างข้อสอบรายวิชาโปรแกรมประยุกต์ด้านสถิติและวิจัยให้สอดคล้องกับตารางวิเคราะห์จุดประสงค์การเรียนรู้ ให้มีข้อสอบที่มากพออย่างน้อย 100 - 120 ข้อ (Weiss, 1988) ดังนั้นผู้วิจัยจึงสร้างข้อสอบ จำแนกเป็นคลังข้อสอบสามโมดูล ได้แก่ 1) สถิติและการนำเสนอข้อมูล จำนวน 125 ข้อ 2) การเลือกใช้สถิติเชิงพรรณนาและการวิเคราะห์ข้อมูล จำนวน 125 ข้อ และ 3) การทดสอบสมมติฐานและการวิเคราะห์สถิติเชิงอนุมาน จำนวน 125 ข้อ รวมทั้งหมดจำนวน 375 ข้อ

3. นำไปตรวจสอบความตรงเชิงเนื้อหาโดยผู้เชี่ยวชาญ จำนวน 5 คน โดยมีเกณฑ์การพิจารณาความเที่ยงตรงเชิงเนื้อหาที่ใช้ได้ ตั้งแต่ 0.8 ขึ้นไป (Davis, L. L., 1992) และแก้ไขปรับปรุงได้ข้อสอบในโมดูลที่ 1) จำนวน 120 ข้อ ข้อสอบโมดูลที่ 2) จำนวน 118 ข้อ และ คลังข้อสอบโมดูลที่ 3) จำนวน 115 ข้อ โดยคลังข้อสอบรวม จำนวน 353 ข้อ

4. นำข้อสอบที่ผ่านการพิจารณาความตรงเชิงเนื้อหาจากผู้เชี่ยวชาญไปทดลองใช้ (Try out) นำไปทดสอบกับนักศึกษาจำนวน 400 คน ที่ได้มาจากการเลือกตัวอย่างแบบเจาะจง (purposive sampling) โดยเลือกมาจากกลุ่มนักศึกษาปริญญาตรี สาขาวิชาคณิตศาสตร์ สาขาวิชาคอมพิวเตอร์ สาขาวิชาภาษาไทย สาขาวิชาภาษาอังกฤษ สาขาวิชารัฐศาสตร์ และสาขาวิชารัฐประศาสนศาสตร์ มหาวิทยาลัยกาฬสินธุ์ ชั้นปีที่ 3 จำนวน 100 คน ชั้นปีที่ 4 จำนวน 200 คน และชั้นปี 5 จำนวน 100 คน ที่เคยเรียนรายวิชาโปรแกรมประยุกต์ทางด้านสถิติและวิจัย รายวิชาการวิจัยทางสังคมศาสตร์ หรือรายวิชาการวิจัยเพื่อพัฒนาการเรียนรู้มาแล้ว

5. การวิเคราะห์คุณภาพของข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ โดยใช้โปรแกรมสำเร็จรูป IRTPRO Version 2.1 โดยคัดเลือกข้อสอบที่ผ่านเกณฑ์การคัดเลือกข้อสอบ (Urry, V. W., 1977) ดังนี้ ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง 2.50 ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ - 2.50 ถึง 2.50 และค่าการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.30 และคัดเลือกข้อสอบที่ผ่านเกณฑ์ จัดทำคลังข้อสอบสามโมดูล โมดูลละ 100 ข้อ รวม 300 ข้อ

6. การวิเคราะห์หาค่าความเที่ยง ทั้งรายด้านและรวมทั้งฉบับ

7. การจัดทำคลังข้อสอบได้นำข้อสอบผ่านตัวเลือก โดยใช้ฐานข้อมูล MySQL



ตอนที่ 2 การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์
สำหรับการจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชนตามแนวคิดการเรียนรู้ของ
เครื่องจักร การดำเนินการตามแนวคิดของวงจรการพัฒนาซอฟต์แวร์ (Software Development Life Cycle: SDLC) ซึ่งประกอบด้วย 5 ขั้นตอน (Ruparelia, N. B., 2010) ดังนี้

1. การวิเคราะห์ความต้องการ (Requirement Analysis) เป็นการกำหนดส่วนประกอบของซอฟต์แวร์ที่เหมาะสมกับการทดสอบแบบปรับเหมาะ ที่ผู้ใช้งานต้องการทั้งวิธีการบันทึกข้อมูลของผู้ทดสอบ จำนวนข้อสอบแต่ละโมดูล วิธีการทดสอบ และการแสดงผลการทดสอบ จะได้รับการพัฒนาเพื่อใช้เป็นแนวทางสำหรับขั้นตอนต่อไป

2. การออกแบบ (Designs) ผู้วิจัยได้ออกแบบซอฟต์แวร์และการออกแบบระบบให้ได้ตามความต้องการของผู้ใช้ ขั้นตอนการออกแบบกำหนดฮาร์ดแวร์และข้อกำหนดของระบบที่จำเป็นตลอดจนสถาปัตยกรรมระบบทั้งหมด ผลลัพธ์จากขั้นตอนนี้จะเป็นอินพุตสำหรับขั้นตอนถัดไป โดยมีการออกแบบวิธีการคัดเลือกข้อสอบตามแนวคิดการตัดสินใจเพื่อให้ได้ข้อสอบข้อแรกที่เหมาะสมกับความสามารถของผู้สอบ ซึ่งจะเขียนในรูปของผังงาน (Flowchart) ที่มีระดับขั้นตอนอย่างชัดเจนตั้งแต่การนำเข้าข้อมูลต่าง ๆ จนถึงการแสดงผลการประเมินจากการทดสอบ

3. การสร้างและการเขียนคำสั่ง (Implementation & Coding) ขั้นตอนนี้การเข้ารหัสจริงจะเสร็จสิ้นและสร้างรหัสตามข้อกำหนดการออกแบบ นี่เป็นขั้นตอนที่สำคัญที่สุดและเป็นระยะที่ยาวที่สุด ซึ่งผู้วิจัยและโปรแกรมเมอร์ร่วมกันเขียนคำสั่งโปรแกรมโดยใช้ภาษา PHP ให้ครอบคลุมข้อกำหนดที่ได้ออกแบบไว้

4. การทดสอบ (Testing) หลังจากพัฒนาคำสั่งโปรแกรมแล้วจะมีการทดสอบเพื่อให้ประเมินว่าเป็นไปตามข้อกำหนดทั้งหมดที่กำหนดไว้ในขั้นแรกหรือไม่ โปรแกรมสามารถใช้งานได้ถูกต้องและรวดเร็ว การประเมินผลโปรแกรมคอมพิวเตอร์ด้วยวิธี Black Box Testing จำแนกเป็นสองส่วน คือ ประเมินความเหมาะสมโดยผู้เชี่ยวชาญ จำนวน 5 คน แบ่งเป็น 5 ด้าน ได้แก่ 1) ด้านความต้องการของผู้ใช้โปรแกรม (Functional Requirement) 2) ด้านการทำงานของโปรแกรม (Functional) 3) ด้านการใช้งาน (Usability) 4) ด้านการรักษาความปลอดภัย (Security) และ 5) ด้านความชัดเจนของคู่มือการใช้โปรแกรม (Program Manual) และประเมินความคิดเห็นโดยผู้ใช้งานซึ่งเป็นนักศึกษาจำนวน 30 คน แบ่งเป็น 2 ด้าน ได้แก่ 1) ด้านการใช้งาน (Usability) และ 2) ด้านความชัดเจนของคู่มือการใช้โปรแกรม (Program Manual)

5. การบำรุงรักษา (Maintenance) เป็นขั้นตอนสุดท้ายที่ซอฟต์แวร์สำเร็จรูปจะถูกส่งมอบให้กับผู้ใช้ ปัญหาที่พบจากการใช้งานจะได้รับการแก้ไขเป็นกรณี ๆ ไป โปรแกรมจะมีการแนะนำให้ความช่วยเหลือแก่ผู้ใช้งาน เพื่อให้สามารถใช้งานได้อย่างต่อเนื่องและ



มีประสิทธิภาพ ได้ดำเนินการเมื่อการประเมินผลระบบเสร็จสิ้น จนกระทั่งสิ้นสุดการดำเนินการวิจัย และซึ่งได้จัดทำคู่มือการใช้งานไว้ในโปรแกรมด้วย

การวิเคราะห์ข้อมูล วิเคราะห์ความเหมาะสม และความคิดเห็นด้วยค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน โดยใช้เกณฑ์ประมาณค่าความคิดเห็น 5 ระดับ ตามวิธีของลิเคิร์ต (Likert) (บุญชม ศรีสะอาด, 2553) ดังนี้ ระดับ 1 2 3 4 และ 5 หมายถึง มีความเหมาะสมน้อยที่สุด น้อย ปานกลาง มาก และมากที่สุด ตามลำดับ และมีเกณฑ์การแปลความหมายค่าเฉลี่ยรายช่วง ดังนี้ ความเหมาะสมในระดับน้อยที่สุด มีค่าตั้งแต่ 1.00 ถึง 1.50 ความเหมาะสมในระดับน้อย มีค่าตั้งแต่ 1.51 ถึง 2.50 ความเหมาะสมในระดับปานกลาง มีค่าตั้งแต่ 2.51 ถึง 3.50 ความเหมาะสมในระดับมาก มีค่าตั้งแต่ 3.51 ถึง 4.50 และความเหมาะสมในระดับมากที่สุด มีค่าตั้งแต่ 4.51 ถึง 5.00

ผลการวิจัย

การจัดทำคลังข้อสอบรายวิชาโปรแกรมประยุกต์ด้านสถิติและวิจัยในการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์สำหรับการจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชนตามแนวคิดการเรียนรู้ของเครื่องจักร มีดังนี้

ตอนที่ 1 ผลการจัดทำคลังข้อสอบรายวิชาโปรแกรมประยุกต์ด้านสถิติและวิจัย

ผลการวิเคราะห์คุณภาพข้อสอบรายวิชาโปรแกรมประยุกต์ด้านสถิติและวิจัย โดยใช้โปรแกรมสำเร็จรูป IRTPRO Version 2.1 มีรายละเอียด ดังนี้

คลังข้อสอบโมดูลที่ 1) สถิติและข้อมูล จำนวนข้อสอบที่คัดเลือก จำนวน 100 ข้อ พบว่า มีค่าความยากของข้อสอบเฉลี่ย เท่ากับ 0.36 ค่าอำนาจจำแนกของข้อสอบเฉลี่ย เท่ากับ 0.99 และค่าการคาดเดาของข้อสอบเฉลี่ย เท่ากับ 0.21

คลังข้อสอบโมดูลที่ 2) การเลือกใช้สถิติเชิงพรรณนาและการวิเคราะห์ข้อมูล จำนวนข้อสอบที่คัดเลือก จำนวน 100 ข้อ พบว่า มีค่าความยากของข้อสอบเฉลี่ยเท่ากับ 0.59 ค่าอำนาจจำแนกของข้อสอบเฉลี่ยเท่ากับ 1.14 และค่าการคาดเดาเฉลี่ย เท่ากับ 0.20

คลังข้อสอบโมดูลที่ 3) การทดสอบสมมติฐานและการวิเคราะห์สถิติเชิงอนุมาน จำนวนข้อสอบที่คัดเลือก จำนวน 100 ข้อ พบว่า มีค่าความยากของข้อสอบเฉลี่ย เท่ากับ 1.03 ค่าอำนาจจำแนกของข้อสอบเฉลี่ย เท่ากับ 2.03 และค่าการคาดเดาของข้อสอบเฉลี่ยเท่ากับ 0.21

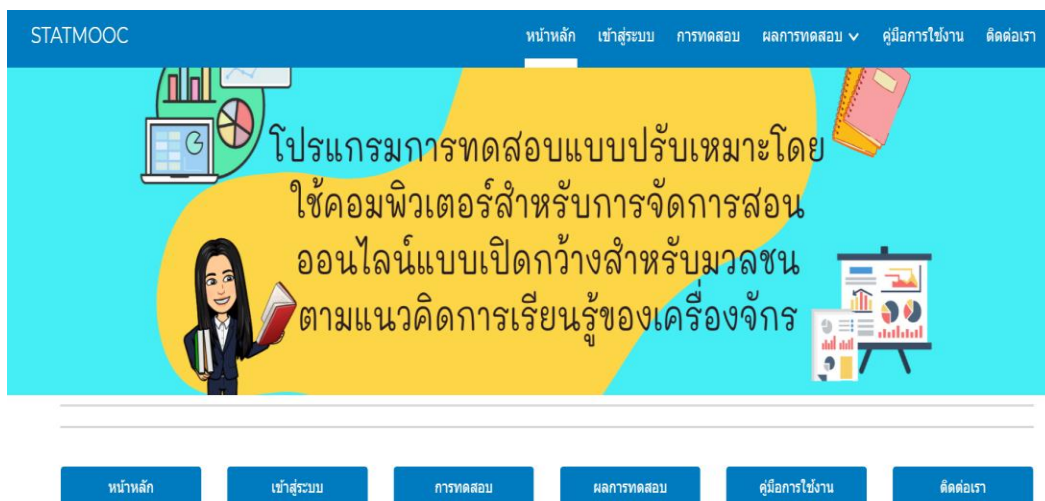
ผลการวิเคราะห์หาความเที่ยง คลังข้อสอบสามโมดูล ประกอบด้วยดังนี้ คลังข้อสอบโมดูลที่ 1) มีค่าความเที่ยงเท่ากับ 0.96 คลังข้อสอบโมดูลที่ 2) มีค่าความเที่ยงเท่ากับ 0.98 และ คลังข้อสอบโมดูลที่ 3) มีค่าความเที่ยง เท่ากับ 0.98 และมีค่าความเที่ยงทั้งฉบับ เท่ากับ 0.99



ตอนที่ 2 ผลการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ สำหรับการจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชนตามแนวความคิดการเรียนรู้ของ เครื่องจักร

การพัฒนาโปรแกรมการทดสอบในรูปแบบเว็บแอปพลิเคชัน สามารถใช้งานได้ที่
<http://www.statmooc.org> โดยมีรายละเอียดดังนี้

1. รูปแบบโปรแกรมการทดสอบ โดยมีหน้าจอหลักในการทำงาน แสดงดังภาพที่ 1



ภาพที่ 1 หน้าจอหลักของโปรแกรม

จากภาพที่ 1 โปรแกรมการทดสอบ ประกอบด้วย เมนูหลัก และส่วนรายละเอียดของโปรแกรมการทดสอบ สำหรับส่วนของเมนูหลัก ประกอบด้วย 6 เมนู ดังนี้ 1) หน้าหลัก เป็นเมนูที่แสดงหน้าหลักของโปรแกรมการทดสอบ 2) เข้าสู่ระบบ เป็นเมนูสำหรับผู้ที่จะเข้าสอบ ทำการลงทะเบียนสอบ 3) การทดสอบ เป็นเมนูสำหรับการทดสอบ ประกอบด้วย 3 โมดูล ได้แก่ 1) สถิติและการนำเสนอข้อมูล 2) การเลือกใช้สถิติเชิงพรรณนาและการวิเคราะห์ข้อมูล 3) การทดสอบสมมติฐานและการวิเคราะห์สถิติเชิงอนุมาน 4) ผลการทดสอบ เป็นเมนูที่แสดงส่วนตรวจสอบผลการทดสอบ 5) คู่มือการใช้งาน เป็นเมนูอำนวยความสะดวกให้กับเข้าสอบ 6) ติดต่อเรา เป็นเมนูที่แสดงข้อมูล แผนที่ ที่อยู่ของข้อมูลของผู้วิจัย

2. ผลการประเมินความเหมาะสม (Black - Box Testing) ในการนำไปใช้ของโปรแกรมสำหรับผู้เชี่ยวชาญ 5 ด้าน พบว่า โดยภาพรวมมีความเหมาะสมอยู่ในระดับมาก ($\bar{X} = 4.19$, S.D. = 0.79) เมื่อพิจารณาเป็นรายด้าน พบว่า ด้านความต้องการของผู้ใช้โปรแกรมมีความเหมาะสมอยู่ในระดับมาก ($\bar{X} = 4.28$, S.D. = 0.74) รองลงมาคือ ด้านการทำงาน of โปรแกรม ($\bar{X} = 4.24$, S.D. = 0.77) ด้านการใช้งาน ($\bar{X} = 4.22$, S.D. = 0.79)



ด้านความชัดเจนของคู่มือการใช้โปรแกรม ($\bar{X}$ = 4.12, S.D. = 0.83) และด้านการรักษาความปลอดภัย ($\bar{X}$ = 4.04, S.D. = 0.84) ตามลำดับ

3. ผลการประเมินความคิดเห็น (Black-Box Testing) ในการนำไปใช้ของโปรแกรมสำหรับผู้ใช้งานจำนวน 400 คน พบว่า โดยภาพรวมมีความเหมาะสมอยู่ในระดับมาก ($\bar{X}$ = 4.17, S.D. = 0.77) เมื่อพิจารณาเป็นรายด้านพบว่า ด้านการใช้งาน มีความเหมาะสมอยู่ในระดับมาก ($\bar{X}$ = 4.18, S.D. = 0.79) รองลงมาคือ ด้านความชัดเจนของคู่มือการใช้โปรแกรม ($\bar{X}$ = 4.16, S.D. = 0.72) ตามลำดับ

อภิปรายผล

การจัดทำคลังข้อสอบรายวิชาโปรแกรมประยุกต์ด้านสถิติและวิจัย เป็นขั้นตอนหนึ่งที่มีความสำคัญเนื่องจากคลังข้อสอบจะส่งผลกระทบต่อการจัดการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ ดังนั้นควรจัดทำคลังข้อสอบให้มีคุณภาพ ซึ่งข้อสอบที่นำมาบรรจุในคลังข้อสอบต้องผ่านการวิเคราะห์คุณภาพของข้อสอบตามหลักเกณฑ์และเงื่อนไขของทฤษฎีการตอบสนองข้อสอบ (ทิพย์ ขำอยู่ และคณะ, 2556) แบบ 3 พารามิเตอร์ มีค่าอำนาจจำแนกของข้อสอบ (a) ระหว่าง 0.50 ถึง + 2.50 ค่าความยากของข้อสอบ (b) ระหว่าง - 2.50 ถึง + 2.50 และค่าการเดาของข้อสอบ (c) ไม่เกิน 0.30 (Urry, V. W., 1977) สำหรับข้อสอบที่นำมาบรรจุในคลังข้อสอบควรมีความละเอียดมากที่สุด (สุชาติดา กรเพชรปานิ และคณะ, 2559) และคลังข้อสอบที่นำมาใช้กับโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ ต้องมีขนาดใหญ่พอสมควร โดยปกติต้องมีข้อสอบในคลังข้อสอบอย่างน้อย 100 - 200 ข้อ ซึ่งคลังข้อสอบที่ได้บรรจุข้อสอบจำนวน 300 ข้อ และครอบคลุมคุณลักษณะและเนื้อหาที่ต้องการวัดความสามารถของผู้เข้าสอบ (Weiss, R. S., 1988)

การคัดเลือกข้อสอบข้อถัดไปเป็นขั้นตอนที่มีความสำคัญขั้นตอนหนึ่ง ในการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ เนื่องจากข้อสอบที่เลือกมาเป็นข้อสอบข้อถัดไปนั้นส่งผลกระทบต่อการประมาณค่าความสามารถของผู้สอบ ดังนั้นควรเลือกข้อสอบที่มีความสอดคล้องกับความสามารถของผู้สอบ ณ ขณะนั้น (Nurakhmetov, D., 2019) การประมาณค่าความสามารถของผู้สอบ โดยใช้หลักการของทฤษฎีการตัดสินใจ (Decision theory) เข้ามาช่วยในการตัดสินใจเลือกข้อสอบข้อถัดไปที่มีความเหมาะสมมากที่สุดและมีโอกาสผิดพลาดน้อยลง สอดคล้องกับงานวิจัยของ Ueno, M. & Songmuang, P. และ โสฬส สุขานนท์สวัสต์ และคณะ ที่วิธีการคัดเลือกข้อสอบข้อถัดไปสามารถลดจำนวนข้อสอบที่ใช้ในการทดสอบลง และใช้เวลาในการทดสอบน้อยกว่าวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ค่าสารสนเทศสูงสุดที่ใกล้เคียงกับค่าความสามารถของผู้สอบ (Ueno, M. & Songmuang, P., 2010); (โสฬส สุขานนท์สวัสต์ และคณะ, 2556)



การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ในครั้งนี้เป็นการพัฒนาโปรแกรมในรูปแบบของ Web Application โดยใช้ภาษา PHP ในการพัฒนาโปรแกรม ใช้ MySQL เป็นระบบจัดการฐานข้อมูล และใช้ภาษา SQL เป็นคำสั่งในการเชื่อมโยงข้อมูลในฐานข้อมูล (โสฬส สุขานนท์สวัสดิ์ และคณะ, 2556) การทดลองใช้งานจากผู้ทดลองใช้ที่เป็นอาจารย์ ผู้ทดลองใช้ที่เป็นนักศึกษา และผู้เชี่ยวชาญแล้ว ซึ่งได้ประเมินโปรแกรมว่า มีระดับความเหมาะสมมาก มีความเหมาะสมที่จะนำไปใช้เป็นเครื่องมือในการจัดการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ จึงทำให้โปรแกรมพัฒนา ขึ้นให้มีประสิทธิภาพ สามารถไปใช้งานจริงได้ แต่การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ที่มีวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ทฤษฎีการตัดสินใจนี้ ยังไม่มีการตรวจสอบอัตราส่วนการใช้ข้อสอบซ้ำในการทดสอบแต่ละครั้ง ดังนั้น ควรมีการวิจัยตรวจสอบอัตราส่วนการใช้ข้อสอบซ้ำในการทดสอบแต่ละครั้งด้วย

องค์ความรู้ใหม่

การพัฒนาคัลลิกข้อสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ (Embretson, S. E., & Reise, S. P., 2000) ที่เหมาะสมกับการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ และได้ออกแบบวิธีการคัดเลือกข้อสอบข้อถัดไปด้วยแนวคิดการเรียนรู้ของเครื่องจักร (Machine Learning) โดยใช้ทฤษฎีการตัดสินใจ (Decision Theory) เข้ามาช่วยในการตัดสินใจ เลือกข้อสอบข้อถัดไปที่มีความเหมาะสมมากที่สุด มีโอกาสผิดพลาดน้อยลง เพื่อวัดความรู้ด้านสถิติ และวิจัยตามความเหมาะสมกับความสามารถของผู้สอบ และสามารถนำไปเป็นส่วนหนึ่งสำหรับการจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชน

สรุป/ข้อเสนอแนะ

วิธีการคัดเลือกข้อสอบข้อถัดไปในการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ โดยใช้ทฤษฎีการตัดสินใจที่สามารถลดจำนวนข้อสอบที่ใช้ในการทดสอบลง และใช้เวลาในการทดสอบน้อยกว่าวิธีการคัดเลือกข้อสอบข้อถัดไป โดยใช้ค่าสารสนเทศสูงสุด ที่ใกล้เคียงกับค่าความสามารถของผู้สอบ เพื่อการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ในการวิจัยนี้ผ่านการตรวจสอบ และทดลองใช้งานจากผู้ทดลองใช้ที่เป็นอาจารย์ ผู้ทดลองใช้ที่เป็นนักศึกษา และผู้เชี่ยวชาญแล้ว ซึ่งได้ประเมินโปรแกรมว่ามีระดับความเหมาะสมมากที่จะนำไปใช้เป็นเครื่องมือในการจัดการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ จึงทำให้โปรแกรมที่พัฒนาขึ้นมีประสิทธิภาพ และทำให้ได้เครื่องมือที่มีคุณภาพสามารถนำไปให้กับผู้ที่สนใจศึกษาทางด้านสถิติและวิจัยและเป็นประโยชน์ในด้านการจัดการทดสอบตามความเหมาะสมกับความสามารถของผู้สอบต่อไป



กิตติกรรมประกาศ

บทความวิจัยนี้เป็นส่วนหนึ่งของงานวิจัยเรื่อง “การจัดการสอนออนไลน์แบบเปิดกว้างสำหรับมวลชนด้วยแนวคิดการเรียนรู้ของเครื่องจักร” ขอกราบขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.ปิยะทิพย์ ประดุจพรม รองศาสตราจารย์ ดร.พูลพงศ์ สุขสว่าง เป็นอย่างสูง ที่กรุณาให้คำแนะนำและขอกราบขอบพระคุณคณาจารย์ประจำหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพาทุกท่านที่ให้ความช่วยเหลืออย่างดี

เอกสารอ้างอิง

- ทิพย์ ขำอยู่ และคณะ. (2556). การวินิจฉัยทักษะการอ่านภาษาอังกฤษโดยประยุกต์โมเดลลำดับชั้นคุณลักษณะและการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์. วารสารวิทยาการวิจัยและวิทยาการปัญญา, 10(2), 55-70.
- บุญชม ศรีสะอาด. (2553). การวิจัยเบื้องต้น (พิมพ์ครั้งที่ 8). กรุงเทพมหานคร: สุวีริยาสาส์น.
- วณิช ฟังชมภู และคณะ. (2560). การพัฒนานวัตกรรมทางการศึกษาพยาบาล: การสอนออนไลน์แบบเปิดสำหรับมหาชนในกระบวนวิชาการพยาบาลผู้สูงอายุ. พยาบาลสาร, 44(พิเศษ), 103-110.
- สำนักงานคณะกรรมการการอุดมศึกษา. (2561). แผนอุดมศึกษาระยะยาว 20 ปี พ.ศ. 2561-2580. กรุงเทพมหานคร: บริษัท 프리ทวอนกราฟฟิค จำกัด.
- สุชาติ กรเพชรปาณี และคณะ. (2559). การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการจัดสอบ O-NET. วิทยาการวิจัยและวิทยาการปัญญา, 14(1), 14-31.
- โสฬส สุขานนท์สวัสดิ์ และคณะ. (2556). การพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ทฤษฎีการตัดสินใจในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์. วิทยาการวิจัยและวิทยาการปัญญา, 10(2), 10(2), 71-85.
- Atiaja, L. & Proenza, R. (2016). Moocs: Problems and challenges in higher education. In *Advances in Education. Teaching & Technology*.
- Bunderson, V. et. al. (1989). Four generations of computer-based testing. *Handbook on educational measurement*, 18(9), 114-146.
- Chen, T. et al. (2020). Analysis of user satisfaction with online education platforms. In China during the COVID- 19 pandemic. Paper presented at the Healthcare.
- Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied nursing research*, 5(4), 194-197.



- Delgado - Gómez, D. et. al. (2019). Computerized adaptive test and decision trees: A unifying approach. *Expert Systems with Applications*, 117(1), 358-366.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Lu, P. et. al. (2012). The study of item selection method in CAT. Retrieved January 8, 2021, from https://link.springer.com/chapter/10.1007/978-3-642-34289-9_45
- Nurakhmetov, D. (2019). Reinforcement Learning Applied to Adaptive Classification Testing. In *Theoretical and Practical Advances in Computer-based Educational Measurement* (pp. 325-336). Springer, Cham.
- Purkayastha, N. & Sinha, M. K. (2016). UNSTOPPABLE STUDY WITH MOOCs DURING COVID 19 PANDEMIC: A STUDY. Library. Retrieved January 8, 2021, from https://digitalcommons.unl.edu/libphilprac/4791/?utm_source=digitalcommons.unl.edu%2Flibphilprac%2F4791&utm_medium=PDF&utm_campaign=PDFCoverPages
- Ruparelia, N. B. (2010). Software development lifecycle models. *ACM SIGSOFT Software Engineering Notes*, 35(3), 8-13.
- Ueno, M. & Songmuang, P. (2010). Computerized adaptive testing based on decision tree. Retrieved January 10, 2021, from https://www.researchgate.net/publication/221425258_Computerized_Adaptive_Testing_Based_on_Decision_Tree
- Urry, V. W. (1977). Tailored testing: A successful application of latent trait theory. *Journal of Educational Measurement*, 14(2), 181-196.
- Weiss, R. S. (1988). Loss and recovery. *Journal of Social Issues*, 44(3), 37-52.