

## การพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอริวริคซ์และการควบคุมการใช้ข้อสอบ สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

### Development of the Next Item Selection Procedure Using Hurwicz Criterion and Item Exposure Control for Computerized Adaptive Testing

ประพล เปรมทองสุข (Prapon Premthongsuk)<sup>\*</sup>

เสรี ชัดเข้ม (Seree Chadcham)<sup>\*\*</sup>

ปิยะทิพย์ ประดุงพรม (Piyathip Pradujprom)<sup>\*\*</sup>

#### บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอริวริคซ์และการควบคุมการใช้ข้อสอบ สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ และศึกษาประสิทธิภาพของวิธีการที่พัฒนาขึ้น โดยเปรียบเทียบประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไป 4 วิธี ได้แก่ 1) วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ค่าสารสนเทศสูงสุดที่ใกล้เคียงกับค่าความสามารถของผู้สอบ (MIC) 2) วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ทฤษฎีการตัดสินใจในสภาวะการณ์ความเสี่ยง (RDM) 3) วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอริวริคซ์ (HC) และ 4) วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอริวริคซ์และการควบคุมการใช้ข้อสอบ (HC-Ex) โดยเปรียบเทียบประสิทธิภาพด้าน 1) การประมาณค่าความสามารถของผู้สอบซึ่งพิจารณาจากค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าความลำเอียงเฉลี่ย 2) ความยาวของแบบทดสอบ และ 3) จำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 ดำเนินการศึกษาในสถานการณ์จำลองแบบมอนติคาร์โล โดยจำลองค่าความสามารถที่แท้จริงของผู้สอบ 10,000 ค่า และคลังข้อสอบ 500 ข้อ วิเคราะห์ข้อมูลด้วยการวิเคราะห์ Brown-Forsythe และ Kruskal-Wallis

ผลการวิจัย ปรากฏว่า 1) วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอริวริคซ์ และการควบคุมการใช้ข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบมีระบบ สามารถใช้คัดเลือกข้อสอบและควบคุมการใช้ข้อสอบได้ 2) การเปรียบเทียบประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบ เมื่อพิจารณาจากค่า RMSE ปรากฏว่า วิธีการ RDM มีประสิทธิภาพสูงสุด แต่เมื่อพิจารณาจากค่าความลำเอียงเฉลี่ย ปรากฏว่า วิธีการ RDM, HC และ HC-Ex มีประสิทธิภาพสูงสุดเทียบเท่ากัน การเปรียบเทียบประสิทธิภาพด้านความยาวของแบบทดสอบ

<sup>\*</sup> นักศึกษาปริญญาเอก หลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา และอาจารย์ประจำคณะวิทยาการจัดการ มหาวิทยาลัยศิลปากร

Ph.D. Candidate, Doctor of Philosophy (Research and Statistics in Cognitive Science) College of Research Methodology and Cognitive Science, Burapha University and Lecturer at Faculty of Management Science, Silpakorn University; E-mail: prapon@ms.su.ac.th

<sup>\*\*</sup> อาจารย์ประจำวิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา

Lecturer at College of Research Methodology and Cognitive Science, Burapha University

ปรากฏว่า วิธีการ HC มีประสิทธิภาพสูงสุดเทียบกับวิธีการ RDM และการเปรียบเทียบประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 ปรากฏว่า วิธีการ MIC มีประสิทธิภาพสูงสุด

**คำสำคัญ:** การคัดเลือกข้อสอบข้อถัดไป/ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์/เกณฑ์ของเฮอริวริชซ์/ การควบคุมการใช้ข้อสอบ

### Abstract

This research aimed at developing the next item selection procedure using Hurwicz Criterion and the item exposure control for computerized adaptive testing which the developed procedure was used to compare the efficiency of 4 next item selection procedures: 1) Maximum Information Criterion (MIC), 2) Risk Decision Making (RDM), 3) Hurwicz Criterion (HC), and 4) Hurwicz Criterion and the item exposure control (HC-Ex) and by comparing the efficacy of 1) the examinee ability estimate considering the Root Mean Square Error (RMSE) and the average bias, 2) the test length, and 3) the number of item with the Item Exposure Rate of more than 0.2 conducted in Monte Carlo Simulation Study by simulating the 10,000 true ability of the examinees and the 500-item bank, and the data was analyzed with Brown-Forsythe and Kruskal-Wallis.

The results showed that 1) the next item selection procedure using Hurwicz Criterion and the item exposure control using System Random Sampling could be used to select the items and control the item exposure. 2) In terms of comparing the efficiency of the examinee ability estimate, considering RMSE, RDM was found to be the most efficient. However, considering the average bias, RDM, HC and HC-Ex represented equally high efficiency. Moreover, comparing the efficiency of the test length, HC was found to be as efficient as that of RDM. For the comparison of the efficiency of the number of item with the Item Exposure Rate of more than 0.2, MIC was the most efficient.

**Keywords:** Next Item Selection/ Computerized Adaptive Testing/ Hurwicz Criterion/ Item Exposure Control

## บทนำ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) เป็นการทดสอบที่มีการจัดข้อสอบให้เหมาะสมกับความสามารถของผู้สอบ ซึ่งพิจารณาจากผลการตอบข้อสอบในข้อที่ผ่านมา กล่าวคือ เมื่อผู้สอบทำข้อสอบข้อแรกหรือชุดเริ่มต้นเสร็จลง จะนำผลการตอบมาวิเคราะห์ระดับความสามารถของผู้สอบ เพื่อใช้คัดเลือกข้อสอบข้อถัดไปที่มีความเหมาะสมกับความสามารถของผู้สอบ โดยการวิเคราะห์นี้ใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) เป็นพื้นฐาน (de Ayala, 2009, pp. 373-381; Wainer, Dorans, Flaugher, Green, & Mislevy, 2000, pp. 105-112)

การดำเนินการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ มีทั้งหมด 5 ขั้นตอน ได้แก่ 1) การสร้างคลังข้อสอบ (Create Item Bank) เป็นขั้นตอนในการจัดทำข้อสอบ เพื่อใช้ในการทดสอบ ซึ่งต้องเป็นไปตามทฤษฎีการตอบสนองข้อสอบ 2) การคัดเลือกข้อสอบข้อแรก (First Item Selection) เป็นการเลือกข้อสอบที่มีความยากระดับปานกลางให้ผู้สอบ แล้วนำผลสอบที่ได้ไปประมาณค่าความสามารถของผู้สอบ เพื่อใช้ในการสอบข้อต่อไป 3) การคัดเลือกข้อสอบข้อถัดไป (Next Item Selection) เป็นการคัดเลือกข้อสอบที่มีความสอดคล้องหรือเหมาะสมกับระดับความสามารถของผู้สอบ 4) การประมาณค่าความสามารถของผู้สอบ (Calculate Possible Ability Level) เป็นการคำนวณค่าประมาณความสามารถของผู้สอบ ณ ขณะนั้น โดยอาศัยผลการตอบข้อสอบในข้อที่ผ่านมา และ 5) เกณฑ์ยุติการทดสอบ (Termination Criterion) เป็นการกำหนดเกณฑ์สิ้นสุดของการทดสอบ โดยการทดสอบจะดำเนินการซ้ำตั้งแต่ขั้นตอนที่ 3 ถึงขั้นตอนที่ 5 จนกระทั่งการทดสอบเป็นไปตามเกณฑ์ยุติการทดสอบที่กำหนดไว้ (Thompson & Weiss, 2011)

ขั้นตอนการคัดเลือกข้อสอบข้อถัดไปเป็นขั้นตอนที่มีความสำคัญ เนื่องจากการคัดเลือกข้อสอบข้อถัดไปที่มีความเหมาะสมกับความสามารถของผู้สอบ จะช่วยลดความยาวของแบบทดสอบ และเพิ่มประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบ (Luecht & Sireci, 2011, pp. 9-10; Lunz, Bergstrom, & Wright, 1992) หลักการทั่วไปของวิธีการคัดเลือกข้อสอบข้อถัดไป จะนำผลการตอบข้อสอบในข้อที่ผ่านมา มาใช้คัดเลือกข้อสอบ คือ ถ้าตอบถูก ผู้สอบจะได้รับข้อสอบข้อถัดไปที่มีค่าความยากของข้อสอบเพิ่มขึ้น ในทางกลับกัน ถ้าตอบผิด ผู้สอบจะได้รับข้อสอบที่มีค่าความยากของข้อสอบลดลง ซึ่งมีวิธีการคัดเลือกข้อสอบข้อถัดไปหลายวิธี แต่ละวิธีมีรายละเอียด จุดเด่น และจุดด้อย แตกต่างกันขึ้นอยู่กับผู้พัฒนา เช่น วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ค่าสารสนเทศสูงสุดที่ใกล้เคียงกับค่าความสามารถของผู้สอบ (Maximum Information Criterion: MIC) เสนอโดย Birnbaum (1968 cited in van der Linden & Glas, 2002, pp. 9-10) หรือวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ทฤษฎีการตัดสินใจในสภาวะการณ์ความเสี่ยงของ โสพส สุขานนท์สวัสดิ์, เสรี ชัดเข้ม และกฤษณะ ชินสาร (2556)

ทฤษฎีการตัดสินใจ (Decision Theory) เป็นทฤษฎีหนึ่งที่ใช้เป็นเครื่องมือในการตัดสินใจ เพื่อให้ผู้ตัดสินใจมีโอกาสตัดสินใจผิดพลาดน้อยลง หรือตัดสินใจได้ถูกต้องมากขึ้น โดยเกณฑ์ของเฮอริวิกซ์ (Hurwicz Criterion) เป็นเกณฑ์หนึ่งในทฤษฎีการตัดสินใจในสภาวะการณ์ที่ไม่แน่นอน (Uncertainly Decision Making) ซึ่งเปิดโอกาสให้ผู้ตัดสินใจระดับของผลตอบแทนที่ผู้ตัดสินใจคิดว่าเหมาะสม โดยระบุจากค่าสัมประสิทธิ์ของการมองโลกในแง่ดี (Coefficient of Optimism:  $\alpha$ ) ซึ่งใช้แทนค่าน้ำหนักในการคำนวณค่าเฉลี่ยถ่วงน้ำหนัก

ระหว่างผลตอบแทนสูงสุดกับผลตอบแทนต่ำสุด แล้วตัดสินใจเลือกทางเลือกที่มีค่าเฉลี่ยถ่วงน้ำหนักสูงสุด (กิตติ ภัคดีวัฒนะกุล และพินิตา พานิชกุล, 2554, หน้า 541-554; Taylor, 2009, pp. 539-572)

จากการศึกษาทฤษฎีการตัดสินใจ รวมทั้งเอกสารและงานวิจัยที่เกี่ยวข้องกับวิธีการคัดเลือกข้อสอบข้อถัดไป ปรากฏว่า เกณฑ์ของเฮอร์วิคซ์สามารถประยุกต์กับการคัดเลือกข้อสอบข้อถัดไปได้ โดยคัดเลือกข้อสอบที่มีค่าเฉลี่ยถ่วงน้ำหนักของค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบ (Standard Error of Estimate: SEE) ต่ำสุด เป็นข้อสอบข้อถัดไป ซึ่งการคำนวณค่าเฉลี่ยถ่วงน้ำหนักของค่า SEE นี้ คำนวณจากค่า SEE ในกรณีที่คาดว่าผู้สอบจะตอบข้อสอบถูก และกรณีที่คาดว่าผู้สอบจะตอบข้อสอบผิด โดยวิธีการนี้ ผู้สอบจะได้รับข้อสอบที่มีความเหมาะสมกับระดับความสามารถของตนเอง เพราะเป็นข้อสอบที่มีค่า SEE ต่ำสุด

วิธีการคัดเลือกข้อสอบข้อถัดไปในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ส่วนใหญ่คัดเลือกข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงมาให้ผู้สอบใช้ในการทดสอบ ทำให้ข้อสอบบางข้อในคลังข้อสอบถูกเลือกใช้ในการทดสอบบ่อยครั้งเกินไป จนเป็นผลให้ผู้สอบรู้ทันความสามารถคาดเดาล่วงหน้าได้ว่าจะได้รับข้อสอบข้อใดในการทดสอบ (Chang & van der Linden, 2003; Wainer et al., 2000, pp. 119-120) ด้วยเหตุนี้ จึงมีการพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปที่มีกระบวนการควบคุมความถี่ในการใช้ข้อสอบให้มีความเหมาะสม เรียกว่า การควบคุมการใช้ข้อสอบ (Item Exposure Control) เพื่อไม่ให้ข้อสอบข้อใดข้อหนึ่งถูกเลือกใช้ในการทดสอบบ่อยครั้งเกินไป เช่น วิธี 5-4-3-2-1 เสนอโดย McBride and Martin (1983) หรือวิธีซิมสันเฮกเตอร์ เสนอโดย Simpson and Hetter (1985)

จากการศึกษาวิธีการควบคุมการใช้ข้อสอบ ปรากฏว่า วิธีการควบคุมการใช้ข้อสอบในกลุ่มวิธีการคัดเลือกอย่างสุ่ม (Randomization Strategies) เป็นกลุ่มวิธีที่มีการดำเนินการไม่ยุ่งยากซับซ้อน เพราะลดโอกาสที่ข้อสอบข้อใดข้อหนึ่งจะถูกเลือกใช้ซ้ำด้วยวิธีการสุ่ม วิธีการในกลุ่มนี้จึงง่ายต่อการนำไปใช้ควบคุมการใช้ข้อสอบ โดยอาศัยแนวคิดนี้ ผู้วิจัยจึงพัฒนาวิธีการควบคุมการใช้ข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบมีระบบ (Systematic Random Sampling) มาควบคุมการใช้ข้อสอบ ซึ่งวิธีการนี้ทำให้ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงมีโอกาสถูกเลือกใช้เท่า ๆ กัน ดังนั้น โอกาสที่ข้อสอบข้อใดข้อหนึ่งจะถูกเลือกใช้ซ้ำบ่อยจึงลดลง (Georgiadou, Triantafillou, & Economides, 2007)

จากการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับวิธีการคัดเลือกข้อสอบข้อถัดไป และวิธีการควบคุมการใช้ข้อสอบ ปรากฏว่า ยังไม่พบงานวิจัยใดที่นำเกณฑ์ของเฮอร์วิคซ์มาใช้ในการคัดเลือกข้อสอบข้อถัดไป และนำวิธีการสุ่มตัวอย่างแบบมีระบบมาใช้ควบคุมการใช้ข้อสอบ ทำให้ผู้วิจัยสนใจพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอร์วิคซ์ และวิธีการควบคุมการใช้ข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบมีระบบ

### วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอร์วิคซ์และการควบคุมการใช้ข้อสอบ สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

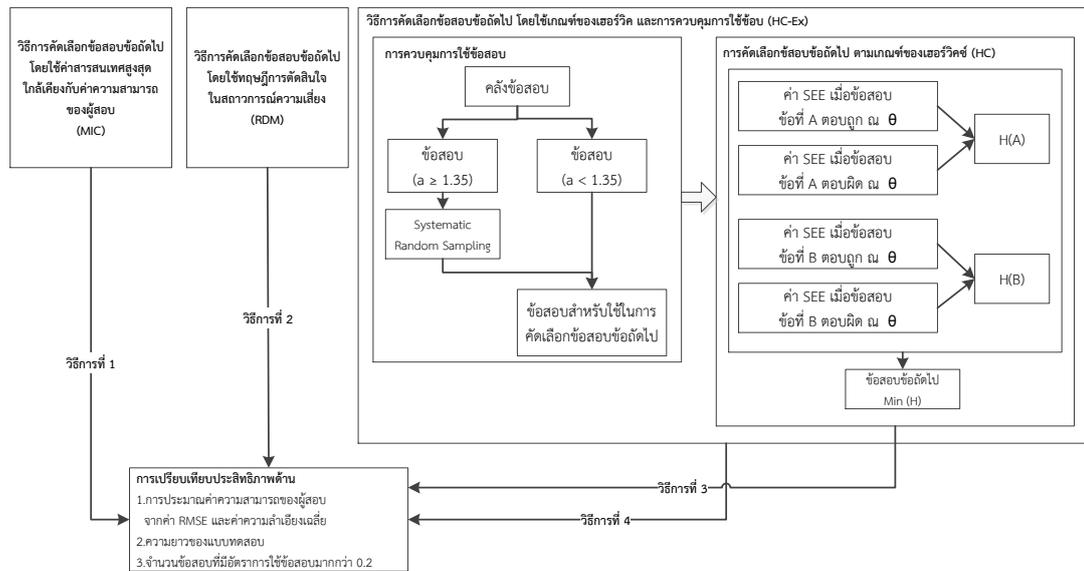
2. เพื่อเปรียบเทียบประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไป 4 วิธี คือ 1) วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ค่าสารสนเทศสูงสุดที่ใกล้เคียงกับค่าความสามารถของผู้สอบ 2) วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ทฤษฎีการตัดสินใจในสภาวะการณ์ความเสี่ยง 3) วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอร์

วิคซ์ และ 4) วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอร์วิคซ์และการควบคุมการใช้ข้อสอบ โดยเปรียบเทียบประสิทธิภาพด้าน 1) การประมาณค่าความสามารถของผู้สอบ 2) ความยาวของแบบทดสอบ และ 3) จำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2

### กรอบแนวคิดการวิจัย

การพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไป โดยประยุกต์ทฤษฎีการตัดสินใจในสภาวะการณ์ที่ไม่แน่นอน ตามเกณฑ์ของเฮอร์วิคซ์ มาคัดเลือกข้อสอบในคลังข้อสอบที่ยังไม่เคยถูกใช้ในการทดสอบ และเป็นข้อสอบที่มี ค่าเฉลี่ยถ่วงน้ำหนักของค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบ (Standard Error of Estimate: SEE) ต่ำสุดมาเป็นข้อสอบข้อถัดไป ซึ่งการคำนวณค่าเฉลี่ยถ่วงน้ำหนักนี้จะอาศัยค่า SEE ทั้งกรณีที่คาดว่าผู้สอบตอบข้อสอบถูก และกรณีที่คาดว่าผู้สอบตอบข้อสอบผิด และใช้ผลต่างระหว่างค่าความ ยากของข้อสอบกับค่าประมาณความสามารถของผู้สอบขณะนั้นมาเป็นค่าถ่วงน้ำหนัก และนำวิธีการสุ่มตัวอย่าง แบบมีระบบ (Systematic Random Sampling) มาควบคุมการใช้ข้อสอบ (Item Exposure Control) โดยข้อสอบที่นำเข้าสู่ขั้นตอนการคัดเลือกข้อสอบข้อถัดไป เป็นข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบต่ำถึง ปานกลาง (มีค่าไม่เกิน 1.35) และข้อสอบบางส่วนที่ถูกสุ่มด้วยวิธีการสุ่มตัวอย่างแบบมีระบบจากข้อสอบที่มีค่า อำนาจจำแนกสูง (มีค่าตั้งแต่ 1.35 ขึ้นไป)

การตรวจสอบประสิทธิภาพของวิธีการที่พัฒนาขึ้น ดำเนินการโดยนำวิธีการคัดเลือกข้อสอบข้อถัดไป โดยใช้ค่าสารสนเทศสูงสุดที่ใกล้เคียงกับค่าความสามารถของผู้สอบ (Maximum Information Criterion: MIC) (Birnbbaum, 1968 cited in van der Linden & Glas, 2002, pp. 9-10) และวิธีการคัดเลือกข้อสอบข้อถัดไป โดยใช้ทฤษฎีการตัดสินใจในสภาวะการณ์ความเสี่ยง (Risk Decision Making: RDM) (โสฬส สุขานนท์สวัสดิ์ และ คณะ, 2556) มาเป็นฐานสำหรับเปรียบเทียบกับวิธีการที่ผู้วิจัยพัฒนาขึ้นทั้งในกรณีที่ไม่มี การควบคุมการใช้ข้อสอบ (Hurwicz Criterion: HC) และกรณีที่มีการควบคุมการใช้ข้อสอบ (Hurwicz Criterion and Exposure Control: HC-Ex) โดยพิจารณาประสิทธิภาพด้าน 1) การประมาณค่าความสามารถของผู้สอบ 2) ความยาวของ แบบทดสอบ และ 3) จำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 กรอบแนวคิดการวิจัยแสดงดังภาพที่ 1



ภาพที่ 1 กรอบแนวคิดการวิจัย

โดยที่

- $\theta$  คือ ค่าความสามารถของผู้สอบ
- $a$  คือ ค่าอำนาจจำแนกของข้อสอบ
- SEE คือ ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ
- $H(A)$  คือ ค่าเฉลี่ยถ่วงน้ำหนักของค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของข้อสอบข้อที่ A
- $Min(H)$  คือ ค่าเฉลี่ยถ่วงน้ำหนักของค่าความคลาดเคลื่อนมาตรฐานของการประมาณความสามารถที่มีค่าต่ำสุด ซึ่งใช้ในการเลือกข้อสอบข้อถัดไป
- RMSE คือ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย

**สมมติฐานการวิจัย**

1. การประมาณค่าความสามารถของผู้สอบ แบ่งได้ 2 ข้อ ดังนี้
  - 1.1 วิธีการ HC มีค่า RMSE น้อยกว่าวิธีการ MIC, RDM และ HC-Ex
  - 1.2 วิธีการ HC มีค่าความลำเอียงเฉลี่ยน้อยกว่าวิธีการ MIC, RDM และ HC-Ex
2. วิธีการ HC มีความยาวของแบบทดสอบน้อยกว่าวิธีการ MIC, RDM และ HC-Ex
3. วิธีการ HC-Ex มีข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 จำนวนน้อยกว่าวิธีการ MIC, RDM

และ HC

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การวิจัยนี้ใช้ทฤษฎีการตอบสนองข้อสอบเป็นพื้นฐาน ซึ่งเป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ภายในบุคคลกับผลการตอบข้อสอบหรือข้อคำถาม โดยเลือกใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (Three-Parameter Logistic Model: 3PL) ซึ่งเป็นโมเดลที่แสดงความน่าจะเป็นของผู้สอบที่มีค่าความสามารถ  $\theta$  สามารถตอบข้อสอบได้ถูกต้อง ประกอบด้วย ค่าความยากของข้อสอบ (Difficulty Parameter:  $b$ ) ค่าอำนาจจำแนกของข้อสอบ (Discrimination Power Parameter:  $a$ ) และค่าการเดาของข้อสอบ (Guess Parameter:  $c$ ) แสดงได้ดังสมการ

$$P(\theta) = c + \frac{1-c}{1+e^{-1.7a(\theta-b)}} \quad (1)$$

ประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไป (Efficiency of the Next Item Selection Procedures) เป็นการวัดความสามารถของวิธีการคัดเลือกข้อสอบข้อถัดไป ซึ่งพิจารณา 3 ด้าน ได้แก่ 1) การประมาณค่าความสามารถของผู้สอบ (Estimative Efficiency) พิจารณาได้จากค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) และค่าความลำเอียงเฉลี่ย (Average Bias) จำนวนตามสมการที่ 2 และ 3 ตามลำดับ 2) ความยาวของแบบทดสอบ (Test Length) และ 3) จำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 (Item Exposure Rate more than 0.2) จำนวนตามสมการที่ 4

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (\hat{\theta}_j - \theta_j)^2} \quad (2)$$

$$Average Bias = \frac{1}{n} \sum_{j=1}^n (\hat{\theta}_j - \theta_j) \quad (3)$$

$$\text{อัตราการใช้ข้อสอบ} = \frac{\text{จำนวนการใช้ข้อสอบ}}{\text{จำนวนผู้สอบทั้งหมด}} \quad (4)$$

เมื่อ  $n$  คือ จำนวนผู้สอบ,  $\hat{\theta}_j$  และ  $\theta_j$  คือ ค่าประมาณและค่าความสามารถที่แท้จริงของผู้สอบ ทฤษฎีการตัดสินใจเป็นหลักเกณฑ์ที่ช่วยตัดสินใจ เพื่อให้ผู้ตัดสินใจมีโอกาสผิดพลาดน้อยลงหรือตัดสินใจได้ถูกต้องมากขึ้น โดยเกณฑ์ของเฮอร์วิคซ์เป็นเกณฑ์หนึ่งในทฤษฎีการตัดสินใจในสภาวะการณ์ที่ไม่แน่นอน คำเนินการโดยในแต่ละทางเลือกจะคำนวณค่าเฉลี่ยถ่วงน้ำหนักระหว่างผลตอบแทนสูงสุดกับผลตอบแทนต่ำสุด แล้วตัดสินใจเลือกทางเลือกที่มีค่าเฉลี่ยถ่วงน้ำหนักสูงสุด

## วิธีดำเนินการวิจัย

**ระยะที่ 1 การพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอร์วิคและการควบคุมการใช้ข้อสอบ** มีขั้นตอนการดำเนินการดังนี้

1. ศึกษาหลักการและวิธีการของทฤษฎีการตัดสินใจ เพื่อค้นหาเกณฑ์การตัดสินใจที่เหมาะสม สำหรับพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไป

2. ศึกษาหลักการและวิธีการควบคุมการใช้ข้อสอบ เพื่อให้ทราบจุดเด่นและจุดด้อยของวิธีการในปัจจุบัน แล้วนำมาใช้เป็นแนวทางพัฒนาวิธีการควบคุมการใช้ข้อสอบ โดยมุ่งพัฒนาวิธีการที่มีแนวคิดและการดำเนินการที่ไม่ยุ่งยากซับซ้อน

3. พัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไป และวิธีการควบคุมการใช้ข้อสอบ โดยการนำผลการวิเคราะห์ที่ได้จากการศึกษาในขั้นตอนที่ 1 และ 2 มาพัฒนาเป็นวิธีการคัดเลือกข้อสอบข้อถัดไป และวิธีการควบคุมการใช้ข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ดังนี้

3.1 การพัฒนาวิธีการควบคุมการใช้ข้อสอบ ได้นำแนวคิดของวิธีการควบคุมการใช้ข้อสอบในกลุ่มวิธีการคัดเลือกอย่างสุ่ม (Randomization Strategies) มาเป็นฐานในการพัฒนา โดยสุ่มข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงด้วยวิธีการสุ่มตัวอย่างแบบมีระบบ ข้อสอบที่สุ่มได้ถูกนำไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป โดยดำเนินการสุ่มข้อสอบใหม่ทุกครั้ง ก่อนเข้าสู่ขั้นตอนการคัดเลือกข้อสอบข้อถัดไป ทำให้ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงถูกหมุนเวียนนำไปใช้ในการคัดเลือกข้อสอบข้อถัดไป เป็นผลให้ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงถูกเลือกใช้ลดลง (โดยทั่วไปข้อสอบที่มีค่าอำนาจจำแนกสูง จะถูกนำไปใช้ในการทดสอบบ่อย)

3.2 การพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไป จากการศึกษาทฤษฎีการตัดสินใจ ปรากฏว่าทฤษฎีการตัดสินใจในสภาวะการณ์ที่ไม่แน่นอน ตามเกณฑ์ของเฮอร์วิคซ์ มีความเหมาะสมในการนำมาประยุกต์กับการคัดเลือกข้อสอบข้อถัดไป เนื่องจาก วิธีการนี้จะคัดเลือกข้อสอบที่ให้ค่าเฉลี่ยถ่วงน้ำหนักของค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถต่ำสุด มาเป็นข้อสอบข้อถัดไป เป็นผลให้ผู้สอบได้รับข้อสอบที่มีความเหมาะสมกับระดับความสามารถของตนเอง

4. ปรับปรุงและแก้ไขวิธีการคัดเลือกข้อสอบข้อถัดไปและวิธีการควบคุมการใช้ข้อสอบ โดยการนำไปทดลองใช้ในสถานการณ์จำลอง แล้วปรับปรุงจนกว่าจะได้วิธีการที่มีความเหมาะสม

## ระยะที่ 2 การเปรียบเทียบประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไป

การเปรียบเทียบประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไปที่พัฒนาขึ้น ได้ศึกษาในสถานการณ์จำลองแบบมอนติคาร์โล (Monte Carlo Simulation Study) ซึ่งจำลองสถานการณ์การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ที่ใช้วิธีการคัดเลือกข้อสอบข้อถัดไปแตกต่างกัน 4 วิธี ดังนี้

1. ตัวแปรที่ใช้ในการศึกษาวิจัย แบ่งได้ 2 ส่วน ดังนี้

- 1) ตัวแปรต้น คือ วิธีการคัดเลือกข้อสอบข้อถัดไป มี 4 วิธี ได้แก่ 1) วิธีการ MIC 2) วิธีการ RDM 3) วิธีการ HC และ 4) วิธีการ HC-Ex

2) ตัวแปรตาม คือ ประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไป มี 3 ด้าน ได้แก่ 1) การประมาณค่าความสามารถของผู้สอบ พิจารณาจากค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) และค่าความลำเอียงเฉลี่ย (Average Bias) 2) ความยาวของแบบทดสอบ และ 3) จำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2

## 2. การจำลองข้อมูลที่ใช้ศึกษาในสถานการณ์จำลองด้วยโปรแกรม WinGen3 ดังนี้

1) ค่าความสามารถที่แท้จริงของผู้สอบ (True Ability:  $\theta$ ) จำลองโดยสุ่มข้อมูลจากเลขสุ่มที่มีการแจกแจงแบบปกติ  $N(0, 1)$  จำนวน 10,000 ค่า แล้วแบ่งเป็น 10 ชุด ๆ ละ 1,000 ค่า

2) คลังข้อสอบ (Item Bank) ขนาด 500 ข้อ แต่ละข้อประกอบด้วยค่าพารามิเตอร์ของข้อสอบ 3 ค่า ดังนี้ ก) ค่าความยากของข้อสอบ จำลองโดยสุ่มข้อมูลจากเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม มีค่าอยู่ระหว่าง -2.50 ถึง 2.50 ข) ค่าอำนาจจำแนกของข้อสอบ จำลองโดยสุ่มข้อมูลจากเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม มีค่าอยู่ระหว่าง 0.50 ถึง 2.50 และ ค) ค่าการเดาของข้อสอบ จำลองโดยสุ่มข้อมูลจากเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม มีค่าไม่เกิน 0.30

3) ผลการตอบข้อสอบของผู้สอบ ดำเนินการโดยสุ่มเลขที่มีการแจกแจงแบบยูนิฟอร์ม  $U(0, 1)$  มา 1 ค่า แล้วนำมาเปรียบเทียบกับค่าความน่าจะเป็นของผู้สอบที่มีค่าความสามารถ  $\theta$  ตอบข้อสอบถูกต้อง (จากสมการที่ 1) หากค่าความน่าจะเป็นมีค่ามากกว่าหรือเท่ากับเลขสุ่ม ให้กำหนดผลการตอบเป็น 1 (ตอบถูก) แต่หากค่าความน่าจะเป็นมีค่าน้อยกว่าเลขสุ่ม ให้กำหนดผลการตอบเป็น 0 (ตอบผิด) (Thompson & Weiss, 2011)

3. การเขียนโปรแกรมเพื่อจำลองสถานการณ์การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ที่มีวิธีการคัดเลือกข้อสอบข้อถัดไปแตกต่างกัน 4 วิธี ซึ่งทุกวิธีถูกทำซ้ำทั้งหมด 10 รอบ แต่ละรอบจะใช้คลังข้อสอบขนาด 500 ข้อ ชุดเดียวกัน แต่ใช้ค่าความสามารถที่แท้จริงของผู้สอบที่ละชุด ชุดละ 1,000 ค่า โดยกำหนดรายละเอียดขั้นตอนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ดังนี้

ขั้นที่ 1 การคัดเลือกข้อสอบข้อแรก โดยการสุ่มเลือกข้อสอบในคลังข้อสอบที่มีค่าความยากของข้อสอบตั้งแต่ -1.00 ถึง 1.00

ขั้นที่ 2 วิธีการคัดเลือกข้อสอบข้อถัดไป ใช้วิธีการใดวิธีการหนึ่งใน 4 วิธี ได้แก่ 1) วิธีการ MIC 2) วิธีการ RMD 3) วิธีการ HC และ 4) วิธีการ HC-Ex ซึ่งขึ้นอยู่กับวิธีการจำลองสถานการณ์ขณะนั้นว่า กำลังวัดประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไปด้วยวิธีการใด

ขั้นที่ 3 การประมาณค่าความสามารถของผู้สอบ เลือกใช้วิธีการประมาณค่าของเบส์แบบปรับใหม่ (Bayesian Updating) ศึกษารายละเอียดได้จาก Owen (1975)

ขั้นที่ 4 เกณฑ์ยุติการทดสอบ เมื่อค่า SEE มีค่าต่ำกว่า 0.3 หรือทำข้อสอบครบ 30 ข้อ

### การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลจากการศึกษาในสถานการณ์จำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ที่มีวิธีการคัดเลือกข้อสอบข้อถัดไปแตกต่างกัน 4 วิธี โดยแต่ละวิธีถูกทำซ้ำทั้งหมด 10 รอบ ผลการศึกษาในแต่ละรอบจะได้รับ 1) ค่าประมาณความสามารถของผู้สอบ เพื่อนำไปคำนวณค่า RMSE และค่าความลำเอียงเฉลี่ย 2) ความยาวของแบบทดสอบ และ 3) จำนวนการใช้ข้อสอบแต่ละข้อในคลังข้อสอบ เพื่อนำไปคำนวณค่าอัตราการใช้ข้อสอบ

### การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูล สำหรับเปรียบเทียบประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไป แบ่งได้ 3 ส่วน ได้แก่

1) การเปรียบเทียบประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบ โดยพิจารณาจากค่า RMSE และค่าความลำเอียงเฉลี่ย ใช้การวิเคราะห์ Brown-Forsythe และหากผลการวิเคราะห์มีนัยสำคัญทางสถิติ จะเปรียบเทียบพหุคูณ (Multiple Comparison) ด้วยการทดสอบของแทมเฮน (Tamhane's Test)

2) การเปรียบเทียบประสิทธิภาพด้านความยาวของแบบทดสอบ ใช้การทดสอบ Kruskal-Wallis และหากผลการวิเคราะห์มีนัยสำคัญทางสถิติ จะเปรียบเทียบค่าเฉลี่ยของอันดับ (Average Rank) เป็นรายคู่ ด้วยการทดสอบ Dunn-Bonferroni

3) การเปรียบเทียบประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 ใช้การทดสอบ Kruskal-Wallis และหากผลการวิเคราะห์มีนัยสำคัญทางสถิติ จะเปรียบเทียบค่าเฉลี่ยของอันดับเป็นรายคู่ ด้วยการทดสอบ Dunn-Bonferroni

### ผลการวิจัย

1. ผลการพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอร์วิคซ์และการควบคุมการใช้ข้อสอบ มีดังนี้

1.1 ผลการพัฒนาวิธีการควบคุมการใช้ข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบมีระบบ มีขั้นตอนการดำเนินการดังนี้

ขั้นที่ 1 แบ่งข้อสอบในคลังข้อสอบออกเป็น 2 กลุ่ม คือ ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบต่ำถึงปานกลาง ซึ่งมีค่าน้อยกว่า 1.35 (Baker, 2001, p. 34) กำหนดให้เป็นกลุ่ม A และข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง ซึ่งมีค่าตั้งแต่ 1.35 ขึ้นไป กำหนดให้เป็นกลุ่ม B

ขั้นที่ 2 สุ่มเลือกข้อสอบในกลุ่ม B ด้วยวิธีการสุ่มตัวอย่างแบบมีระบบ ซึ่งมีขั้นตอนการสุ่มดังนี้

ขั้นที่ 2.1 กำหนดจำนวนข้อสอบที่ต้องการสุ่มในกลุ่ม B สมมติว่าต้องการสุ่มข้อสอบจำนวน  $n$  ข้อ โดยที่  $n$  เป็นจำนวนเต็มที่สุ่มจากช่วงจำนวนตั้งแต่ 5 ถึง 10 ข้อ

ขั้นที่ 2.2 หาช่วงของการสุ่ม ( $k$ ) โดยคำนวณจาก  $k = \frac{N}{n}$  เมื่อ  $N$  คือ จำนวนข้อสอบทั้งหมดในกลุ่ม B และ  $n$  คือ ข้อสอบที่ต้องสุ่ม

ขั้นที่ 2.3 หาข้อสอบที่ใช้เป็นตัวสุ่มเริ่มต้น (r) โดยสุ่มค่า r จากค่าตั้งแต่ 1 ถึง k จากนั้นจึงเลือกข้อสอบโดยเริ่มตั้งแต่ข้อที่ r, r+k, r+2k ตามลำดับจนครบ n ข้อ

ขั้นที่ 3 รวมข้อสอบในกลุ่ม A กับข้อสอบที่สุ่มได้ในขั้นที่ 2 เข้าด้วยกัน ซึ่งข้อสอบทั้งหมดในขั้นนี้ จะนำไปใช้ในกระบวนการคัดเลือกข้อสอบข้อถัดไป

1.2 ผลการพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอร์วิคซ์ มีขั้นตอนการดำเนินการดังนี้

ขั้นที่ 1 นำข้อสอบที่ผ่านการควบคุมการใช้ข้อสอบ (จากข้อ 1.1) มาคำนวณหาค่าเฉลี่ยถ่วงน้ำหนักของค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ H(i) ตามเกณฑ์ของเฮอร์วิคซ์ ซึ่งมีสูตรดังนี้

$$H(i) = \alpha \times \min(SEE_{\text{true}}(\hat{\theta}), SEE_{\text{false}}(\hat{\theta})) + (1-\alpha) \times \max(SEE_{\text{true}}(\hat{\theta}), SEE_{\text{false}}(\hat{\theta})) \quad (5)$$

โดยที่

H(i)	คือ	ค่าเฉลี่ยถ่วงน้ำหนักของค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ ณ ข้อสอบข้อที่ i
$SEE_{\text{true}}(\hat{\theta})$	คือ	ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ ในกรณีที่ผู้สอบตอบถูก
$SEE_{\text{false}}(\hat{\theta})$	คือ	ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ ในกรณีที่ผู้สอบตอบผิด
$\hat{\theta}$	คือ	ค่าประมาณความสามารถของผู้สอบ
$\alpha$	คือ	สัมประสิทธิ์ของการมองโลกในแง่ดี มีเท่ากับ $1 - \frac{1}{5.5}  \hat{\theta} - b $
b	คือ	ค่าความยากของข้อสอบ

ขั้นที่ 2 เลือกข้อสอบที่มีค่า H(i) ต่ำสุด มาเป็นข้อสอบข้อถัดไป

ค่า  $\alpha$  มีค่าตั้งแต่ 0 ถึง 1 กำหนดได้จากที่ทราบ ว่า ข้อสอบที่มีคุณภาพสูงเป็นข้อสอบที่มีค่าความยากของข้อสอบใกล้เคียงกับค่าความสามารถของผู้สอบ โดยค่าความยากของข้อสอบในทางปฏิบัติจะมีค่าอยู่ระหว่าง -2.5 ถึง 2.5 ส่วนค่าความสามารถของผู้สอบในทางปฏิบัติจะมีค่าอยู่ระหว่าง -3 ถึง 3 ซึ่งค่าความยากของข้อสอบและค่าความสามารถของผู้สอบถูกวัดในสเกลเดียวกัน ดังนั้น ค่าความยากของข้อสอบและค่าความสามารถของผู้สอบจึงมีค่าแตกต่างกันต่ำสุด 0 หน่วย และมีค่าแตกต่างกันสูงสุด 5.5 หน่วย และการเทียบสัดส่วนตามหลักคณิตศาสตร์ จึงกำหนดให้  $\alpha$  มีค่าเท่ากับ  $1 - \frac{1}{5.5} |\hat{\theta} - b|$

2. ผลการเปรียบเทียบประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไป แบ่งได้ 3 ประเด็นดังนี้

2.1 ผลการเปรียบเทียบประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบ แสดงดังตารางที่ 1 - 3

**ตารางที่ 1** ผลการวิเคราะห์ความแตกต่างของค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าความลำเอียงเฉลี่ย จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป

	<i>Asymp F</i>	<i>df1</i>	<i>df2</i>	<i>p-value</i>
RMSE	414.261	3	16.097	.000**
Average Bias	54.655	3	21.212	.000**

\*\*  $p < .01$

จากตารางที่ 1 แสดงผลการวิเคราะห์ความแตกต่างของค่า RMSE และค่าความลำเอียงเฉลี่ย (Average Bias) จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป ด้วยการวิเคราะห์ Brown-Forsythe ปรากฏว่ามีวิธีการคัดเลือกข้อสอบข้อถัดไปอย่างน้อย 1 คู่ ที่มีค่า RMSE และค่าความลำเอียงเฉลี่ยแตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .01 ผู้วิจัยจึงเปรียบเทียบพหุคูณของค่า RMSE และค่าความลำเอียงเฉลี่ย ด้วยการทดสอบของแทมเฮน แสดงดังตารางที่ 2 และ 3

**ตารางที่ 2** ผลการเปรียบเทียบพหุคูณของค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป

วิธีการ	Mean	ผลต่างเฉลี่ย (Mean Difference)			
		MIC	RDM	HC	HC-Ex
MIC	.564	-	.258**	.233**	.223**
RDM	.306		-	-.025**	-.035**
HC	.331			-	-.010
HC-Ex	.341				-

\*\*  $p < .01$

จากตารางที่ 2 แสดงผลการเปรียบเทียบพหุคูณของค่า RMSE จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป 4 วิธี ปรากฏว่า วิธีการที่มีค่า RMSE แตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .01 มีทั้งหมด 5 คู่ ได้แก่ วิธีการ MIC กับ RDM, วิธีการ MIC กับ HC, วิธีการ MIC กับ HC-Ex, วิธีการ RDM กับ HC และวิธีการ RDM กับ HC-Ex ส่วนวิธีการ HC กับวิธีการ HC-Ex มีค่า RMSE ไม่แตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .01 เมื่อพิจารณาจากค่าเฉลี่ยของ RMSE จึงสรุปได้ว่า ค่า RMSE ของวิธีการ RDM มีค่าต่ำสุด รองลงมา ได้แก่ วิธีการ HC ซึ่งมีค่าไม่แตกต่างกับวิธีการ HC-Ex ส่วนวิธี MIC มีค่า RMSE สูงสุด แสดงว่า วิธีการ RDM มีประสิทธิภาพสูงสุด

รองลงมา คือ วิธีการ HC ซึ่งมีประสิทธิภาพเทียบเท่ากับวิธีการ HC-Ex ส่วนวิธีการ MIC มีประสิทธิภาพต่ำสุด ซึ่งเป็นไปตามสมมติฐานการวิจัยข้อ 1.1 บางส่วน

**ตารางที่ 3** ผลการเปรียบเทียบพหุคูณของค่าความลำเอียงเฉลี่ย จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป

วิธีการ	Mean	ผลต่างเฉลี่ย (Mean Difference)			
		MIC	RDM	HC	HC-Ex
MIC	.058	-	.058**	.054**	.059**
RDM	.000		-	-.004	.001
HC	.004			-	.005
HC-Ex	-.001				-

\*\*  $p < .01$

จากตารางที่ 3 แสดงผลการเปรียบเทียบพหุคูณของค่าความลำเอียงเฉลี่ย จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป ปรากฏว่า วิธีการที่มีค่าความลำเอียงเฉลี่ยแตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .01 มีทั้งหมด 3 คู่ ได้แก่ วิธีการ MIC กับ RDM, วิธีการ MIC กับ HC และวิธีการ MIC กับ HC-Ex วิธีการที่มีค่าความลำเอียงเฉลี่ยไม่แตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .01 มี 3 คู่ ได้แก่ วิธีการ RDM กับ HC, วิธีการ RDM กับ HC-Ex และวิธีการ HC กับ HC-Ex เมื่อพิจารณาจากค่าเฉลี่ยของค่าความลำเอียงเฉลี่ย สรุปได้ว่า วิธีการ RDM, HC และ HC-Ex มีค่าความลำเอียงเฉลี่ยไม่แตกต่างกัน และมีค่าต่ำกว่าวิธีการ MIC แสดงว่า วิธีการ RDM, HC และ HC-Ex มีประสิทธิภาพสูงสุดเทียบเท่ากัน ส่วนวิธีการ MIC มีประสิทธิภาพต่ำสุด ซึ่งเป็นไปตามสมมติฐานการวิจัยข้อ 1.2 บางส่วน

2.2 ผลการเปรียบเทียบประสิทธิภาพด้านความยาวของแบบทดสอบ แสดงดังตารางที่ 4 - 5

**ตารางที่ 4** ผลการวิเคราะห์ความแตกต่างของความยาวของแบบทดสอบ จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป

วิธีการ	Median	$X^2$	$df$	$p$ -value
MIC	11	13,398.43	3	.000**
RDM	7			
HC	7			
HC-Ex	10			

\*\* $p < .01$

จากตารางที่ 4 แสดงผลการวิเคราะห์ความแตกต่างของความยาวของแบบทดสอบ จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป 4 วิธี ด้วยการทดสอบ Kruskal-Wallis ปรากฏว่า มีวิธีการคัดเลือกข้อสอบข้อถัดไป อย่างน้อย 1 คู่ ที่มีความยาวของแบบทดสอบแตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .01 ดังนั้น จึงเปรียบเทียบค่าเฉลี่ยของอันดับเป็นรายคู่ ด้วยการทดสอบ Dunn-Bonferroni แสดงดังตารางที่ 5

**ตารางที่ 5** ผลการเปรียบเทียบค่าเฉลี่ยของอันดับเป็นรายคู่ของความยาวของแบบทดสอบ จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป

วิธีการ	ค่าเฉลี่ยของอันดับ	ผลต่างค่าเฉลี่ยของอันดับ			
		MIC	RDM	HC	HC-Ex
MIC	28,734.58	-	15,209.19**	15,137.87**	4,589.26**
RDM	13,525.39		-	-71.32	-10,619.93**
HC	13,596.71			-	-10,548.61**
HC-Ex	24,145.32				-

\*\* $p < .01$

จากตารางที่ 5 แสดงผลการเปรียบเทียบค่าเฉลี่ยของอันดับเป็นรายคู่ของความยาวของแบบทดสอบ จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป ด้วยการทดสอบ Dunn-Bonferroni ปรากฏว่า วิธีการที่มีความยาวของแบบทดสอบแตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .01 มีทั้งหมด 5 คู่ ได้แก่ วิธีการ MIC กับ RDM, วิธีการ MIC กับ HC, วิธีการ MIC กับ HC-Ex, วิธีการ RDM กับ HC-Ex และวิธีการ HC กับ HC-Ex ส่วนวิธีการ RDM กับ HC มีความยาวของแบบทดสอบไม่แตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .01 เมื่อพิจารณาจากค่าเฉลี่ยของอันดับ สรุปได้ว่า ค่าเฉลี่ยของอันดับจากวิธีการ RDM มีค่าต่ำสุด ซึ่งไม่แตกต่างกับวิธีการ HC รองลงมา ได้แก่ วิธีการ HC-Ex และ MIC ตามลำดับ แสดงว่า วิธีการ RDM มีประสิทธิภาพสูงสุดเทียบเท่ากับวิธีการ HC รองลงมา ได้แก่ วิธีการ HC-Ex และวิธีการ MIC ตามลำดับ ซึ่งเป็นไปตามสมมติฐานการวิจัยข้อ 2 บางส่วน

2.3 ผลการเปรียบเทียบประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 แสดงดังตารางที่ 6 และตารางที่ 7

**ตารางที่ 6** ผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป

วิธีการ	Median	$\chi^2$	df	p-value
MIC	0	30.486	3	.000**
RDM	10			
HC	10			
HC-Ex	2			

\*\*  $p < .01$

จากตารางที่ 6 แสดงผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป 4 วิธี ด้วยการทดสอบ Kruskal-Wallis ปรากฏว่ามีวิธีการคัดเลือกข้อสอบข้อถัดไปอย่างน้อย 1 คู่ ที่มีจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 แตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .01 ดังนั้น จึงเปรียบเทียบค่าเฉลี่ยของอันดับเป็นรายคู่ของจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 ด้วยการทดสอบ Dunn-Bonferroni แสดงดังตารางที่ 7

**ตารางที่ 7** ผลการเปรียบเทียบค่าเฉลี่ยของอันดับเป็นรายคู่ของจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป

วิธีการ	ค่าเฉลี่ยของ อันดับ	ผลต่างค่าเฉลี่ยของอันดับ			
		MIC	RDM	HC	HC-Ex
MIC	6.20	-	-21.55**	-25.50**	-10.15**
RDM	27.75		-	-3.95	11.40**
HC	31.70			-	15.35**
HC-Ex	16.35				-

\*\* $p < .01$

จากตารางที่ 7 แสดงผลการเปรียบเทียบค่าเฉลี่ยของอันดับเป็นรายคู่ของจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 จำแนกตามวิธีการคัดเลือกข้อสอบข้อถัดไป 4 วิธี โดยใช้การทดสอบ Dunn-Bonferoni ปรากฏว่า วิธีการที่มีจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 แตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .01 มีทั้งหมด 5 คู่ ได้แก่ วิธีการ MIC กับ RDM, วิธีการ MIC กับ HC, วิธีการ MIC กับ HC-Ex, วิธีการ RDM กับ HC-Ex และวิธีการ HC กับ HC-Ex ส่วนวิธีการ RDM กับ HC มีจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบไม่แตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .01 เมื่อพิจารณาจากค่าเฉลี่ยของอันดับ สรุปได้ว่าค่าเฉลี่ยของอันดับจากวิธีการ MIC มีค่าต่ำสุด รองลงมา ได้แก่ วิธีการ HC-Ex ส่วนวิธีการ RDM มีค่าสูงสุด ซึ่งไม่แตกต่างกับวิธีการ HC แสดงว่า วิธีการ MIC มีประสิทธิภาพสูงสุด รองลงมา ได้แก่ วิธีการ HC-Ex ส่วนวิธีการ RDM ซึ่งมีประสิทธิภาพต่ำสุด เทียบเท่ากับวิธีการ HC ซึ่งเป็นไปตามสมมติฐานการวิจัยข้อ 3 บางส่วน

สรุปได้ว่า วิธีการ HC ซึ่งเป็นวิธีการคัดเลือกข้อสอบข้อถัดไปที่พัฒนาขึ้น มีประสิทธิภาพสูงในด้านการประมาณค่าความสามารถของผู้สอบ และด้านความยาวของแบบทดสอบ ถึงแม้จะมีประสิทธิภาพต่ำกว่าวิธีการ RDM แต่มีประสิทธิภาพสูงกว่าวิธีการ MIC ซึ่งเป็นวิธีการที่นิยมใช้ทั่วไป ส่วนวิธีการ HC-Ex ซึ่งเป็นวิธีการคัดเลือกข้อสอบข้อถัดไปที่มีการควบคุมการใช้ข้อสอบ ถึงแม้จะมีประสิทธิภาพต่ำกว่าวิธีการ MIC แต่สามารถควบคุมการใช้ข้อสอบได้ค่อนข้างมีประสิทธิภาพ เนื่องจากข้อสอบในคลังข้อสอบขนาด 500 ข้อ มีเพียง 2 ข้อ เท่านั้น ที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 ดังนั้น วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอร์วิคซ์และการควบคุมการใช้ข้อสอบที่พัฒนาขึ้น จึงเป็นอีกทางเลือกหนึ่งสำหรับนำไปใช้คัดเลือกข้อสอบข้อถัดไปในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

### อภิปรายผลการวิจัย

1. การพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอร์วิคซ์และการควบคุมการใช้ข้อสอบ การนำเกณฑ์ของเฮอร์วิคซ์ซึ่งเป็นเกณฑ์หนึ่งภายใต้ทฤษฎีการตัดสินใจในสภาวะการณ์ไม่แน่นอน มาใช้คัดเลือกข้อสอบข้อถัดไป โดยคำนวณค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบ (SEE) ของข้อสอบทุกข้อในคลังข้อสอบ ที่ยังไม่ถูกเลือกใช้ในการทดสอบ ซึ่งคำนวณค่า SEE ทั้งในกรณีที่คาดว่าผู้สอบจะตอบข้อสอบถูก และกรณีที่คาดว่าผู้สอบจะตอบข้อสอบผิด แล้วนำค่า SEE ทั้งสองกรณี มาคำนวณค่าเฉลี่ยถ่วงน้ำหนักของค่า SEE จากนั้นตัดสินใจเลือกข้อสอบข้อที่มีค่าเฉลี่ยถ่วงน้ำหนักของค่า SEE ต่ำสุด เป็น

ข้อสอบข้อถัดไป ตามกระบวนการนี้ เกณฑ์ของเฮอร์วิคซ์จึงมีความเหมาะสมสำหรับใช้คัดเลือกข้อสอบข้อถัดไป เนื่องจากผู้สอบจะได้รับข้อสอบที่มีความเหมาะสมกับระดับความสามารถของตนเอง (เป็นข้อสอบที่มีค่าเฉลี่ยของค่า SEE ต่ำ) นอกจากนี้ เกณฑ์ของเฮอร์วิคซ์เป็นเกณฑ์ที่มีประสิทธิภาพค่อนข้างสูง เมื่อเทียบกับเกณฑ์อื่น ๆ ในทฤษฎีการตัดสินใจในสภาวะการณ์ที่ไม่แน่นอน สอดคล้องกับงานวิจัยของ Groenewald and Pretorius (2011) ที่ได้ศึกษากลยุทธ์ในการตัดสินใจเลือกพอร์ตการลงทุน โดยใช้เกณฑ์ที่แตกต่างกันของการตัดสินใจในสภาวะการณ์ที่ไม่แน่นอน ผลปรากฏว่า การตัดสินใจโดยใช้เกณฑ์ของเฮอร์วิคซ์ให้ผลตอบแทนรายเดือนของตลาดเงินที่เหมาะสมที่สุด

การคัดเลือกข้อสอบข้อถัดไป ส่วนใหญ่จะคัดเลือกข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง มาเป็นข้อสอบข้อถัดไป ดังนั้น ข้อสอบเหล่านี้จึงถูกเลือกใช้ในการทดสอบบ่อยครั้งจนเกินไป เป็นผลให้ผู้สอบในรุ่นถัดไปสามารถคาดเดาข้อสอบที่ตนเองจะได้รับล่วงหน้าได้ ในการวิจัยนี้ ได้พัฒนาวิธีการควบคุมการใช้ข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบมีระบบ (Systematic Random Sampling) มาควบคุมการใช้ข้อสอบ โดยข้อสอบที่นำไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป คือ ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบต่ำถึงปานกลาง และข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงจำนวน 5-10 ข้อ ซึ่งถูกสุ่มจากกลุ่มข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง ด้วยวิธีการสุ่มตัวอย่างแบบมีระบบ ด้วยกระบวนการนี้ ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง จึงมีโอกาสถูกสุ่มเท่า ๆ กัน และการสุ่มข้อสอบจะกระทำทุกครั้ง ก่อนเข้าสู่ขั้นตอนการคัดเลือกข้อสอบข้อถัดไป ทำให้ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงถูกหมุนเวียนไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป วิธีการนี้จึงลดจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 ลงได้ (Georgiadou et al., 2007)

## 2. การเปรียบเทียบประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไป

การเปรียบเทียบประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบ ปรากฏว่า เมื่อพิจารณาจากค่า RMSE วิธีการ RDM มีประสิทธิภาพสูงสุด รองลงมา ได้แก่ วิธีการ HC ซึ่งมีประสิทธิภาพเทียบเท่ากับวิธีการ HC-Ex ส่วนวิธีการ MIC มีประสิทธิภาพต่ำสุด ซึ่งเป็นไปตามสมมติฐานการวิจัยข้อที่ 1.1 บางส่วน แต่เมื่อพิจารณาจากค่าความลำเอียงเฉลี่ย วิธีการ RDM, HC และ HC-Ex มีประสิทธิภาพสูงสุดเทียบเท่ากัน ส่วนวิธีการ MIC มีประสิทธิภาพต่ำสุด แสดงให้เห็นว่า วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ทฤษฎีการตัดสินใจ ได้แก่ วิธีการ RDM, HC และ HC-Ex มีประสิทธิภาพสูงกว่าวิธีการ MIC ซึ่งเป็นวิธีการที่นิยมใช้ทั่วไป เนื่องจาก วิธีการ RDM, HC และ HC-Ex ได้ถูกพัฒนาขึ้นโดยใช้ทฤษฎีการตัดสินใจ ควบคู่กับทฤษฎี IRT ซึ่งทฤษฎีการตัดสินใจจะคำนวณผลตอบแทนที่คาดว่าจะได้รับของทุกทางเลือก แล้วตัดสินใจเลือกทางเลือกที่ให้ผลตอบแทนที่ดีที่สุด เมื่อนำหลักการนี้มาใช้ในการคัดเลือกข้อสอบข้อถัดไป อย่างเช่น วิธีการ HC กับวิธีการ HC-Ex ที่พัฒนาขึ้นได้คำนวณค่า SEE ที่คาดว่าจะเกิดขึ้น เมื่อผู้สอบทำข้อสอบแต่ละข้อในคลังข้อสอบ โดยคำนวณค่า SEE ทั้งกรณีที่คาดว่าจะผู้สอบตอบถูก และกรณีที่คาดว่าจะผู้สอบตอบผิด แล้วนำมาคำนวณค่าเฉลี่ยถ่วงน้ำหนักของค่า SEE และตัดสินใจเลือกข้อสอบที่มีค่าเฉลี่ยถ่วงน้ำหนักของค่า SEE ต่ำสุด เป็นข้อสอบข้อถัดไป ด้วยกระบวนการนี้ข้อสอบข้อถัดไปที่ผู้สอบได้รับ จึงเป็นข้อสอบที่มีค่า SEE ต่ำสุด ประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบ จึงสูงกว่าวิธีการ MIC ที่พิจารณาเฉพาะค่าสารสนเทศสูงสุดที่ใกล้เคียงกับค่าความสามารถของผู้สอบ ณ ขณะนั้น (Taylor, 2009, pp. 538-539; Birnbaum, 1968 cited in van der Linden & Glas, 2002, pp. 9-10)

เมื่อพิจารณาวิธีการ RDM, HC และ HC-Ex ซึ่งใช้ทฤษฎีการตัดสินใจ ผลการวิจัยแสดงให้เห็นว่าวิธีการ HC และ HC-Ex ที่พัฒนาขึ้นมีประสิทธิภาพต่ำกว่าวิธีการ RDM เป็นผลมาจาก วิธีการ RDM เป็นวิธีการที่พัฒนาจากทฤษฎีการตัดสินใจในสภาวะการตัดสินใจความเสี่ยง (Risk Decision Making) ซึ่งใช้ค่าความน่าจะเป็นมาคำนวณผลตอบแทนที่คาดว่าจะเกิดขึ้น ผลลัพธ์ที่ได้จึงมีความแม่นยำมากกว่า วิธีการ HC และ HC-Ex ที่พัฒนาจากทฤษฎีการตัดสินใจในสภาวะการตัดสินใจที่ไม่แน่นอน (Uncertainly Decision Making) ที่ใช้เพียงค่าเฉลี่ยถ่วงน้ำหนักมาคำนวณผลตอบแทนที่คาดว่าจะเกิดขึ้น (Taylor, 2009, pp. 539-547) แต่งานวิจัยนี้ ผู้วิจัยมุ่งพัฒนาโดยใช้ทฤษฎีการตัดสินใจ ที่มีการคำนวณไม่ยุ่งยากซับซ้อนเท่ากับการใช้ทฤษฎีการตัดสินใจในสภาวะการตัดสินใจความเสี่ยง ซึ่งเกณฑ์ของเฮอร์วิคซ์ในทฤษฎีการตัดสินใจในสภาวะการตัดสินใจที่ไม่แน่นอน ใช้เพียงค่าเฉลี่ยถ่วงน้ำหนัก ถึงแม้จะมีความแม่นยำน้อยกว่า แต่ให้ประสิทธิภาพใกล้เคียงกัน เห็นได้จาก เมื่อพิจารณาจากค่าความลำเอียงเฉลี่ยประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบของวิธีการ RDM, HC และ HC-Ex มีประสิทธิภาพสูงเทียบเท่ากัน

การเปรียบเทียบประสิทธิภาพด้านความยาวของแบบทดสอบ ปรากฏว่า วิธีการ RDM มีประสิทธิภาพสูงสุดเทียบเท่ากับวิธีการ HC รองลงมา ได้แก่ วิธีการ HC-Ex และ MIC ตามลำดับ ซึ่งเป็นไปตามสมมติฐานการวิจัยข้อที่ 2 บางส่วน แสดงให้เห็นว่า วิธีการ HC ที่พัฒนาขึ้นใช้แบบทดสอบที่มีความยาวน้อย เนื่องจากวิธีการนี้ใช้ทฤษฎีการตัดสินใจในสภาวะการตัดสินใจที่ไม่แน่นอน ตามเกณฑ์ของเฮอร์วิคซ์ โดยเลือกข้อสอบที่มีค่าเฉลี่ยถ่วงน้ำหนักของค่า SEE ต่ำสุด เป็นข้อสอบข้อถัดไป ทำให้ข้อสอบที่ผู้สอบได้รับมีความเหมาะสมกับระดับความสามารถของผู้สอบ สอดคล้องกับงานวิจัยของ โสฬส สุขานนท์สวัสดิ์ และคณะ (2556) ที่ได้พัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ทฤษฎีการตัดสินใจในสภาวะการตัดสินใจความเสี่ยงผลการวิจัยปรากฏว่า วิธีการที่พัฒนาขึ้นสามารถลดจำนวนข้อสอบและเวลาที่ใช้ในการทดสอบลงได้

การเปรียบเทียบประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 ปรากฏว่าวิธีการ MIC มีประสิทธิภาพสูงสุด รองลงมา ได้แก่ วิธีการ HC-Ex ส่วนวิธีการ RDM มีประสิทธิภาพต่ำสุดเทียบเท่าวิธีการ HC ซึ่งเป็นไปตามสมมติฐานการวิจัยข้อที่ 3 บางส่วน แสดงให้เห็นว่า วิธีการ HC-Ex เป็นวิธีการที่มีประสิทธิภาพค่อนข้างสูง สามารถควบคุมข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 ให้มีจำนวนลดลงได้โดยข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 ส่วนใหญ่เป็นข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง ซึ่งวิธีการนี้ ทำให้ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง มีโอกาสถูกเลือกใช้ในการทดสอบน้อยลง โดยการสุ่มข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงจำนวน 5-10 ข้อ ด้วยวิธีการสุ่มตัวอย่างแบบมีระบบ แล้วนำข้อสอบที่สุ่มได้นี้ไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป ซึ่งการสุ่มข้อสอบนี้ จะทำการสุ่มใหม่ทุกครั้งก่อนเข้าสู่ขั้นตอนการคัดเลือกข้อสอบข้อถัดไป ทำให้ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง ถูกหมุนเวียนนำไปใช้ในการคัดเลือกข้อสอบข้อถัดไป โอกาสถูกเลือกใช้จึงลดลง ซึ่งสอดคล้องกับงานวิจัยของ Ozturk and Dogan (2015) ที่ได้เปรียบเทียบประสิทธิภาพของวิธีการควบคุมการใช้ข้อสอบ ปรากฏว่า วิธีการ Randomesque ซึ่งเป็นวิธีการควบคุมการใช้ข้อสอบด้วยวิธีการสุ่มสามารถลดจำนวนข้อสอบที่มีอัตราการใช้ข้อสอบมากกว่า 0.2 ให้น้อยลงได้

## ข้อเสนอแนะ

### ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. ผู้ที่สนใจพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไป สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สามารถนำผลการพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้เกณฑ์ของเฮอร์วิคซ์ไปใช้เป็นแนวทางในการพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปวิธีการใหม่ ที่ใช้เกณฑ์อื่นในทฤษฎีการตัดสินใจในสถานการณ์ที่ไม่แน่นอนได้ เช่น ใช้เกณฑ์แมกซิแมก เกณฑ์แมกซิมิน หรือเกณฑ์ลาปลาซ

2. ผู้ที่สนใจพัฒนาวิธีการควบคุมการใช้ข้อสอบ ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สามารถนำผลการพัฒนาวิธีการควบคุมการใช้ข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบมีระบบ ไปใช้เป็นแนวทางในการพัฒนาวิธีการควบคุมการใช้ข้อสอบวิธีการใหม่ ที่ใช้วิธีการสุ่มตัวอย่างวิธีการอื่นๆ ได้ เช่น วิธีการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (Stratified Random Sampling)

### ข้อเสนอแนะในการวิจัยต่อไป

1. วิธีการควบคุมการใช้ข้อสอบที่พัฒนาขึ้นนี้ มุ่งเน้นเข้าควบคุมการใช้ข้อสอบที่มีอัตราการใช้ข้อสอบสูง (Over-Exposure Item) ให้ลดลง แต่ไม่ได้ควบคุมข้อสอบที่ถูกเลือกใช้น้อยหรือไม่ถูกเลือกใช้เลย (Under-Utilized Item) ให้มีอัตราการใช้ข้อสอบเพิ่มขึ้น จึงควรพัฒนาวิธีการควบคุมการใช้ข้อสอบให้ครอบคลุมกรณีข้อสอบที่ถูกเลือกใช้น้อยหรือไม่ถูกเลือกใช้เลย

2. การศึกษาเปรียบเทียบประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไปนี้ ใช้คลังข้อสอบขนาดคงที่ 500 ข้อ ซึ่งมีขนาดค่อนข้างใหญ่ จึงควรศึกษาในกรณีที่ใช้คลังข้อสอบขนาดต่าง ๆ กัน เช่น ขนาดเล็ก หรือขนาดปานกลาง เพื่อให้ทราบประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไปครอบคลุมกับคลังข้อสอบทุกขนาด

## เอกสารอ้างอิง

### ภาษาไทย

กิตติ ภัคดีวัฒนะกุล และพนิดา พานิชกุล. (2554). *การวิเคราะห์เชิงปริมาณเพื่อการตัดสินใจ*. กรุงเทพฯ:

เคทีพี คอมพ์ แอนด์ คอนซัลท์.

โสฬส สุขานนท์สวัสดิ์, เสรี ชัดเข้ม และ กฤษณะ ชินสาร. (2556). การพัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไป

โดยใช้ทฤษฎีการตัดสินใจในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์. *วิทยาการวิจัยและ*

*วิทยาการปัญญา*, 10(2), 71-85.

### ภาษาต่างประเทศ

Baker, F. B. (2001). *The Basics of Item Response Theory* (2<sup>nd</sup> ed.). USA: ERIC Clearinghouse on Assessment and Evaluation.

Chang, H.H., & van der Linden, W. J. (2003). Optimal Stratification of Item Pools in a-Stratified Computerized Adaptive Testion. *Applied Psychological Measurement*, 27(4), 262 – 274.

- de Ayala, R. J. (2009). *The theory and practice of item response theory*. London: The Guilford Press.
- Georgiadou, E., Triantafillou, E., & Economides, A. (2007). A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *Journal of Technology, Learning, and Assessment*, 5(8), 1-38.
- Groenewald, M. E., & Pretorius, P. D. (2011). Comparison of Decision-making under Uncertainty Investment Strategies with the Money Market. *Journal of Financial Studies and Research*, 1-16.
- Luecht, R. M., & Sireci, S. G. (2011). *A Preview of Models for Computer-Based Testing*. New York: The College Board. Retrieved from <https://research.collegeboard.org/publications/content/2012/05/review-models-computer-based-testing>
- Lunz, M. E., Bergstrom, B. A., & Wright, B. D. (1992). The Effect of Review on Student Ability and Test Efficiency for Computerized Adaptive Tests. *Applied Psychological Measurement*. 16(1): 33-40.
- McBride, J. R., & Martin, J. T. (1983). Reliability and validity of adaptive ability tests in a military setting. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 223-226). New York: Academic Press.
- Owen, R. J. (1975). A Bayesian Sequential Procedure for Quantal Response in the Context of Adaptive Mental Testing. *Journal of the American Statistical Association*, 70(350), 351-356.
- Ozturk, N. B., & Dogan, N. (2015). Investigating Item Exposure Control Methods in Computerized Adaptive Testing. *Educational Sciences: Theory & Practice*, 15(1), 85-98.
- Sympson, J. B., & Hetter, R. D. (1985). Controlling item-exposure rates in computerized adaptive testing. In Proceedings of the 27<sup>th</sup> annual meeting of the Military Testing Association (pp. 973-977). San Diego, CA: Navy Personnel Research and Development Center.
- Taylor, B. W. (2009). *Introduction to Management Science* (11<sup>th</sup> ed.). Upper Saddle River, NJ: Prentice Hall.
- Thompson, N. A., & Weiss, D. J. (2011). A Framework for the Development of Computerized Adaptive Tests. *Practical Assessment, Research & Evaluation*, 16(1), 1-9.
- van der Linden, W. J., & Glas, G. A. W. (2002). *Computerized Adaptive Testing: Theory and Practice*. New York: Kluwer Academic.
- Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., & Mislevy, R. J. (2000). *Computerized Adaptive Testing: A Primer* (2<sup>nd</sup> edit.). New Jersey: Lawrence Erlbaum Associates.