

การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด
ภายใต้วิธีการคัดเลือกข้อสอบ และวิธีการควบคุมการใช้ข้อสอบซ้ำที่แตกต่างกัน

DEVELOPMENT OF HYBRID COMPUTERIZED ADAPTIVE TESTING SYSTEM
BASED ON DIFFERENT ITEM SELECTION
AND ITEM EXPOSURE CONTROL METHODS

รังสิมาภรณ์ หนูน้อย¹ โชติกา ภาชีผล² ศิริชัย กาญจนาวาสี³

Rungsimaporn Noonoi¹ Shotika Pasiphol² Sirichai Kanjanawasee³

^{1,2,3}คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

^{1,2,3}Faculty of Education, Chulalongkorn University

Received: January 22, 2020 / Revised: March 20, 2020 / Accepted: April 30, 2020

บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดภายใต้วิธีการคัดเลือกข้อสอบระหว่าง วิธีสารสนเทศสูงสุดของฟิชเชอร์ (Maximum Fisher Information: MFI) และวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) และการควบคุมการใช้ข้อสอบซ้ำระหว่างวิธีซิมสันและเฮตเตอร์ (Simpson-Hetter Procedure: SH) และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive Restricted Standard Error: PR-SE) โดยจำลองกลุ่มตัวอย่าง 1,000 คน จำลองคลังข้อสอบ จำนวน 600 ข้อ การคำนวณค่าความน่าจะเป็นในการตอบข้อสอบถูกของกลุ่มผู้สอบที่จำลองขึ้นกับคลังข้อสอบ โดยใช้สูตรแบบ 3 พารามิเตอร์ ศึกษาประสิทธิภาพวิธีการประมาณค่าความสามารถ วิธีการคัดเลือกข้อสอบ วิธีการควบคุมการใช้ข้อสอบซ้ำและเกณฑ์การยุติการสอบน้อยกว่าหรือเท่ากับ 0.3 โดยการวิเคราะห์ประสิทธิภาพด้านความถูกต้องแม่นยำของการประมาณค่าความสามารถของผู้สอบ จากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าความลำเอียงเฉลี่ย และค่าเฉลี่ยความสัมพันธ์ระหว่างค่าความสามารถจริงกับค่าความสามารถประมาณค่า

ผลการวิจัยพบว่า วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-Restricted Standard Error: PR-SE) มีประสิทธิภาพความถูกต้องและความแม่นยำสูงกว่า โดยเกณฑ์ยุติการทดสอบเมื่อความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.3

คำสำคัญ: การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด วิธีการคัดเลือกข้อสอบ วิธีการควบคุมการใช้ข้อสอบซ้ำ

Abstract

The purposes of this experiment were to compare the efficiency of ability estimation different item selection methods between Maximum Fisher Information (MFI) and Maximum Priority Index (MPI) and item exposure control methods between Sympton-Hetter Procedure (SH) Progressive-Restricted Standard Error (PR-SE). There were 1,000 examinees who simulated an item pool of 600 items and estimated the possibility of choosing the correct answers of examinees with item pool by the Three-Parameter Model. Efficiency of ability estimation, item selection method, item exposure control methods, and termination criteria were the standard error of estimation less than or equal to 0.3. There were employed to bias, Mean Squared Error (MSE) and compare the Pearson product-moment correlation in estimating examinees.

The research indicates that the Maximum Priority Index (MPI) with Progressive-Restricted Standard Error (PR-SE) has higher efficiency and accuracy when the standard error of estimation is less than or equal to 0.3.

Keywords: Hybrid Computerized Adaptive Testing, Item Selection, Item Exposure Control

บทนำ

ปัจจุบันได้มีการนำเทคโนโลยีทางคอมพิวเตอร์มาใช้ในการทดสอบ ซึ่งเทคโนโลยีทางคอมพิวเตอร์ได้เข้ามาช่วยเสริมสร้างประสิทธิภาพในการตรวจข้อสอบ การวิเคราะห์ข้อสอบ การตัดเกรด และรายงานผล ซึ่งมีการใช้กันอย่างกว้างขวาง รวมทั้งยังมีการนำมาใช้สร้างคลังข้อสอบ (Item Bank) การทดสอบที่ใช้คอมพิวเตอร์ที่รู้จักกันดีคือ การทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (Computer Adaptive Testing: CAT) ซึ่งเป็นการทดสอบที่มีการคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ โดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ในการประมาณค่าความสามารถที่แท้จริงของผู้สอบ (Kanjana-wasee, 2012) การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ ใช้ข้อสอบเพียง 1 ข้อ ในการประมาณค่าความสามารถของผู้สอบและเลือกข้อถัดไปจากค่าความสามารถที่ได้จากการประมาณค่าในข้อแรก ถ้าทำถูกข้อถัดไปจะมีความยากขึ้น ถ้าทำผิดข้อถัดไปจะมีความยากลดลง ส่วนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบ

หลายขั้นตอน (Multi-Stage Test Design: MST) ใช้การประมาณค่าความสามารถของผู้สอบ เลือกชุดข้อสอบข้อถัดไปโดยดูจากค่าความสามารถในการประมาณค่าในขั้นแรก โดยเลือกชุดที่มีความยากใกล้เคียงกับผู้สอบ ถ้าผู้สอบมีความสามารถต่ำจะเลือกชุดข้อสอบที่ง่ายลง ถ้าผู้สอบมีความสามารถสูงจะเลือกชุดข้อสอบที่ยากขึ้น อย่างไรก็ตามการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) แตกต่างกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ ในจำนวนของข้อคำถามที่ใช้ในการทดสอบตามเวลาที่กำหนดโดยปกติการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จะกำหนดกลุ่มข้อสอบไว้ล่วงหน้าตามระดับความยาก ตัวอย่างเช่น ง่าย ปานกลาง และยาก และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ถูกมองว่าเป็นโมดูลที่ใช้ทดสอบการปรับเหมาะ แม้ว่าวัตถุประสงค์หลักของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) คือ การประเมินประสิทธิภาพ

ความสามารถของผู้สอบกับข้อคำถามน้อยกว่าที่จำเป็นต้องใช้ในการทดสอบกระดาษและดินสอ แต่ทั้งสองวิธีก็มีจุดแข็งและจุดอ่อนที่ส่งผลต่อการประมาณค่าความสามารถ

Wang et al. (2016) จึงนำจุดแข็งในการดำเนินการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อและการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มาศึกษาจนนำไปสู่การพัฒนาการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดเพื่อเพิ่มประสิทธิภาพ การประมาณค่าความสามารถของผู้สอบโดยการรวมแนวคิดของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อเข้าด้วยกัน ซึ่งเป็นเพียงแนวคิดและเป็นการศึกษากระบวนการต้นแบบ ยังไม่เคยมีการนำแนวคิดดังกล่าวมาใช้ในการทดสอบจริง โดยนำวิธีการประมาณค่าความสามารถและวิธีการคัดเลือกข้อสอบในการออกแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (Computer Adaptive Testing: CAT) มาใช้ในการทดสอบ เช่น สารสนเทศสูงสุดของฟิชเชอร์ (Maximum Fisher Information: MFI) และการประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood Estimate: MLE) สามารถนำมาใช้ในขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดได้ สำหรับขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จะมี 2 ขั้นตอน หรือ 3 ขั้นตอนนั้น ขึ้นอยู่กับข้อเสนอนี้และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มีการลดความยาวของข้อสอบ โดยมีการศึกษาวิจัยเรื่องความยาวของโมดูลจะขึ้นอยู่กับขั้นตอนของเส้นทางที่ยาวซึ่งสามารถให้ประมาณการที่แม่นยำมากขึ้น (Hambleton & Xing, 2006; Jodoin, Zenisky & Hambleton, 2006; Patsula, 1999) และในการทดสอบปรับเหมาะด้วยแบบไฮบริดมีการดำเนินการออกแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์

แบบหลายขั้นตอน (MST) โดยมีการกำหนดเส้นทางและลดความยาวโมดูลเพื่อตอบสนองต่อวัตถุประสงค์ในการประมาณค่าความสามารถที่ถูกต้องสำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายข้อ จะใช้วิธี Automatic Test Assembly (ATA) ในการตรวจสอบ

ในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์จะมีองค์ประกอบที่สำคัญ ได้แก่ จุดเริ่มต้นการทดสอบ การประมาณค่าความสามารถ การคัดเลือกข้อสอบ การควบคุมการใช้ข้อสอบซ้ำ และเกณฑ์การยุติการทดสอบ ซึ่งการคัดเลือกข้อสอบนับว่าเป็นหัวใจสำคัญของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์วิธีการเลือกข้อสอบที่ดีต้องพิจารณาประสิทธิภาพสูงสุดในการประมาณค่าความสามารถให้ใกล้เคียงค่าความสามารถจริงของผู้สอบ รวมทั้งพิจารณาการจัดสมดุลของเนื้อหาและการควบคุมการเปิดเผยข้อสอบหรือการควบคุมการใช้ข้อสอบซ้ำไปพร้อมๆ กันด้วย ซึ่งการควบคุมสมดุลเนื้อหาเพื่อต้องการให้แน่ใจว่าแบบทดสอบที่แตกต่างกันในการจัดให้ผู้สอบแต่ละคนครอบคลุมสัดส่วนเนื้อหาเหมือนกัน สามารถอ้างอิงไปยังการวัดผู้สอบคนอื่นๆ ในลักษณะเดียวกัน และการควบคุมการใช้ข้อสอบซ้ำเพื่อไม่ต้องการให้ข้อสอบบางข้อนำไปใช้บ่อยเกินไปจนรู้จักกันทั่วไปในกลุ่มผู้สอบ

เนื่องจาก ส่วนที่ 1 วิธีการคัดเลือกข้อสอบ 2 วิธี ได้แก่ 1) วิธีสารสนเทศสูงสุดของฟิชเชอร์ (Maximum Fisher Information: MFI) 2) วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) และ ส่วนที่ 2 วิธีการควบคุมอัตราการใช้ข้อสอบซ้ำ 2 วิธี ได้แก่ 1) วิธีซิมสันและเฮตเตอร์ (Simpson-Hetter Procedure: SH) 2) วิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-Restricted Standard Error: PR-SE) พัฒนาขึ้นภายใต้บริบทของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ ดังนั้น เพื่อขยายองค์ความรู้ใหม่ ผู้วิจัยจึงสนใจนำวิธีการคัดเลือกข้อสอบทั้ง 2 วิธี มาพัฒนาต่อในบริบทของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ แบบ

ไฮบริดเพื่อให้ประโยชน์ในการประมาณค่าความสามารถที่ถูกต้องแม่นยำ

วัตถุประสงค์การวิจัย

เพื่อเปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดภายใต้

1. วิธีการคัดเลือกข้อสอบระหว่างวิธีสารสนเทศสูงสุดของฟิชเชอร์ (Maximum Fisher Information: MFI) และวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI)

2. วิธีการควบคุมการใช้ข้อสอบซ้ำ ระหว่างวิธีซิมสันและเฮตเตอร์ (Simpson-Hetter Procedure: SH) และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-Restricted Standard Error: PR-SE)

ทบทวนวรรณกรรม

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์เป็นการทดสอบที่มีการคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ ซึ่งเป็นการใช้ระบบคอมพิวเตอร์ในการสอบเพื่อประเมินความสามารถของผู้สอบแต่ละคนมีการคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ เริ่มต้นจากข้อสอบข้อที่ง่าย ปานกลาง และยาก และเมื่อทำถูกจะได้ข้อสอบข้อถัดไปที่ยากขึ้นเรื่อยๆ โดยกระบวนการสอบจะดำเนินการไปจนกระทั่งประเมินความสามารถของผู้สอบได้ (Kanjawasee, 2012)

การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดคือ การรวมกรอบแนวคิดการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) และการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายชื่อ โดยนำวิธีการประมาณค่าความสามารถและวิธีการคัดเลือกข้อสอบในการออกแบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) มาใช้ในการทดสอบ

เช่น สารสนเทศสูงสุดของฟิชเชอร์ (Maximum Fisher Information: MFI) และการประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood Estimate: MLE) สามารถนำมาใช้ในขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดได้ สำหรับขั้นตอนการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) มีการลดความยาวของข้อสอบ โดยมีการกำหนดเส้นทางและลดความยาวโมดูลตอบสนองต่อวัตถุประสงค์ในการประมาณค่าความสามารถที่ถูกต้อง สำหรับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายชื่อจะใช้วิธี Automatic Test Assembly (ATA) ในการตรวจสอบ

สำหรับขั้นตอนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด จะเริ่มต้นการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) ด้วยการเลือกข้อสอบเป็นชุด (Initial Item) ที่มีจำนวน 5 ข้อ โดยการสุ่ม (MST ขั้นที่ 1) ที่มีความยากระดับปานกลาง จากนั้นจะทำการประมาณค่าความสามารถที่แท้จริงของผู้สอบจากการตอบข้อสอบในขั้นแรก และระบบจะคัดเลือกข้อสอบชุดถัดไปที่มีความยากใกล้เคียงกับความสามารถของผู้สอบ (MST ขั้นที่ 2) ในขั้นตอนนี้จะมีการแบ่งโมดูลข้อสอบออกเป็น 3 ระดับ ได้แก่ ระดับง่าย ระดับปานกลาง และระดับยาก เมื่อผู้สอบทำชุดข้อสอบตามระดับความสามารถแล้ว ระบบจะคัดเลือกข้อสอบถัดไปเป็นรายชื่อ ซึ่งเป็นส่วนของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายชื่อ ระบบจะประเมินระดับความคลาดเคลื่อนของการประมาณค่าจนถึงเกณฑ์การยุติการทดสอบสำหรับกระบวนการเตรียมชุดข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (MST) จะกระทำขึ้นหลังจากการประมาณค่าความสามารถของผู้สอบ จะมีการจัดเตรียมชุดข้อสอบไว้ล่วงหน้า และการเตรียมข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบรายชื่อ จะใช้คอมพิวเตอร์ในการคัดเลือกข้อสอบแต่ละข้อ (Wang et al., 2016)

กระบวนการในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด (Wang et al., 2016) ประกอบไปด้วยรายละเอียดดังนี้

ขั้นตอนที่ 1 จุดเริ่มต้นการทดสอบ (Starting Point or Entry Level) การทดสอบ การคัดเลือกข้อสอบข้อแรก (initial item) ด้วยวิธีการสุ่มชุดข้อสอบที่มีจำนวน 5 ข้อ มีความยากปานกลาง จัดให้กับผู้สอบแต่ละคนเมื่อไม่ทราบค่าสารสนเทศเบื้องต้นของผู้สอบ

ขั้นตอนที่ 2 การประมาณค่าความสามารถสำหรับขั้นตอนการออกแบบระบบการประมาณค่าใช้การประมาณค่าร่วมกัน วิธีการประมาณค่าด้วยความเป็นไปได้สูงสุด (Maximum Likelihood) คัดเลือกข้อสอบที่มีตำแหน่งสารสนเทศสูงสุด (m_i) สอดคล้องกับระดับความสามารถที่ประมาณได้ (θ) [Match m_i to θ] สำหรับโมเดล 3 พารามิเตอร์

ขั้นตอนที่ 3 การคัดเลือกข้อสอบ (Item Selection) แบบทดสอบที่มีการควบคุมเงื่อนไขทางสถิติ (tests without nonstatistical constraints) เช่น ความครอบคลุมของเนื้อหา กระบวนการเลือกข้อสอบสามารถนำมาใช้กับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด ซึ่งในแต่ละขั้น (stage) จะทำการเลือกข้อสอบที่มีสารสนเทศสูงสุดภายหลังการประมาณค่าความสามารถของผู้สอบ ในงานวิจัยนี้มีการเปรียบเทียบวิธีการคัดเลือกข้อสอบ 2 วิธี ได้แก่

3.1 วิธีสารสนเทศสูงสุดของฟิชเชอร์ (Maximum Fisher Information: MFI) เป็นวิธีที่ใช้กันอย่างแพร่หลายในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ที่มีประสิทธิภาพในการประมาณค่าความสามารถ การทดสอบข้อสอบก่อนหน้านี้ในการบริหารจัดการทดสอบของผู้เข้าสอบ โดยทั่วไปเป็นการค้นหาข้อสอบ x ทำให้สูงสุด $I[\theta_{m-1}]$ สำหรับผู้เข้าสอบ ในการประมาณค่าความสามารถ θ และ $m - 1$ ของจำนวนข้อสอบในการบริหาร การทดสอบต่อไปในคลังข้อสอบแบบตัวเลือกมีการกำหนดรูปแบบโมเดลโลจิสติกส์ 3 พารามิเตอร์ (3PL) วิธีการคัดเลือกข้อสอบตามเกณฑ์ MFI จะคัดเลือกข้อสอบสำหรับข้อสอบ 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 เป็นพารามิเตอร์ค่าอำนาจจำแนกข้อสอบ

b_i เป็นพารามิเตอร์ค่าความยากของข้อสอบ

c_i เป็นพารามิเตอร์การเดาข้อสอบได้ถูก

D เป็นค่าคงที่เท่ากับ 1.702

(Deng, Ansley & Chang, 2010)

3.2 วิธีดัชนีลำดับความสำคัญสูงสุดจะพิจารณาความแปรปรวนของค่าสารสนเทศโดยการนำดัชนีไปเป็นตัวคูณให้กับค่าสารสนเทศของข้อสอบ ดังนั้นในการคัดเลือกข้อสอบแทนที่จะพิจารณาจากค่าสารสนเทศเพียงอย่างเดียว วิธีการนี้จะนำลำดับความสำคัญซึ่งกำหนดโดยผู้เชี่ยวชาญที่พิจารณาร่วมด้วย การคัดเลือกข้อสอบจะพิจารณาจากผลคูณของค่าสารสนเทศของข้อสอบกับดัชนีความสำคัญ ข้อใดให้ผลคูณมากที่สุด ข้อนั้นจะได้รับการคัดเลือกให้นำไปให้กับผู้สอบ ดังนั้นวิธีดัชนีลำดับความสำคัญสูงสุดจึงเกี่ยวข้องกับเมทริกซ์เงื่อนไขบังคับ ซึ่งจะบอกว่าข้อสอบแต่ละข้อประกอบด้วยเงื่อนไขบังคับใดบ้าง (Cheng & Chang, 2009)

ขั้นตอนที่ 4 การควบคุมการใช้ข้อสอบซ้ำ (exposure control)

ในการควบคุมการเปิดเผยของข้อสอบสามารถควบคุมได้ด้วยกระบวนการอัตโนมัติเหมือนกับการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ โดยเราสามารถป้องกันข้อสอบที่ถูกนำมาใช้บ่อย (over-exposed) จากกรณีที่มีการเปิดเผยและมีการแบ่งปันข้อสอบกับผู้สอบอื่น ในงานวิจัยนี้มีการเปรียบเทียบการควบคุมการใช้ข้อสอบซ้ำ 2 วิธี ได้แก่

4.1 วิธีซิมสันและเฮตเตอร์ (Simpson-Hetter Procedure: SH) พัฒนาโดยซิมสันและเฮตเตอร์ (1985) การใช้ข้อสอบซ้ำสามารถที่จะควบคุมได้เป็นอย่างดีภายใต้วิธีซิมสันและเฮตเตอร์ ซึ่งจะเป็นพื้นฐานสำหรับวิธีการในการควบคุมการใช้ข้อสอบซ้ำที่มีประสิทธิภาพมากยิ่งขึ้น วิธีซิมสันและเฮตเตอร์ จะควบคุมการใช้ข้อสอบซ้ำโดยตรงในลักษณะของความเป็นไปได้ วิธีซิมสันและเฮตเตอร์ยอมให้มีการกำหนดอัตราการแสดงข้อสอบล่วงหน้า เพื่อยืนยัน

ว่าการแจกแจงอัตราการแสดงข้อสอบควรมีลักษณะเดียวกัน ซึ่งเป็นวิธีพื้นฐานในการประยุกต์ไปสู่วิธีการคัดเลือกข้อสอบอย่างมีเงื่อนไขอื่นๆ (Elissavet & Economides, 2007)

4.2 วิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-Restricted Standard Error: PR-SE) ใช้สูตรคำนวณเดียวกับวิธีจำกัดความก้าวหน้า (Progressive-restricted Procedure) แต่มีนิยาม S ใหม่เป็นอัตราส่วนของความคลาดเคลื่อนมาตรฐาน (SE) กฎการหยุด มากกว่าความคลาดเคลื่อนมาตรฐาน (SE) ปัจจุบันการใช้ความคลาดเคลื่อนมาตรฐาน (SE) ไม่ได้ขึ้นอยู่กับตำแหน่งตามลำดับ แต่ใช้ความคลาดเคลื่อนมาตรฐาน (SE) เพื่อตรวจสอบองค์ประกอบการสุ่มหรือสารสนเทศที่มีอิทธิพลต่อการถ่วงน้ำหนัก เช่นเดียวกับขั้นตอนการสุ่ม ขั้นตอนการทำงานร่วมกันง่ายกว่าวิธีการแบบเงื่อนไข (McClarty, Sperling & Dodd, 2006)

ขั้นตอนที่ 5 เกณฑ์การยุติการสอบ (termination criteria or Stopping criteria)

การกำหนดค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ในทางปฏิบัติการทดสอบปรับเหมาะกับความสามารถของผู้สอบโดยใช้คอมพิวเตอร์ ควรให้ดำเนินการไปเรื่อยจนกว่าการประมาณค่า θ มีความคลาดเคลื่อนมาตรฐาน (Standard Error: SE (θ)) ลดต่ำลงจนถึงระดับที่ยอมรับได้ การทดสอบจึงควรยุติลง ซึ่งการกำหนดค่าความคลาดเคลื่อนมาตรฐานของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดได้กำหนดการประมาณค่าความสามารถของผู้สอบน้อยกว่าหรือเท่ากับ 0.30 (Promsit, 2012)

วิธีการวิจัย

1. สถานการณ์จำลองที่ใช้ในการศึกษา

1.1 ตัวอย่างสำหรับการศึกษาในครั้งนี้คือ พารามิเตอร์ผู้สอบ ได้แก่ ค่าความสามารถจริง (θ_i) ที่สร้างขึ้นโดยสุ่มการแจกแจงโค้งปกติมาตรฐานมีค่า

เฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 [$\theta \sim N(0,1)$] และกำหนดให้มีค่าตั้งแต่ -4.0 ถึง 4.0 สำหรับการทดสอบในแต่ละเงื่อนไขผู้วิจัยกำหนดผู้เข้าสอบไว้ที่ 1,000 คน ในการทดสอบหนึ่งครั้งจะต้องมีพารามิเตอร์ผู้สอบจำนวน 1,000 ชุด เมื่อเปลี่ยนเงื่อนไขในการทดสอบระบบจะทำการสุ่มค่าพารามิเตอร์ผู้สอบใหม่ทุกครั้ง ในการทดสอบแต่ละเงื่อนไขจะสุ่มตัวอย่างเพื่อทำซ้ำ 10 รอบ (Koedsri, 2014) ดังนั้นตัวอย่างที่ใช้ในสถานการณ์จำลองข้อมูลแต่ละเงื่อนไขจะมี 10,000 ค่า และการศึกษานี้มี 4 เงื่อนไข ได้แก่ Maximum Fisher information (MFI) กับ Sympton-Hetter (SH), Maximum Fisher information (MFI) กับ Progressive-Restricted Standard Error (PR-SE), Maximum Priority Index (MPI) กับ Sympton-Hetter (SH) และ Maximum Priority Index (MPI) กับ Progressive-Restricted Standard Error (PR-SE) ดังนั้นจึงใช้จำนวนตัวอย่างทั้งหมด 40,000 ตัวอย่าง

1.2 การจำลองข้อมูล

การพัฒนาระบบการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริด กรณีการจำลองข้อมูล โดยใช้โปรแกรม R มี 4 ขั้นตอนหลัก ได้แก่ 1) จำลองกลุ่มตัวอย่าง ความสามารถจริง (θ) 2) จำลองคลังข้อสอบ โครงสร้างของคลัง 3) การคำนวณค่าความน่าจะเป็นในการตอบข้อสอบถูก 4) การคำนวณผลการตอบ มีรายละเอียดดังนี้

ขั้นตอนที่ 1 จำลองกลุ่มตัวอย่างความสามารถจริง (θ) กลุ่มตัวอย่างสำหรับการศึกษาในครั้งนี้ ได้แก่ ค่าความสามารถจริงที่สร้างขึ้นโดยสุ่มการแจกแจงโค้งปกติมาตรฐานมีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 [$\theta \sim N(0,1)$] และกำหนดให้มีค่าตั้งแต่ -4.0 ถึง 4.0 สำหรับการทดสอบในแต่ละเงื่อนไขผู้วิจัยกำหนดผู้เข้าสอบไว้ที่ 1,000 คน ในการทดสอบหนึ่งครั้งจะต้องมีพารามิเตอร์ผู้สอบจำนวน 1,000 ชุด เมื่อเปลี่ยนเงื่อนไขในการทดสอบระบบจะทำการสุ่มค่าพารามิเตอร์ผู้สอบใหม่ทุกครั้ง

ในการทดสอบแต่ละเงื่อนไขจะสุ่มตัวอย่างเพื่อทำซ้ำ 100 รอบ (Koedsri, 2014) ดังนั้นตัวอย่างที่ใช้ในสถานการณ์จำลองข้อมูลแต่ละเงื่อนไขจะมี 10,000 ค่า และการศึกษาครั้งนี้มี 4 เงื่อนไข ดังนั้นจึงใช้จำนวนตัวอย่างทั้งหมด 40,000 ตัวอย่าง

ขั้นตอนที่ 2 จำลองคลังข้อสอบ โครงสร้างของคลัง โดยจำลองคลังข้อสอบ จำนวน 600 ข้อ ซึ่งเป็นคลังข้อสอบขนาดปานกลางและมีความเป็นไปได้ในการนำไปในสถานการณ์การทดสอบที่เป็นจริง และกำหนดแนวทางการทดสอบตามหลักการทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) ใช้รูปแบบโลจิสติกส์ 3 พารามิเตอร์ (3 Parameter Logistic Model: 3 PLM) (Kanjawasee, 2012) คือ ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -4 ถึง 4 ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง ∞ ค่าการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.50

ขั้นตอนที่ 3 การคำนวณค่าความน่าจะเป็นในการตอบข้อสอบถูกของกลุ่มผู้สอบที่จำลองขึ้นกับคลังข้อสอบ โดยใช้สูตรแบบ 3 พารามิเตอร์ (Three-Parameter Model)

ขั้นตอนที่ 4 การคำนวณผลการตอบสามารถทำได้โดยนำความน่าจะเป็นในการตอบข้อสอบไปคำนวณหาผลการตอบของผู้สอบ โดยการนำค่าความน่าจะเป็นในการตอบข้อสอบถูก $P_i(\theta)$ ไปพิจารณาผลการตอบข้อสอบ ตอบถูกต้องได้คะแนนเท่ากับ 1 และตอบผิดได้คะแนนเท่ากับ 0

2. ประสิทธิภาพด้านความถูกต้องแม่นยำของการประมาณค่าความสามารถของผู้สอบ

ประสิทธิภาพด้านความถูกต้องแม่นยำของการประมาณค่าความสามารถของผู้สอบเพื่อเลือกวิธีการคัดเลือกข้อสอบและวิธีการควบคุมการใช้ข้อสอบซ้ำที่มีการประมาณค่าความสามารถของผู้สอบให้เท่ากับหรือใกล้เคียงค่าความสามารถจริงของผู้สอบมากที่สุด (Promsit, 2012; Koedsri, 2014) เกณฑ์ที่ใช้พิจารณาจาก

2.1 ค่าความลำเอียงเฉลี่ย (Bias) หมายถึง ค่าความคลาดเคลื่อนที่บอกความแตกต่างระหว่างค่าความสามารถจริงของผู้สอบกับค่าความสามารถที่ประมาณค่าจากการทดสอบปรับเหมาะด้วยคอมพิวเตอร์

2.2 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) หมายถึง ค่าที่บอกความถูกต้องในการประมาณค่าความสามารถของผู้สอบ

2.3 ค่าเฉลี่ยความสัมพันธ์ระหว่างค่าความสามารถจริงกับค่าความสามารถประมาณค่า

3. การวิเคราะห์ข้อมูล

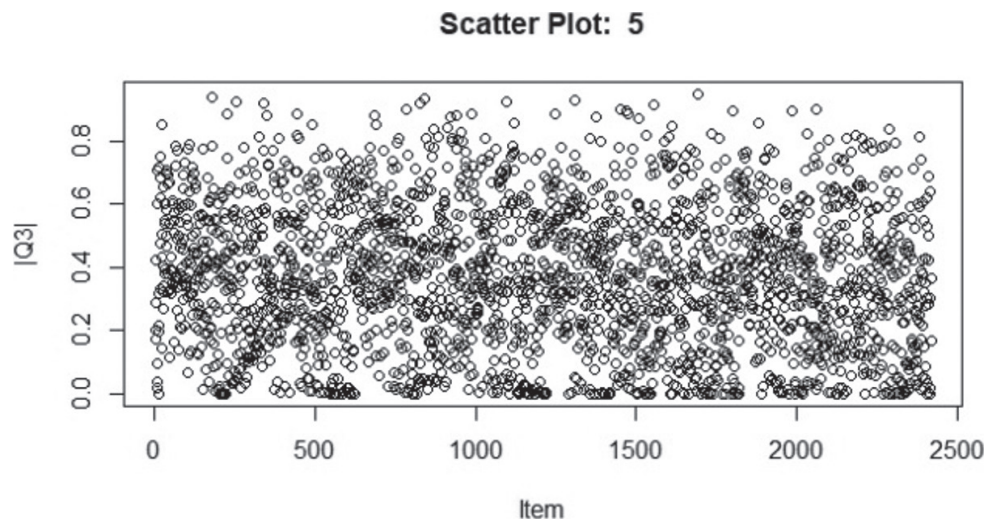
ในการวิจัยใช้ขนาดกลุ่มตัวอย่างในแต่ละเงื่อนไข 1,000 คน ทำการทดลองซ้ำ 100 ครั้ง และจะใช้ผลของการทดลองซ้ำแต่ละครั้งในการวิเคราะห์เปรียบเทียบ โดยในแต่ละครั้งของการทดลองจะคำนวณค่าความลำเอียงเฉลี่ย (Bias) ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) และค่าเฉลี่ยความสัมพันธ์ระหว่างค่าความสามารถจริงกับค่าความสามารถประมาณค่า

4. การตรวจสอบความถูกต้องของข้อมูล

ผู้วิจัยได้ตรวจสอบผลการจำลองของข้อมูลโดยใช้ข้อตกลงเบื้องต้นเกี่ยวกับความไม่เป็นอิสระเฉพาะที่ของข้อสอบ (Local Item Dependence: LID) เพื่อวินิจฉัยว่าข้อสอบที่รวมกันอยู่ภายใต้การทดสอบคอมพิวเตอร์แบบไฮบริดเป็นอิสระต่อกันหรือไม่ โดย Yen (1984) เสนอสถิติ Q3 ซึ่งเป็นดัชนีสำหรับการตรวจสอบความไม่เป็นอิสระเฉพาะที่ของข้อสอบโดยการหาความสัมพันธ์ของส่วนที่เหลือ (residuals) ของคู่ของข้อสอบหลังจากตัดส่วนที่เป็นการประมาณค่าคุณลักษณะ (trait estimate) ออกไป โดยผลการวิเคราะห์ค่าสถิติ Q3 มีค่าเฉลี่ย Q3 เป็น -0.00528 แสดงว่าข้อสอบมีความเป็นอิสระเฉพาะที่ของข้อสอบโดยมีรายละเอียดผลการวิเคราะห์ค่าสถิติ Q3 แสดงในตารางที่ 1 และภาพที่ 1

ตารางที่ 1 ผลการวิเคราะห์ค่าสถิติ Q3

item	Mean	S.D.	Min	Max
600	-0.00528	0.43208	-0.94815	0.90267



ภาพที่ 1 ค่าเฉลี่ยของสถิติ Q3

ผลการวิจัย

ผลการเปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบไฮบริดภายใต้วิธีการคัดเลือกข้อสอบ และวิธีการควบคุมการใช้ข้อสอบซ้ำที่แตกต่างกันพบว่า

วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-Restricted Standard Error: PR-SE) มีค่าความคลาดเคลื่อนที่บอกความแตกต่างระหว่างค่าความสามารถจริงของผู้สอบกับค่าความสามารถที่ประมาณค่าจากการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ โดยมีค่าความลำเอียงเฉลี่ย (BIAS) 0.1995892

วิธีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-Restricted Standard Error: PR-SE) มีความถูกต้องในการประมาณค่าความสามารถของผู้สอบมากที่สุด โดยมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) 0.07008627

วิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) และวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-Restricted Standard Error: PR-SE) มีค่าเฉลี่ยความสัมพันธ์ระหว่างค่าความสามารถจริงกับค่าความสามารถประมาณค่ามากที่สุด (Correlation) 0.9754560 ดังตารางที่ 2

ตารางที่ 2 ผลการวิเคราะห์ความแม่นยำความถูกต้องและความสัมพันธ์

วิธีการ	ค่าความคลาดเคลื่อน กำลังสองเฉลี่ย (Bias)	ค่าความคลาดเคลื่อนกำลังสอง เฉลี่ย (MSE)	ค่าเฉลี่ยความสัมพันธ์ (Corr)
MFI-SH	0.2995328	0.1196238	0.9756660
MPI-SH	0.2501426	0.0925185	0.9756047
MFI-PR-SE	0.2994115	0.1195482	0.9756452
MPI-PR-SE	0.1995892	0.0700862	0.9754560

อภิปรายผล

การคัดเลือกข้อสอบด้วยวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) มีความแม่นยำสูงสุดทั้งนี้อาจเป็นเพราะมีการเพิ่มประสิทธิภาพจากการใช้ค่าฟังก์ชันสารสนเทศสูงสุดและมีการปรับค่าฟังก์ชันสารสนเทศสูงสุดให้สามารถใช้ควบคู่กับกระบวนการในการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบอื่นๆ ทำให้การคัดเลือกข้อสอบมีความแม่นยำมากยิ่งขึ้น สอดคล้องกับงานวิจัยของ Cheng & Chang (2009) วิธีดัชนีลำดับความสำคัญสูงสุด ซึ่งระบุว่าเป็นวิธีเพื่อใช้ควบคุมเงื่อนไขบังคับอย่างเข้มงวดในการคัดเลือกข้อสอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ จะพิจารณาความแปรปรวนของค่าสารสนเทศโดยการนำดัชนีไปเป็นตัวคูณให้กับค่าสารสนเทศของข้อสอบ ดังนั้นในการคัดเลือกข้อสอบแทนที่จะพิจารณาจากค่าสารสนเทศเพียงอย่างเดียว วิธีการนี้จะนำลำดับความสำคัญพิจารณาร่วมด้วย และสามารถปรับให้เข้ากับข้อบังคับที่ไม่ใช่ทางสถิติที่หลากหลาย และสามารถควบคุมได้หลายอย่างพร้อมๆ กัน เช่น ความสมดุลของเนื้อหา การควบคุมการใช้ข้อสอบซ้ำ ความสมดุลของคำตอบ โดยที่วิธีการนี้สามารถนำไปปรับใช้กับโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ที่มีอยู่ในปัจจุบันได้สะดวก โดยไม่จำเป็นต้องปรับค่าน้ำหนักความสัมพันธ์ระหว่างข้อบังคับและสารสนเทศ

วิธีการควบคุมการใช้ข้อสอบซ้ำวิธีค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-Restricted Standard Error: PR-SE) มีประสิทธิภาพสูงสุดทั้งนี้อาจเป็นเพราะมีการใช้อัตราส่วนของความคลาดเคลื่อนมาตรฐาน (SE) ในกฎการหยุดเพื่อใช้ตรวจสอบองค์ประกอบการสุ่มป้องกันไม่ให้มีการเปิดเผยข้อสอบมากเกินไปและเพิ่มอัตราการคัดเลือกข้อสอบที่ไม่ค่อยได้รับการถูกเลือกและเป็นวิธีที่มีการรักษาความแม่นยำในการประมาณค่าความสามารถสอดคล้องกับงานวิจัยของ Leroux et al. (2013) ได้ศึกษาเรื่องการเปรียบเทียบวิธีการควบคุมการใช้ข้อสอบซ้ำใน CATs โดยใช้โมเดล 3 พารามิเตอร์ ได้แก่ ค่าความคลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า (Progressive-Restricted Standard Error: PR-SE) วิธีการสุ่ม (randomesque) และวิธี Sympson-Hetter (SH) พบว่า วิธี PR-SE ให้ประโยชน์ของการควบคุมการใช้ข้อสอบซ้ำโดยภาพรวมสำหรับคลังข้อสอบขนาดใหญ่และเล็ก วิธี PR-SE สามารถบริหารจัดการได้เกือบทั้งหมดของคลังข้อสอบ ในขณะที่วิธีการอื่นๆ สามารถบริหารจัดการได้เพียง 52% หรือน้อยกว่าสำหรับคลังข้อสอบขนาดใหญ่ และทำได้ 80% หรือน้อยกว่าสำหรับคลังข้อสอบขนาดเล็ก วิธี PR-SE ยังทำให้มีจำนวนการซ้อนทับกันน้อยที่สุดของข้อสอบระหว่างแบบสอบข้ามเงื่อนไข

จากการจำลองข้อมูลระบบการทดสอบปรับ
 เหมาะด้วยคอมพิวเตอร์แบบไฮบริด โดยวิธีการคัด
 เลือกข้อสอบด้วยวิธีดัชนีลำดับความสำคัญสูงสุด
 (Maximum Priority Index: MPI) และวิธีการควบคุม
 การใช้ข้อสอบด้วยวิธีค่าความคลาดเคลื่อนมาตรฐาน
 แบบจำกัดความก้าวหน้า (Progressive-Restricted
 Standard Error: PR-SE) ซึ่งทั้งสองวิธีมีประสิทธิภาพ
 สูงสุด เนื่องจากการลดการฝ่าฝืนข้อบังคับและมี
 ประสิทธิภาพสามารถปรับให้เข้ากับข้อบังคับที่ไม่ใช่
 ทางสถิติที่หลากหลายและสามารถควบคุมได้หลาย
 อย่างพร้อมๆ กัน ซึ่งเหมาะสมกับการทดสอบปรับ
 เหมาะด้วยคอมพิวเตอร์แบบไฮบริดเนื่องจากการ
 ใช้คลังข้อสอบขนาดใหญ่ที่มีความยาวแปรผัน และมี
 ขั้นตอนที่หลากหลายที่มีการทดสอบปรับเหมาะสมด้วย
 คอมพิวเตอร์รายข้อและการทดสอบปรับเหมาะสมด้วย
 คอมพิวเตอร์แบบหลายขั้นตอน

สรุปผล

ผลการเปรียบเทียบประสิทธิภาพการ
 ประเมินค่าความสามารถของผู้สอบในการทดสอบ
 ปรับเหมาะสมด้วยคอมพิวเตอร์แบบไฮบริดภายใต้วิธีการ

คัดเลือกข้อสอบและวิธีการควบคุมการใช้ข้อสอบซ้ำ
 ที่แตกต่างกันพบว่า วิธีดัชนีลำดับความสำคัญสูงสุด
 (Maximum Priority Index: MPI) และวิธีค่าความ
 คลาดเคลื่อนมาตรฐานแบบจำกัดความก้าวหน้า
 (Progressive-Restricted Standard Error: PR-SE)
 มีความแม่นยำสูงกว่าและถูกต้องสูงกว่าเกณฑ์ยุติ
 การทดสอบเมื่อความคลาดเคลื่อนมาตรฐานในการ
 ประเมินค่าความสามารถน้อยกว่าหรือเท่ากับ 0.3 มี
 ประสิทธิภาพสูงสุด นั่นคือ การทดสอบการปรับเหมาะ
 ด้วยคอมพิวเตอร์แบบไฮบริดจะให้ประสิทธิภาพสูงสุด
 เมื่อใช้วิธีการคัดเลือกข้อสอบด้วยดัชนีลำดับความ
 สำคัญสูงสุด และใช้วิธีการควบคุมการใช้ข้อสอบซ้ำ
 ด้วยวิธีความคลาดเคลื่อนมาตรฐานแบบจำกัดความ
 ก้าวหน้าและยุติการทดสอบเมื่อความคลาดเคลื่อน
 มาตรฐานในการประเมินค่าความสามารถน้อยกว่า
 หรือเท่ากับ 0.3

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนทุนวิจัยจาก
 “ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย” กองทุน
 รัชดาภิเษกสมโภช

References

- 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.
- 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.
- Davis, L. L. & Dodd, B. G. (2005). *Strategies for controlling item exposure in computerized adaptive testing with partial credit model (PEM Research Report No. 05-01)*. Austin, TX: Pearson Educational Measurement.
- Deng, H., Ansley, T. & Chang, H. H. (2010). Stratified and Maximum Information Item Selection Procedures in Computer Adaptive Testing. *Journal of Educational Measurement*, 47(2), 202-226.
- Elissavet, G. E. 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), 4-37.

- 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.
- 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.
- Kanjanawasee, S. (2013). *Classical Test Theory* (7th ed.). Bangkok: Chulalongkorn University Publishing.
- . (2012). *Modern Test Theories* (4th ed.). Bangkok: Chulalongkorn University Publishing.
- Koedsri, A. (2014). *Efficiency of two item selection methods in computerized adaptive testing for the testlet response model: A comparison between the Monte Carlo CAT method and the Constraint-Weighted a-stratification method*. Doctor of Philosophy Program in Educational Measurement and Evaluation, Chulalongkorn University.
- 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.
- 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. *Annual meeting of the American Educational Research Association 2006*, 7-11 April 2006. San Francisco, CA: AERA.
- Patsula, L. N. (1999). *A comparison of computerized-adaptive testing and multi-stage testing*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.
- Promsit, P. (2012). *A Study of Efficiency of Multidimensional Computerized Adaptive Testing*. Doctor of Philosophy Thesis in Educational Measurement and Evaluation, Graduate School, Khon Kaen University.
- Sympson, J. B. & Hetter, R. D. (1985). Controlling item-exposure rates in computerized adaptive testing. *Proceedings of the 27th annual meeting of the Military Testing Association*. (pp. 973–977). San Diego, CA: Navy Personnel Research and Development Center.
- Wang, S., Lin, H., Chang, H. & Douglas, J. (2016). Hybrid Computerized Adaptive Testing: From Group Sequential Design to Fully Sequential Design. *Journal of Education Measurement*, 53(1), 45-62.



Name and Surname: Rungsimaporn Noonoi

Highest Education: M.Ed. Education Research and Evaluation,
Kasetsart University

University or Agency: Chulalongkorn University

Field of Expertise: Educational Measurement and Evaluation



Name and Surname: Shotiga Pasiphol

Highest Education: Ph.D. Educational Measurement and Evaluation,
Chulalongkorn University

University or Agency: Chulalongkorn University

Field of Expertise: Educational Measurement and Evaluation



Name and Surname: Sirichai Kanjanawasee

Highest Education: Ph.D. Measurement and Evaluation,
University of California Los Angeles, UCLA

University or Agency: Chulalongkorn University

Field of Expertise: Educational Measurement and Evaluation,
Project Evaluation