

การทดสอบแบบปรับเหมาะด้วย คอมพิวเตอร์สำหรับการวัดด้านเจตพิสัย

สุชาดา สกลกิจรุ่งโรจน์*

ปร.ด. (การวัดและเทคโนโลยีทางวิทยาการปัญญา), อาจารย์ประจำ
สำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช

บทคัดย่อ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เป็นการทดสอบรูปแบบหนึ่งที่มีการปรับกระบวนการทดสอบให้เหมาะสมกับระดับคุณลักษณะแฝงของผู้รับการทดสอบ ซึ่งมีองค์ประกอบที่สำคัญ 5 องค์ประกอบ คือ 1) คลังข้อสอบ 2) การเริ่มต้นการทดสอบ 3) การเลือกข้อคำถามข้อถัดไป 4) การประมาณค่าคุณลักษณะแฝง และ 5) การยุติการทดสอบ บทความนี้มุ่งนำเสนอความรู้เบื้องต้นเกี่ยวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สถานการณ์การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในการวัดทางด้านเจตพิสัยทั้งในประเทศและต่างประเทศ รวมทั้งวิเคราะห์ประเด็นความท้าทาย และมุมมองในการพัฒนาที่เกิดขึ้นจากประสบการณ์ในการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ในการวัดทางด้านเจตพิสัย

คำสำคัญ : การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทฤษฎีการตอบสนองข้อสอบ เจตพิสัย

*วท.บ. (จิตวิทยา) (เกียรตินิยมอันดับ 1), วท.ม. (จิตวิทยาคลินิก), ปร.ด. (การวัดและเทคโนโลยีทางวิทยาการปัญญา) มหาวิทยาลัยบูรพา อาจารย์ประจำสำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช

Computerized Adaptive Testing for Measuring Affective Domain

Suchada Sakolkijrunroj

Ph.D. (Measurement and Technology in Cognitive Science)

Lecturer, Office of Registration, Records and Evaluation,

Sukhothai Thammathirat Open University

Abstract

Computerized Adaptive Testing (CAT) is a form of testing that adapts to the examinee's proficiency level (Latent Trait). Computerized Adaptive Testing consists of 5 components including : 1) Item Bank; 2) Starting the First Item; 3) Selecting the Next Item; 4) Scoring; and 5) Termination Criterion. This article aimed to present the basic knowledge of computerized adaptive testing and the situations of building computerized adaptive testing for affective domain. In addition, the challenges and aspects in developing computerized adaptive testing were discussed.

Keywords : Computerized Adaptive Testing, Item Response Theory, Affective domain

บทนำ

CAT ทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing; CAT) เป็นการทดสอบด้วยคอมพิวเตอร์ (Computer-Based Testing; CBT) รูปแบบหนึ่งที่มีการปรับกระบวนการทดสอบให้เหมาะสมกับระดับความสามารถหรือคุณลักษณะแฝงของบุคคล การทดสอบรูปแบบนี้นำเสนอข้อสอบผ่านทางหน้าจอคอมพิวเตอร์และอาศัยทฤษฎีการทดสอบแนวใหม่มาใช้เป็นแนวทางในการวิเคราะห์คุณภาพของข้อสอบและแบบทดสอบ ทำให้การทดสอบมีประสิทธิภาพและมีความแม่นยำในการทดสอบมากยิ่งขึ้น ผู้รับการทดสอบแต่ละคนจะได้รับชุดของข้อสอบที่แตกต่างกัน แต่เป็นข้อสอบที่เหมาะสมกับบุคคลนั้น ๆ ข้อใดที่ไม่เกี่ยวข้องจะไม่ถูกนำมาใช้ในการทดสอบ ทำให้การทดสอบใช้จำนวนข้อสอบลดลง ซึ่งช่วยประหยัดเวลาในการทดสอบได้อย่างมาก (Gibbons *et al.*, 2008)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ถูกนำมาใช้ครั้งแรกในบริบทของการประเมินทักษะความสามารถ เช่น ทักษะทางภาษา ทักษะทางการคำนวณ เป็นต้น แต่ในช่วงหลายปีที่ผ่านมาได้มีการนำการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มาใช้ในบริบทอื่นมากขึ้นโดยเฉพาะอย่างยิ่งในมิติการวัดทางด้านเจตพิสัย โดยพัฒนาแบบคัดกรองสำหรับใช้ในสถานพยาบาลในรูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เช่น แบบวัดภาวะสุขภาพ แบบวัดความวิตกกังวล แบบวัดภาวะซึมเศร้า แบบวัดบุคลิกภาพ เป็นต้น (Patient-Reported Outcome Measurement Information System [PROMIS], 2013) ทำให้การคัดกรองภาวะสุขภาพกายและจิตใจมีความสะดวก รวดเร็ว และ

มีประสิทธิภาพมากยิ่งขึ้น การพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในมิติการวัดทางด้านเจตพิสัยมีได้แตกต่างจากการวัดทางด้านพุทธิพิสัยเพียงด้านเนื้อหาเท่านั้น แต่ในทางปฏิบัติเทคนิควิธีการ ทฤษฎี และรายละเอียดต่าง ๆ ยังมีความแตกต่างกันในหลายประเด็น บทความนี้จึงมีวัตถุประสงค์เพื่อแนะนำผู้อ่านให้รู้จักกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในบริบททั่วไป จากนั้นจึงนำเสนอสถานการณ์ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในมิติการวัดทางด้านเจตพิสัย แล้วจึงนำเสนอการวิเคราะห์ประเด็นความท้าทายและมุมมองในการพัฒนาการวัดทางด้านเจตพิสัยไว้ในบทความนี้ด้วย

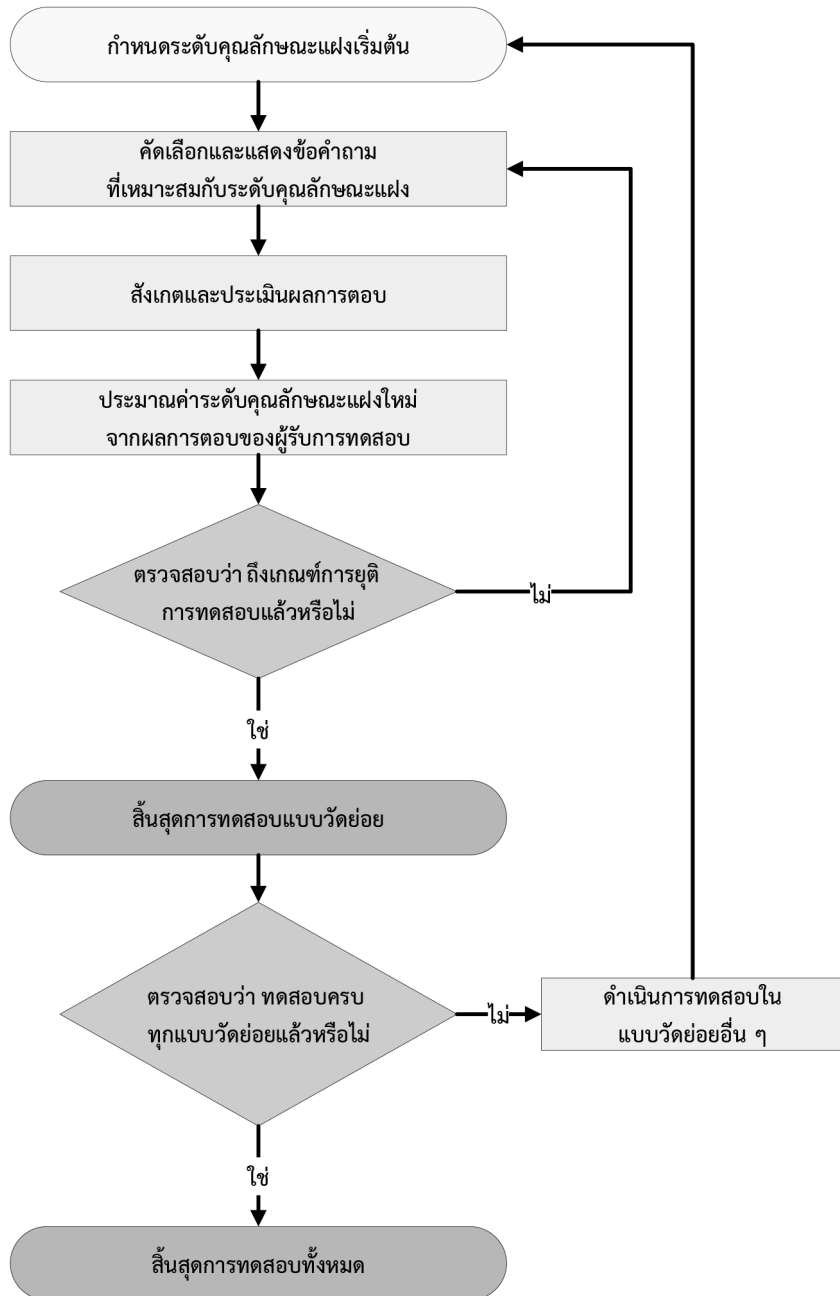
โดยปกติแล้วผู้ที่ให้ความสนใจเกี่ยวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ส่วนใหญ่มีความคุ้นเคยกับการทดสอบและการวัดความสามารถทางด้านพุทธิพิสัย จึงคุ้นชินกับคำว่า “ข้อสอบ แบบสอบ ความสามารถ” อย่างไรก็ดี เพื่อให้เนื้อหาของบทความเหมาะสมกับบริบทการวัดทางด้านเจตพิสัยมากขึ้น ในเนื้อหาต่อจากนี้ผู้เขียนจึงใช้คำว่า “ข้อคำถาม (Item)” แทน “ข้อสอบ (Test item)” ใช้คำว่า “แบบวัด (Scale)” แทน “แบบทดสอบ (Test)” และใช้คำว่า “คุณลักษณะแฝง (Latent Trait)” แทน “ความสามารถ (Ability)” แต่คำศัพท์เฉพาะอื่น ๆ ยังคงไว้ดังเดิม โดยคุณลักษณะแฝง หมายถึง คุณลักษณะของบุคคลที่มีอิทธิพลอยู่เบื้องหลังการตอบข้อคำถามในแบบวัด โดยสะท้อนถึงคุณลักษณะทางจิตวิทยาที่แบบวัดนั้น ๆ มุ่งวัด เช่น ความสุข ความเศร้า ความโกรธ ความเจ็บปวด เป็นต้น

ความรู้เบื้องต้นเกี่ยวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เป็นการทดสอบที่นำเทคโนโลยีคอมพิวเตอร์มาใช้ในการวัดและประเมินผล โดยกระบวนการทดสอบจะมีการเลือกใช้ข้อคำถามจากคลังข้อสอบ (Item Bank) ให้เหมาะสมกับระดับคุณลักษณะแฝงของผู้รับการทดสอบ ทำให้การทดสอบมีความเฉพาะเจาะจงกับแต่ละบุคคลมากขึ้น

การทดสอบตามหลักการของทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory; CTT) เชื่อว่า หากข้อคำถามที่ใช้ในการทดสอบมีจำนวนมาก การทดสอบนั้น ๆ จะมีความแม่นยำในการทดสอบเพิ่มมากขึ้น และผลการทดสอบจะมีความเที่ยง (Reliability) สูง ในขณะที่การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่อาศัยหลักการของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT) เชื่อว่า การดำเนินการทดสอบที่ทำให้ผลการทดสอบมีความเที่ยงสูงหรือมีความแม่นยำในการ

ประมาณค่าระดับคุณลักษณะแฝงของบุคคลสูงไม่ได้ขึ้นอยู่กับความยาว (Test Length) ของแบบวัดเสมอไป แต่ขึ้นอยู่กับ การเลือกข้อคำถามที่ใช้ในการทดสอบให้เหมาะสมกับระดับคุณลักษณะแฝงของบุคคลนั้น จากแนวคิดดังกล่าวส่งผลให้ข้อคำถามที่ไม่มีความสัมพันธ์หรือมีความเกี่ยวข้องกับคุณลักษณะแฝงของผู้รับการทดสอบน้อยจะไม่ถูกเลือกเข้ามาใช้ในการทดสอบ ดังนั้น ข้อคำถามที่ใช้ในการทดสอบจึงมีจำนวนลดลง ระยะเวลาที่ใช้ในการทดสอบลดลง ความเหนื่อยล้าในการทดสอบลดลง ขณะที่ความแม่นยำในการทดสอบไม่แตกต่างจากการทดสอบที่ใช้ข้อคำถามจำนวนมาก และอาจมีความแม่นยำในการทดสอบมากกว่าหากกำหนดระดับความแม่นยำในการทดสอบเป็นเกณฑ์ในการยุติการทดสอบ (Gibbons *et al.*, 2008) กระบวนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เป็นกระบวนการที่เป็นพลวัต (Dynamic) ซึ่งมีขั้นตอนการทดสอบดังภาพที่ 1



ภาพที่ 1 กระบวนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Wainer *et al.*, 2000)

กระบวนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เริ่มด้วยการกำหนดระดับคุณลักษณะแฝงเริ่มต้น จากนั้นจึงเลือกข้อคำถามที่เหมาะสมกับระดับคุณลักษณะแฝงนั้นแล้วนำเสนอข้อคำถามดังกล่าวผ่านทางหน้าจอคอมพิวเตอร์เพื่อให้ผู้รับการทดสอบตอบ การเลือกคำตอบของแต่ละบุคคลจะขึ้นอยู่กับระดับคุณลักษณะแฝงที่ซ่อนอยู่ภายในตัวของบุคคลนั้น เมื่อตอบเรียบร้อยแล้ว โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะนำผลการตอบมาใช้ในการประมาณค่าระดับคุณลักษณะแฝงใหม่เพื่อให้ตรงกับระดับที่เป็นจริงมากที่สุด รวมทั้งตรวจสอบว่า การทดสอบเข้าเกณฑ์การยุติการทดสอบที่กำหนดไว้หรือไม่ หากไม่เข้ากับเกณฑ์จะคัดเลือกข้อคำถามข้อถัดไปเพื่อทดสอบต่อ และดำเนินการตามขั้นตอนที่กล่าวมาซ้ำไปเรื่อย ๆ จนกว่าการทดสอบจะเข้าเกณฑ์การยุติการทดสอบ หากการทดสอบนั้นไม่มีแบบวัดย่อย โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะประมวลผลและรายงานผลการทดสอบให้ผู้รับการทดสอบทราบ แต่หากมีแบบวัดย่อย เมื่อเสร็จสิ้นการทดสอบในแบบวัดย่อยหนึ่งแล้ว จะเริ่มดำเนินการทดสอบในแบบวัดย่อยต่อไปในลักษณะเดียวกัน จนกว่าจะดำเนินการครบทุกแบบวัดย่อย จึงจะรายงานผลการทดสอบให้ผู้รับการทดสอบทราบ กระบวนการดังกล่าวนี้จะดำเนินการได้จะต้องอาศัยองค์ประกอบที่สำคัญของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ 5 ประการ ดังนี้

1. **คลังข้อสอบ (Item Bank)** คือ แหล่งจัดเก็บข้อมูลต่าง ๆ ที่เกี่ยวข้องกับข้อคำถาม (โจทย์และตัวเลือก) รวมทั้งคุณลักษณะต่าง ๆ ของข้อคำถามที่ผ่านการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ คลังข้อสอบควรมีจำนวนข้อคำถามมาก

เพียงพอที่จะครอบคลุมเนื้อหาทั้งหมด ซึ่งในการพัฒนาคลังข้อสอบจะเริ่มต้นจากการนำข้อคำถามที่สร้างขึ้นไปใช้ในรูปแบบกระดาษ-ดินสอ (Paper and Pencil Test) กับกลุ่มตัวอย่างจำนวนมากพอ เพื่อให้สามารถนำผลการทดสอบมาวิเคราะห์คุณภาพของข้อคำถามได้ ซึ่งขนาดของกลุ่มตัวอย่างขึ้นอยู่กับโมเดลการตอบสนองข้อสอบที่ใช้ในการวิเคราะห์ด้วย (Thompson & Weiss, 2011; Wainer *et al.*, 2000) ผลการวิเคราะห์จะช่วยให้สามารถตัดสินใจเลือกข้อคำถามที่มีความเหมาะสมเข้าสู่คลังข้อสอบได้ โดยอาจพิจารณาจากค่าพารามิเตอร์ข้อสอบ (Item Parameter) ค่าความสอดคล้องของโมเดลการตอบสนองข้อสอบกับข้อมูลเชิงประจักษ์ (Item Fit Indices) น้ำหนักองค์ประกอบ (Factor Loadings) ของข้อคำถาม ค่าสหสัมพันธ์เศษเหลือ (Residual Correlations) ที่ได้จากการวิเคราะห์องค์ประกอบเชิงยืนยัน การแจกแจงความถี่ (Frequency Distribution) ของผลคะแนนรวมของผู้รับการทดสอบ ค่าสหสัมพันธ์ปรับแก้ระหว่างข้อคำถามกับแบบวัด (Adjusted Item-Total Correlations) โมโนโทนิซิตี (Monotonicity) ฯลฯ (Fliege *et al.*, 2005; Kocalevant *et al.*, 2009; Olino *et al.*, 2012; Pilkonis *et al.*, 2011; Walter, Becker, & Bjorner, 2007)

คลังข้อสอบเป็นปัจจัยสำคัญที่ทำให้การทดสอบดำเนินไปได้อย่างมีประสิทธิภาพ ดังนั้นการเลือกข้อคำถามที่มีคุณภาพเข้าสู่คลังข้อสอบจึงเป็นสิ่งที่ละเลยไม่ได้ ในคลังข้อสอบอาจมีรูปแบบข้อคำถามที่แตกต่างกันอยู่ในคลังเดียวกันได้ เช่น ข้อคำถามแบบหลายตัวเลือก (Multiple Choice Questions) แบบถูกผิด (True/False Questions) แบบมาตราประมาณค่า (Graded

Response Format) โดยปกติแล้วหากข้อคำถามมีการตรวจให้คะแนนแบบสองค่า (เช่น 0 หรือ 1) ควรมีข้อคำถามอยู่ในคลังข้อสอบไม่ต่ำกว่า 100 ข้อ แต่หากข้อคำถามมีการตรวจให้คะแนนมากกว่าสองค่า จำนวนข้อคำถามในคลังอาจมีจำนวนน้อยกว่านี้ได้ (Embretson & Reise, 2000)

2. การเริ่มต้นการทดสอบ (Starting the First Item) เป็นการกำหนดเกณฑ์ในการเริ่มต้นการทดสอบ เนื่องจากในการทดสอบครั้งแรกจะยังไม่มีข้อมูลระดับคุณลักษณะแฝงของผู้รับการทดสอบ ในส่วนนี้ผู้วิจัยจะต้องเป็นผู้พิจารณาว่า จะใช้แนวทางใดในการกำหนดเกณฑ์เริ่มต้นการทดสอบ ซึ่งหากไม่มีข้อมูลอื่นประกอบการพิจารณา โดยปกติจะนิยมใช้ค่าเฉลี่ยคุณลักษณะแฝงของประชากร (Average Theta) ซึ่งมีค่าเท่ากับ 0.00 เป็นจุดเริ่มต้นในการเลือกข้อคำถามข้อแรกให้เหมาะกับระดับคุณลักษณะแฝงดังกล่าว (Becker *et al.*, 2008; Fliege *et al.*, 2005; Kirisci *et al.*, 2012; Kocalevent *et al.*, 2009; Smits, Cuijpers, & Straten, 2011) อย่างไรก็ตาม วิธีกรรมวิธีนี้มีข้อจำกัดคือ ในการทดสอบแต่ละครั้งผู้รับการทดสอบเกือบทุกคนจะได้รับข้อคำถามข้อเดียวกัน จึงอาจทำให้ข้อคำถามข้อนั้นมีอัตราการใช้ข้อสอบซ้ำ (Item Exposure) มากกว่าข้ออื่น ๆ ดังนั้น อาจกำหนดระดับคุณลักษณะแฝงเป็นช่วงแทนการกำหนดเพียงค่าเดียว ตัวอย่างเช่น กำหนดค่าคุณลักษณะแฝงให้อยู่ระหว่าง -0.5 ถึง +0.5 แล้วใช้กระบวนการสุ่มเลือกข้อคำถามที่มีค่าพารามิเตอร์ตำแหน่ง (Location Parameter) อยู่ในช่วงดังกล่าวมาเป็นข้อคำถามข้อแรก แต่หากมีข้อมูลการทดสอบก่อนหน้านี้ หรือข้อมูลอื่นที่เกี่ยวข้อง เช่น แรงจูงใจ เศรษฐฐานะทางสังคม เกรดเฉลี่ย ตัวชี้วัดระดับความสามารถ ผลการสอบ

วิชาซึ่งจะสามารถนำข้อมูลดังกล่าวมาประกอบการกำหนดระดับคุณลักษณะแฝงเริ่มต้นของผู้รับการทดสอบได้ (Thompson & Weiss, 2011; Wainer *et al.*, 2000)

3. การคัดเลือกข้อคำถามข้อถัดไป (Selecting the Next Item) เป็นการกำหนดวิธีการในการเลือกข้อคำถามจากคลังข้อสอบมาใช้ในกระบวนการทดสอบ สามารถจำแนกวิธีการเลือกข้อคำถามได้เป็นสองกลุ่มคือ แบบที่ 1 แบบที่ไม่มีการควบคุม (Unconstrained Item Selection Methods) มีสองแนวทางหลัก ได้แก่ 1) การเลือกข้อคำถามที่ให้ข้อมูลสารสนเทศสูงที่สุด (Maximum Information) และ 2) การเลือกข้อคำถามที่ทำให้เกิดความแม่นยำในการประมาณค่าสูงที่สุด (Maximum Expected Precision) (Thompson & Weiss, 2011; Wainer *et al.*, 2000) และแบบที่ 2 แบบที่มีการควบคุม (Constrained Item Selection Methods) จะมีการพิจารณาประเด็นอื่น ๆ ร่วมด้วย เช่น การควบคุมอัตราการใช้ข้อสอบซ้ำ (Exposure Control) และการควบคุมสัดส่วนเนื้อหาข้อคำถาม (Content Balancing) (Embretson & Reise, 2000) เป็นต้น

4. การประมาณค่าคุณลักษณะแฝง (Scoring) สามารถประมาณค่าได้โดยอาศัยทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ แต่วิธีการของทฤษฎีการตอบสนองข้อสอบได้รับความนิยมเพราะสามารถประมาณค่าคุณลักษณะแฝงในแต่ละตำแหน่งได้มีประสิทธิภาพมากกว่า วิธีการประมาณค่าคุณลักษณะแฝงที่ส่วนใหญ่นิยมใช้มี 3 วิธี ได้แก่ 1) Maximum Likelihood (ML) 2) Maximum a Posteriori (MAP) และ 3) Expected a Posterior (EAP) (De Ayala, 2009)

5. เกณฑ์การยุติการทดสอบ (Termination Criterion) สามารถเลือกใช้ได้หลากหลายวิธี เช่น

- 1) กำหนดจำนวนข้อคำถามที่จะทดสอบไว้แน่นอน
- 2) ยุติเมื่อคุณลักษณะแฝงที่ประมาณค่าได้มีการเปลี่ยนแปลงเพียงเล็กน้อยหรือไม่เปลี่ยนแปลงเลย
- 3) ยุติเมื่อความคลาดเคลื่อนมาตรฐานจากการประมาณค่า (Standard Error of Estimation; SEE) ต่ำสุดที่ยอมรับได้ถึงเกณฑ์ที่กำหนด และ 4) ยุติเมื่อข้อคำถามที่เหลืออยู่ในคลังให้ข้อมูลสารสนเทศได้น้อย (Thompson & Weiss, 2011; Wainer *et al.*, 2000)

สถานการณ์การพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดด้านเจตพิสัย

ในปัจจุบันมีการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดด้านเจตพิสัยมากขึ้น งานวิจัยแต่ละฉบับจะมีทั้งส่วนที่คล้ายคลึงกันและแตกต่างกันไป ซึ่งนักวิจัยจะเป็นผู้ออกแบบและเลือกกระบวนการและวิธีการให้เหมาะสมกับการวัดในบริบทนั้น ผู้เขียนได้สรุปรายละเอียดของตัวอย่างงานวิจัยโดยจำแนกประเด็นสำคัญต่าง ๆ ที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ไว้ดังตารางที่ 1

ตารางที่ 1 สรุปตัวอย่างงานวิจัยการวัดทางด้านเจตพิสัยในรูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

ผู้วิจัย	คุณลักษณะแฝง	คลังข้อสอบ	IRT Model	การเริ่มต้น	การเลือกข้อถัดไป	การประมาณค่าคุณลักษณะแฝง	การยุติการทดสอบ
Fliege <i>et al.</i> (2005)	ภาวะซึมเศร้า	64	GPCM	Theta mean = 0.00	MI	EAP	SEE \leq 0.32
Becker <i>et al.</i> (2008)	ความวิตกกังวล	50	GPCM	Theta mean = 0.00	MI	EAP	SEE \leq 0.32
Kocalevent <i>et al.</i> (2009)	ความเครียด	69	GPCM	Theta mean = 0.00	MI	EAP	SEE \leq 0.32
Smits, Cuijpers, & Straten (2011)	ภาวะซึมเศร้า	20	GRM	Theta mean = 0.00	MI	MAP	SEE \leq 0.30, 0.40, 0.50, 0.60, 0.70, 0.80
Vogels, Jacobusse, & Reijneveld (2011)	ปัญหาพฤติกรรม	190	PCM	Theta mean = 0.00	Fisher's information Index	Bayesian Approach with a Right-Skewed Log Normal Prior	ครบจำนวนข้อคำถามที่กำหนด และมีความแม่นยำในการประมาณค่ามากกว่าหรือเท่ากับ 95%

ตารางที่ 1 สรุปตัวอย่างงานวิจัยการวัดทางด้านเจตพิสัยในรูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (ต่อ)

ผู้วิจัย	คุณลักษณะ แฝง	คลัง ข้อสอบ	IRT Model	การเริ่มต้น	การเลือก ข้อถัดไป	การประมาณค่า คุณลักษณะ แฝง	การยุติการ ทดสอบ
Smits, Zitman, Cuijpers, Hollander- Gijsman, & Carlier (2012)	ภาวะซึมเศร้า	22	GRM	Theta mean = 0.00	MI	MAP	SEE \leq 0.20, 0.30, 0.40, 0.50, 0.60
Zheng, Chang, & Chang (2013)	สุขภาพ	36	Bifactor GRM	Theta mean = 0.00	MI	ML และ EAP	ครบจำนวนข้อ คำถามที่กำหนด
Kirisci <i>et al.</i> (2012)	การโกหก	65	GRM	The median trait level	item closest median trait level	EAP	SEE \leq 0.30 และ กำหนดจำนวนข้อ คำถามน้อยที่สุด 8 ข้อและมากที่สุด 20 ข้อ
Sakolkijrungrroj <i>et al.</i> (2015a)	ความสุข	67	GRM	Theta mean = 0.00	MI	EAP	SEE \leq 0.32

หมายเหตุ : PCM คือ Partial Credit Model, GPCM คือ Generalized Partial Credit Model, GRM คือ Graded Response Model, MI คือ Maximum Information, EAP คือ Expected a Posteriori, MAP คือ Maximum a Posteriori, ML คือ Maximum Likelihood, SEE คือ Standard Error of Estimation

จากการทบทวนงานวิจัยที่เกี่ยวข้องกับการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ในมิติการวัดด้านเจตพิสัย ตามตารางที่ 1 พบว่า การพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ส่วนใหญ่เกิดขึ้นในบริบทของสถานพยาบาล ที่เน้นพัฒนาแบบประเมินและแบบคัดกรองทางด้านสุขภาพกายและจิตใจ เช่น ภาวะซึมเศร้า ความวิตกกังวล ความเครียด ปัญหาพฤติกรรม เป็นต้น ในขณะที่มีการศึกษาเพียงส่วนน้อยที่สนใจในมุมมองของจิตวิทยาเชิงบวก เช่น ความสุข เป็นต้น นอกจากนี้ยังพบว่า คลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ด้านเจตพิสัยนั้นจะมีขนาดเล็กกว่าการทดสอบโดย

ทั่วไป ซึ่งคลังข้อสอบที่มีขนาดเล็กที่สุดมีข้อคำถามจำนวน 20 ข้อเท่านั้น ทางด้านโมเดลการตอบสนองข้อสอบที่ใช้ในการประมาณค่าพารามิเตอร์ข้อสอบมีสองกลุ่มหลักคือ โมเดลการตอบสนองข้อสอบแบบเอกมิติสำหรับแบบวัดที่มีการตรวจให้คะแนนมากกว่าสองค่า (ได้แก่ PCM, GPCM, GRM) และโมเดลการตอบสนองข้อสอบแบบพหุมิติ (ได้แก่ Bifactor GRM) ส่วนการเริ่มต้นการทดสอบส่วนใหญ่จะกำหนดระดับคุณลักษณะแฝงเริ่มต้นให้มีค่าเท่ากับ 0.00 และเลือกข้อคำถามข้อถัดไป โดยพิจารณาเลือกข้อที่ให้ค่าสารสนเทศสูงที่สุด (Maximum Information) ณ ตำแหน่งคุณลักษณะแฝงที่สนใจ เมื่อพิจารณาวิธีการที่ใช้ใน

การประมาณค่าคุณลักษณะแฝงพบว่า ส่วนใหญ่นิยมใช้วิธี EAP อย่างไรก็ตามงานวิจัยเลือกใช้การประมาณค่าคุณลักษณะแฝงด้วยวิธี MAP และ ML ด้วย ในส่วนของการกำหนดเกณฑ์การยุติการทดสอบนั้น จะเน้นใช้ความคลาดเคลื่อนมาตรฐานจากการประมาณค่าต่ำที่สุดที่ยอมรับได้เป็นเกณฑ์ในการยุติการทดสอบ แต่มีบางงานวิจัยที่ใช้วิธีการกำหนดจำนวนข้อคำถามต่ำสุดและสูงสุดที่ใช้ในการทดสอบเป็นเกณฑ์ในการยุติการทดสอบด้วยเช่นกัน

จากการทบทวนงานวิจัยที่เกี่ยวข้องพบว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีประสิทธิภาพในการทดสอบที่ดี เนื่องจากกระบวนการทดสอบสามารถกำหนดค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าเป็นเกณฑ์ในการยุติการทดสอบได้ ซึ่งค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่ามีความสัมพันธ์โดยตรงกับความแม่นยำในการทดสอบ (Precision) หากค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่ามีค่าต่ำ การทดสอบจะมีความแม่นยำสูง ดังนั้น ผู้วิจัยจึงสามารถกำหนดได้ว่าต้องการให้มีความแม่นยำในการทดสอบมากน้อยเพียงใด นอกจากนี้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ยังสามารถประมาณค่าคุณลักษณะแฝงได้โดยไม่ต้องใช้ผลรวมคะแนนที่ได้จากข้อคำถามในมาตรวัดทั้งหมด ผู้รับการทดสอบแต่ละคนจึงไม่ต้องทดสอบด้วยข้อคำถามชุดเดียวกัน และไม่ต้องทดสอบด้วยข้อคำถามทั้งหมดในคลัง ทำให้ข้อคำถามที่ใช้มีจำนวนลดลงจึงช่วยลดความเหนื่อยล้าในการทดสอบได้เป็นอย่างดี นอกจากนี้งานวิจัยหลายฉบับได้ศึกษาความสัมพันธ์ระหว่างผลการทดสอบที่ได้จากการทดสอบด้วยโปรแกรม

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ กับผลการทดสอบที่ได้จากแบบวัด รูปแบบกระดาษ-ดินสอ ปรากฏว่า ส่วนใหญ่มีความสัมพันธ์กันในระดับสูง ซึ่งแสดงให้เห็นว่านอกจากโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะมีประสิทธิภาพในการทดสอบที่ดีแล้วยังมีความตรงในการวัดที่ดีอีกด้วย

ความท้าทายและมุมมองในการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการวัดด้านเจตพิสัย

ประโยชน์คุณูปการของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ทำให้มีผู้สนใจศึกษา และ มุ่งที่จะปรับเปลี่ยนการทดสอบรูปแบบเดิมมาเป็นการทดสอบรูปแบบใหม่จำนวนมาก อย่างไรก็ตาม การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เป็นการทดสอบที่มีความซับซ้อนและต้องอาศัยความรู้และความสามารถจากผู้เชี่ยวชาญในหลากหลายสาขา ซึ่งการพัฒนานั้นไม่ใช่เรื่องง่าย ในต่างประเทศมีการศึกษาและพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ในการวัดมิติทางด้านพุทธิพิสัยมานานหลายทศวรรษ ทำให้องค์ความรู้และความเข้าใจในการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดความสามารถมีอยู่หลากหลายและกระจ่างชัดขึ้นเป็นลำดับ ขณะที่มิติทางด้านเจตพิสัยนั้น เพิ่งเริ่มต้นมาได้สักระยะหนึ่ง แม้จะพอมองเห็นแนวทางในการพัฒนา แต่องค์ความรู้และความกระจ่างชัดในมิติการวัดด้านนี้ยังมีจำกัด

ในส่วนของประเทศไทยนั้น การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เป็นประเด็นหนึ่งที่ได้รับความสนใจจากนักวัดผลทางการศึกษา โดยมีการ

ศึกษาทั้งในเชิงเทคนิควิธีการ เช่น การพัฒนาเทคนิค การเลือกข้อสอบ (Sukhanonsawat, Chadcham, & Chinnasarn, 2012) การประเมินประสิทธิภาพ การทดสอบ (Homjan, Thayreakum, & Kumpon, 2016) การศึกษาวิธีการประเมินความสามารถ วิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบ (Promsit, Suwannoi, & Maneelek, 2012) รวมทั้งการศึกษาในเชิงการพัฒนาแบบทดสอบจากรูปแบบกระดาษ-ดินสอไปสู่การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อวัดความสามารถด้านต่าง ๆ เช่น การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการจัดสอบ O-NET (Sittiparu, Tinnaworn, & Sukhanonsawat, 2016) การวิจัยทักษะการอ่านภาษาอังกฤษ โดยประยุกต์โมเดลลำดับชั้นคุณลักษณะและการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Khumyoo, Chadcham, & Chinnasarn, 2012) การพัฒนาแบบทดสอบปรับเหมาะโดยใช้คอมพิวเตอร์สาระเทคโนโลยีสารสนเทศและการสื่อสาร (Siribenchana, Pranee, & Seekhieo, 2013) อย่างไรก็ตาม จะเห็นได้ชัดว่า ในประเทศไทยนั้นให้ความสนใจในการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อวัดความสามารถ ซึ่งผู้ที่ศึกษาส่วนใหญ่จะเป็นนักวัดผลทางการศึกษาซึ่งมีความรู้ความสามารถทางด้านการวัดและประเมินผลทางการศึกษาเป็นอย่างดี ในขณะที่การพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทางด้านเจตพิสัยนั้นมียุ่จำกัด อาจเนื่องมาจากการวัดด้านนี้ต้องอาศัยผู้ที่มีความสนใจในหลากหลายศาสตร์ร่วมกัน ทั้งทางด้านการวัดผลและจิตวิทยา ดังนั้น ผู้เขียนซึ่งพอมิพื้นความรู้ในสองศาสตร์นี้จึงได้สนใจและศึกษาการพัฒนาการทดสอบแบบปรับเหมาะ

ด้วยคอมพิวเตอร์สำหรับการวัดความสุขขึ้น (Sakolkijrungrroj, Chadcham, & Sudhasani, 2015a; Sakolkijrungrroj, Chadcham, & Sudhasani, 2015b; Sakolkijrungrroj, Chadcham, & Sudhasani, 2016) อย่างไรก็ตาม การศึกษาวิจัยที่ผ่านมาทำให้เห็นถึงความท้าทายต่าง ๆ ที่เกิดขึ้นระหว่างการศึกษาวิจัย ผู้เขียนจึงต้องการนำเสนอข้อมูลที่ได้จากการถ่วงน้ำหนักการประเมินในการทำวิจัย โดยคาดหวังให้เป็นแหล่งข้อมูลหรือข้อควรพิจารณาสำหรับผู้สนใจในอนาคต การเสนอเนื้อหาส่วนนี้จำแนกตามประเด็นสำคัญต่าง ๆ ดังนี้

1. คลังข้อสอบ

การพัฒนาคลังข้อสอบสำหรับการวัดทางด้านเจตพิสัยมีความท้าทายเกิดขึ้นหลายประการ เริ่มต้นตั้งแต่ความยากในการสร้างข้อคำถามให้มีจำนวนมากพอ เพราะการวัดด้านนี้ไม่สามารถนำเทคนิคการสร้างข้อสอบทางด้านพุทธิพิสัยมาใช้ได้ นอกจากนี้ ข้อคำถามทุกข้อที่จะนำเข้าคลังข้อสอบจะต้องผ่านการวิเคราะห์คุณภาพก่อน ดังนั้น ข้อคำถามที่สร้างขึ้นใหม่จึงต้องถูกนำไปใช้กับกลุ่มตัวอย่างที่มีขนาดใหญ่พอสำหรับการวิเคราะห์ข้อมูล และกลุ่มตัวอย่างต้องมีระดับคุณลักษณะแฝงกระจายครอบคลุมตลอดช่วงคุณลักษณะแฝงที่มุ่งวัด ซึ่งการเก็บข้อมูลทำได้ยาก เพราะกลุ่มตัวอย่างต้องเป็นตัวแทนของประชากรทั้งหมด ตัวอย่างเช่น หากต้องการวัดความสุข กลุ่มตัวอย่างจะต้องประกอบด้วยคนที่มีความสุขระดับต่าง ๆ ตั้งแต่ น้อยที่สุด ปานกลาง จนถึงมากที่สุด (ในการวัดความสามารถอาจใช้เกรดเฉลี่ยมาเป็นแนวทางในการกำหนดกลุ่มตัวอย่างให้มีระดับความสามารถกระจายได้ แต่ในการวัดทางด้านเจตพิสัยนั้นหาตัวช่วยได้ยาก) ในส่วนของการวิเคราะห์คุณภาพ

ของข้อคำถามนับได้ว่า เป็นความท้าทายที่สำคัญประการหนึ่ง เนื่องจากเกณฑ์ที่ใช้ในการพิจารณาคุณภาพมีหลากหลายเกณฑ์ และการที่ข้อคำถามจะผ่านเกณฑ์ต่าง ๆ ไม่ใช่เรื่องง่าย ซึ่งการศึกษาของ Sakolkijrunroj *et al.* (2015b) เรื่องการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับมาตรวัดความสุขของคนไทยพบว่า มีข้อคำถามที่ไม่ผ่านเกณฑ์การพิจารณาคุณภาพจำนวนมาก จากการสร้างข้อคำถามเริ่มต้นจำนวน 191 ข้อ แต่สุดท้ายแล้วคงเหลือข้อคำถามในคลังเพียง 61 ข้อ วิธีการที่ง่ายที่สุดในการจัดการกับปัญหา คือ การคัดข้อคำถามดังกล่าวออกจากคลังข้อสอบ แต่การทำเช่นนั้นอาจส่งผลกระทบต่อประเด็นอื่น ๆ ตามมาได้ ดังนั้น ผู้ที่สนใจศึกษาควรมีการเตรียมพร้อมสำหรับการรับมือปัญหาดังกล่าวไว้ล่วงหน้า เช่น การพิจารณายุบตัวเลือกให้มีจำนวนลดลง จากมาตรประมาณค่า 5 ระดับ อาจต้องลดลงเหลือ 3 หรือ 4 ระดับขึ้นอยู่กับผลการตอบของผู้รับการทดสอบด้วย วิธีการนี้เป็นอีกวิธีหนึ่งที่ช่วยบริหารจัดการปัญหาดังกล่าวได้

2. โมเดลการตอบสนองข้อสอบ

โมเดลการตอบสนองข้อสอบที่ใช้วิเคราะห์ข้อคำถามสำหรับแบบวัดทางด้านเจตพิสัยนั้น มีให้เลือกใช้หลากหลายโมเดล ซึ่งผู้วิจัยต้องเลือกโมเดลให้มีความเหมาะสมกับลักษณะข้อคำถามแบบวัด และประเด็นอื่นที่เกี่ยวข้อง อย่างไรก็ตาม แบบวัดและข้อคำถามฉบับหนึ่งอาจสามารถเลือกใช้โมเดลในการวิเคราะห์ได้มากกว่าหนึ่งโมเดล โมเดลแต่ละโมเดลจะมีความเฉพาะ มีจุดเด่นและข้อจำกัดที่แตกต่างกัน การเลือกโมเดลที่เหมาะสมมิใช่พิจารณาจากค่าสถิติเท่านั้น แต่ควรคำนึงถึงประเด็นทางการนำไปใช้ บริบทการทดสอบ ความคุ้ม

ค่า และประสิทธิภาพพร้อมด้วย เพราะการเลือกโมเดลการตอบสนองข้อสอบไม่ได้ส่งผลเพียงแค่ผลการวิเคราะห์คุณภาพเท่านั้น แต่ยังส่งผลต่อกระบวนการทดสอบและอัลกอริทึมที่ใช้ในการทดสอบด้วย

3. การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดทางด้านพุทธิพิสัยค่อนข้างมีความชัดเจน และมีผู้วิเคราะห์สูตรอัลกอริทึมต่าง ๆ ที่ใช้ในเขียนโปรแกรมเอาไว้เรียบร้อยแล้ว ผู้พัฒนาโปรแกรมสามารถหยิบสูตรดังกล่าวมาใช้ได้ไม่ยาก ต่างจากการวัดทางด้านเจตพิสัย ซึ่งยังมีข้อจำกัดที่ผู้พัฒนาโปรแกรมจะต้องวิเคราะห์สูตรที่ใช้ในกระบวนการต่าง ๆ ด้วยตนเอง ซึ่งบางครั้งสูตรที่มีในตำรา หรืองานวิจัยต่าง ๆ ไม่ได้ระบุรายละเอียดมากพอที่ผู้วิจัยจะสามารถดึงมาใช้ได้ทันที การวิเคราะห์สูตรจึงไม่ใช่เรื่องง่าย เพราะไม่เพียงแต่ผู้วิจัยจะต้องเข้าใจ แต่ยังต้องสื่อสารกับผู้เขียนโปรแกรมให้เข้าใจตรงกันด้วย โดยผู้วิจัยต้องสามารถอธิบายขั้นตอนการคำนวณต่าง ๆ ได้อย่างละเอียด และต้องมีการตรวจสอบความถูกต้องของกระบวนการทดสอบเป็นระยะด้วย

4. ประสิทธิภาพและความคุ้มค่าของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

แม้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะมีประโยชน์และจุดเด่นที่ดึงดูดใจนักวิจัยจำนวนมาก แต่การพัฒนาโปรแกรมขึ้นมาั้นไม่ใช่เรื่องง่าย เพราะการดำเนินการล้วนต้องอาศัยงบประมาณ ระยะเวลา และบุคลากรวิชาชีพต่าง ๆ จำนวนมาก ดังนั้น ก่อนที่จะตัดสินใจปรับเปลี่ยนรูปแบบการทดสอบจากรูปแบบกระดาษ-

ดินสอ มาเป็นการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เป็นต้น ผู้วิจัยควรมีการพิจารณาถึงความพร้อมด้านผู้เชี่ยวชาญทางการวัดผล จำนวนข้อคำถามที่จะนำมาจัดทำคลังข้อสอบ ทรัพยากรสำหรับการพัฒนาระบบการทดสอบ ประโยชน์ที่ได้จากการพัฒนาระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในด้านของมูลค่า เช่น จำนวนข้อสอบที่ใช้ในการทดสอบ ระยะเวลาที่ใช้ในการทดสอบ ค่าใช้จ่ายต่าง ๆ ที่เกี่ยวข้อง เป็นต้น และประโยชน์ที่ได้จากการพัฒนาระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในด้านของประสิทธิภาพในการทดสอบ เช่น ความแม่นยำในการทดสอบความปลอดภัยของข้อคำถาม เป็นต้น การพิจารณาอาจใช้เทคนิค Monte Carlo Simulation ซึ่งเป็นการจำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในเงื่อนไขต่าง ๆ โดยใช้กลุ่มผู้สอบ และข้อสอบที่สมมติขึ้น ซึ่งผลที่ได้จากการทดสอบสามารถนำมาเปรียบเทียบเพื่อช่วยในการตัดสินใจได้ อย่างไรก็ตาม Monte Carlo Simulation ต้องใช้โปรแกรมเฉพาะในการจำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์โดยโปรแกรมที่จำเป็น

มีสองส่วน คือ 1) โปรแกรมสำหรับการจำลองข้อมูลผลการทดสอบ เช่น WINGEN (Han, 2007) และ PARDSIM (Yoes, 1997) และ 2) โปรแกรมสำหรับการจำลองสถานการณ์การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เช่น FireStar (Choi, 2009) หรือ CATSim (Weiss & Guyer, 2010) เป็นต้น

ประเด็นที่น่าเสนาในที่นี่เป็นเพียงบางส่วนของปัญหาและความท้าทายที่ผู้วิจัยต้องเผชิญระหว่างการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในการดำเนินการจริงยังมีสถานการณ์ที่ต้องเผชิญอีกมาก อย่างไรก็ตาม ความก้าวหน้าทางวิทยาการและการวิจัยทำให้องค์ความรู้ในด้านนี้มีความชัดเจนมากขึ้นเรื่อย ๆ นักวิจัยต่างพยายามหาทางลดข้อจำกัดต่าง ๆ และพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ให้มีความสมบูรณ์มากยิ่งขึ้น หากมีนักวิจัยให้ความสนใจในการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในมิติการวัดด้านเจตพิสัยมากขึ้น องค์ความรู้ทางการวัดในประเทศไทยจะมีความก้าวหน้าและแผ่ขยายอย่างไม่หยุดยั้ง ซึ่งจะก่อให้เกิดประโยชน์ต่อสังคมสืบไป 📖

References

- Becker, J., Fliege, H. Kocalevent, R. D., Bjorner, J. B, Rose, M., Walter, O. B., & Klapp, B. F. (2008). Functioning and Validity of a Computerized Adaptive Test to Measure Anxiety (A-Cat). *Depression and Anxiety, 25*, 182-194.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP Estimation of Ability in a Microcomputer Environment. *Applied Psychological Measurement, 6*(4), 431-444.
- Choi, S. W. (2009). Firestar: Computerized adaptive testing (CAT) simulation program for polytomous IRT Models (Computer software). *Applied Psychological Measurement, 33*, 644-645.

- De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York: The Guildford Press.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Fliege, H., Becker, J., Walter, O. B., Bjorner, J. B., Klapp, B. F., & Rose, M. (2005). Development of a computer-adaptive test for depression (D-CAT). *Quality of Life Research, 14*, 2277-2291.
- Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., Bhaumik, D. K., Stover, A., Bock, R. D., & Immekus, J. C. (2008). Using Computerized Adaptive Testing to Reduce the Burden of Mental Health Assessment. *Psychiatric Services, 59*(4), 361-368.
- Han, K. T. (2007). WinGen: Windows software that generates IRT parameters and item responses. *Applied Psychological Measurement, 31*(5), 457-459.
- Homjan, S., Thayreakum, S., & Kumpon, B. (2016). The efficiency of a Computerized Adaptive Testing (CAT) by applying multidimensional item response models. *Social Sciences Research and Academic Journal, 11*(31), 73-88. (in Thai)
- Khomyoo, T., Chadcham, S., & Chinnasarn, K. (2012). Diagnostic of English reading skills by applying the attribute hierarchy model and computerized adaptive testing. *Research Methodology & Cognitive Science, 10*(2), 54-69. (in Thai)
- Kirisci, L., Tarter, R., Reynolds, M., Ridenour, T., Stone, C., & Vanyukov, M. (2012). Computer adaptive testing of liability to addiction: Identifying individuals at risk. *Drug and Alcohol Dependence, 123S*, S79-S86.
- Kocalevent, R. D., Rose, M., Becker, J., Walter, O. B., Fliege, H., Bjorner, J. B., Kleiberg, D., & Klapp, B. F. (2009). An evaluation of patient-reported outcomes found computerized adaptive testing was efficient in assessing stress perception. *Journal of Clinical Epidemiology, 62*, 278-287.
- Olino, T. M., Yu, L., Klein, D. N., Rohde, P., Seeley, J. R., Pilkonis, P. A., & Lewinsohn, P. M. (2012). Measuring depression using item response theory: an examination of three measures of depressive symptomatology. *International Journal of Methods in Psychiatric Research, 21*(1), 76-85.
- Patient-Reported Outcome Measurement Information System (PROMIS). (2013, November). *PROMIS Overview*. Retrieved November 25, 2013, from the PROMIS Dynamic Tools to Measure Health Outcomes for the Patient Perspective Web site: <http://www.nihpromis.org/about/overview>

- Pilkonis, P. A., Choi, S. W., Reise, S. P., Stover, A. M., Riley, W. T., & Cella, D. (2011). Item Banks for Measuring Emotional Distress from the Patient-Reported Outcomes Measurement Information System (PROMIS): Depression, Anxiety, and Anger. *Assessment, 18*(3), 263-283.
- Promsit, P., Suwannoi, P., & Maneelek, R. (2012). A study of ability estimation, item selection methods and test termination of multidimensional computerized adaptive testing. *Journal of Education Khon Kaen University, 35*(1), 38-47. (in Thai)
- Rose, M., Bjorner, J. B., Becker, J., Fries, J. F., & Ware, J. E. (2006). Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *Journal of Clinical Epidemiology, 61*, 17-33.
- Sakolkijrungrroj, S., Chadcham, S., & Sudhasani, S. (2015a). The development of computerized adaptive testing program for Thai happiness scale. *Research Methodology & Cognitive Science, 13*(1), 1-17. (in Thai)
- _____. (2015b). *The Item Bank Construction for Thai Happiness Scale (Computerized Adaptive Testing Version)*. In Burapha University International Conference 2015 Committee (Eds.), *Burapha University International Conference 2015: Moving Forward to a Prosperous and Sustainable Community, July 10-12, 2015 Bangsaen Heritage Hotel Chon Buri* (pp.33-45). Thailand: Burapha University. Retrieved from <http://www.buuconference.buu.ac.th/buureg/getfile.php?file=file/BUU2015/proceedings.pdf>
- _____. (2016). The efficiency of computerized adaptive testing program for Thai happiness scale. *Educational Measurement Mahasarakham University, 22*(2), 321-331. (in Thai)
- Siribench, S., Pranee, C., & Seekhieo, D. (2013). A development of computerized adaptive testing of information and communication technology matter by estimation of ability by bayesian for students matthayom 3 students. *Social Sciences Research and Academic Journal, 8*(22), 87-102. (in Thai)
- Sittiparu, J., Tinnaworn, P., & Sukhanonsawat, S. (2016). Development of the computerized adaptive testing program for O-NET at the grade nine level. *Educational Measurement Mahasarakham University, 22*(1), 47-62. (in Thai)
- Smits, N., Cuijpers, P., & Straten, A. V. (2011). Applying computerized adaptive testing to the CES-D scale: A simulation study. *Psychiatry Research, 188*, 147-155.

- Smits, N., Zitman, F. G., Cuijpers, P., Hollander-Gijsman, M. E., & Carlier, I. V. (2012). A proof of principle for using adaptive testing in routine Outcome Monitoring: the efficiency of the Mood and Anxiety Symptoms Questionnaire -Anhedonic Depression CAT. *BMC Medical Research Methodology*, 12(4).
- Sukhanonsawat, S., Chadcham, S., & Chinnasarn, K. (2012). Development of the next item selection procedure using decision theory in computerized adaptive testing. *Research Methodology & Cognitive Science*, 10(2), 71-85. (in Thai)
- Thompson, N. A., & Weiss, D. J. (2011). A Framework for the Development of Computerized Adaptive Tests. *Practical Assessment, Research & Evaluation*, 16(1). Available online: <http://pareonline.net/getvn.asp?v=16&n=1>.
- Vogels, A. G., Jacobusse, G. W., & Reijneveld, S. A. (2011). An accurate and efficient identification of children with psychosocial problems by means of computerized adaptive testing. *BMC Medical Research Methodology*, 11, 1-9.
- Wainer, H., Dorans, N. J., Daniel, E., Flaugher, R., Green, B. F., Mislevy, R. J., Steinberg, L., & Thissen, D. (2000). *Computerized Adaptive Testing: A Primer* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Walter, O. B., Becker, J., Bjorner, J. B., Fliege, H., Klapp, B. F., & Rose M. (2007). Development and evaluation of a computer adaptive test for anxiety (Anxiety-CAT). *Quality of Life Research*, 16, 143-155.
- Weiss, D. J., & Guyer, R. (2012). *Manual for CATSim: Comprehensive simulation of computerized adaptive testing*. St. Paul MN: Assessment Systems Corporation.
- Yoes, M. (1997). *PARDSIM* (Computer software). Saint Paul, MN: Assessment Systems Corporation.
- Zheng, Y., Chang, C. H., & Chang, H. H. (2013). Content-balancing strategy in bifactor computerized adaptive patient-reported outcome measurement. *Quality of life research*, 22(3), 491-499.