

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing : MCAT)

ชัยมงคล ปินะสา¹ รัตนะ บัวสนธ์²

Chaimongkhon Pinasa¹ Rattana Buosonte²

¹ นิสิตปริญญาเอก หลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชานวัตกรรมทางการวัดผลการเรียนรู้

คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร

Students at Naresuan University, Doctor of Philosophy Program in Innovation of Learning

Measurement Faculty of Education Naresuan University

Corresponding author, E – mail : chaimongkhonp63@nu.ac.th

² ศาสตราจารย์ ดร. คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร

Lecturer, Ph.D. Professor, Faculty of Education Naresuan University

Corresponding author, E – mail : rattanab@nu.ac.th

Received: November 13, 2020; Revised: February 16, 2022; Accepted: April 11, 2022

บทคัดย่อ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing : MCAT) เป็นการผสมผสานแนวคิดระหว่างทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory : MIRT) กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing : CAT) ซึ่งเป็นการทดสอบที่มีประสิทธิภาพสูง มีความแม่นยำ และลดจำนวนข้อสอบ แบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) มี 5 ขั้นตอน ได้แก่ 1) การสร้างคลังข้อสอบ (Item pools) 2) การเริ่มต้นการทดสอบ (Starting point) 3) การคัดเลือกข้อสอบข้อถัดไป (Item selection) 4) การประมาณความสามารถ (Scoring algorithm) และ 5) เกณฑ์การยุติการทดสอบ (Stopping rules) โมเดลการตอบสนองข้อสอบแบบพหุมิติระหว่างข้อสอบ (Between-Items MIRT Model) นิยมใช้การประมาณค่าความสามารถของการทดสอบ แบบวิธีการประมาณค่าแบบความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) ส่วนโมเดลการตอบสนองข้อสอบแบบพหุมิติภายในข้อสอบ (Within-Items MIRT Model) นิยมใช้การประมาณค่าความสามารถด้วยวิธีของเบย์ที่ปรับเปลี่ยน (Bayesian Updating) และเกณฑ์ยุติการสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) นิยมใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard Error of estimation: SE)

คำสำคัญ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ, วิธีของเบย์ปรับเปลี่ยน

Abstract

Multidimensional computerized adaptive testing (MCAT) is a conceptual combination of Multidimensional Item Response Theory (MIRT) with Computerized Adaptive Testing (CAT), which is highly effective, accurate and reduces the number of questions. Multidimensional computerized adaptive testing has 5 steps: 1) to build the item pools 2) Starting point 3) Item selection 4) Scoring algorithm and 5) Stopping rules. Between-Items MIRT Model is the most commonly used for estimation the test, the Maximum Likelihood Estimation method, and Within-Items MIRT Model is the most commonly used for the Bayesian Updating method, stopping criteria of Multidimensional computerized adaptive testing (MCAT) is popularly used in the standard error of estimation (SE)

Keyword : Multidimensional Computerized Adaptive Testing, Bayesian Updating

บทนำ

ในปัจจุบันนี้การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ได้รับความนิยมมากทั้งในการทดสอบทางการศึกษาและจิตวิทยา และเป็นวิธีการเฉพาะที่ใช้ในการประเมินความสามารถและคุณลักษณะแฝงอื่น ๆ ในการเลือกข้อสอบจากคลังข้อสอบที่ให้ผู้สอบได้รับข้อสอบตามความสามารถที่แท้จริงของตนเอง และได้พัฒนาแบบทดสอบแบบปรับเหมาะขึ้นมาเป็นการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing: MCAT) เป็นวิธีการทดสอบที่มีการขยายแนวคิดมาจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ (Unidimensional Computerized Adaptive Testing: UCAT) ซึ่งเป็นวิธีการเฉพาะที่ใช้ในการประเมินความสามารถและคุณลักษณะแฝงอื่น ๆ ในการเลือกข้อสอบให้เหมาะกับระดับความสามารถของผู้เข้าสอบ (Frey & Seitz, 2009) เป็นการลดจำนวนข้อสอบให้ผู้สอบตอบข้อสอบและเพิ่มความแม่นยำของการวัด เมื่อเปรียบเทียบกับ การทดสอบแบบดั้งเดิมที่มีจำนวนข้อสอบคงที่ในแบบทดสอบ (Fixed Item Test: FIT) และเป็นการขยายแนวคิด จากโมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model : MIRT) ซึ่งถือว่าคุณลักษณะแฝงของบุคคลมากกว่า 1 องค์ประกอบ ส่งผลต่อการตอบข้อสอบ และการทดสอบแบบปรับเหมาะกำลังได้รับความนิยมอย่างมากในการนำไปปฏิบัติ จึงเกิดการรวมกันของทั้งสองวิธีนี้ เรียกว่า การทดสอบปรับเหมาะแบบพหุมิติ (Multidimensional Adaptive Testing : MAT) (Segall, 2010, p. 62, Reckase, 2010, p. 95)

เมื่อนำคอมพิวเตอร์มาใช้ในกระบวนการทดสอบ จึงเรียกว่า การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing: MCAT) ซึ่งวิธีการทดสอบดังกล่าวนี้ เป็นการทดสอบที่มีประสิทธิภาพในการวัดสูง ไม่ว่าจะเป็นความแม่นยำในการวัดการลดความยาวของข้อสอบ และเป็นการวัดที่ตรงกับสภาพจริงของลักษณะข้อสอบ ดังนั้น ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ จึงสามารถลดจำนวนข้อสอบลงได้มากกว่าการทดสอบด้วยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เอกมิติ (CAT) ประมาณ 30-50% และลดจำนวนข้อสอบได้มากกว่า การทดสอบแบบดั้งเดิม ที่กำหนดจำนวนข้อสอบไว้ประมาณ 70% โดยไม่สูญเสียความแม่นยำ (Frey & Seitz, 2009) และการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติมีประสิทธิภาพ ที่สูงกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ ถึง 1.3 เท่า (Frey & Seitz, 2009) อย่างไรก็ตาม ถ้ากระบวนการทดสอบได้ใช้วิธีการเลือกข้อสอบแบบสุ่ม (random) การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) จะส่งผลให้การทดสอบมีประสิทธิภาพสูงกว่าวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ (CAT) ถึง 3.7 เท่า (Frey & Seitz, 2009)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติเป็นการขยายมิติความสามารถการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ (CAT) สู่การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) ซึ่งผู้เขียนได้นำเสนอสาระสำคัญเกี่ยวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ คือ แนวคิดพื้นฐานของลักษณะพหุมิติ, การวิเคราะห์โมเดลการตอบสนองข้อสอบแบบพหุมิติ, ลักษณะของโมเดลการตอบสนองข้อสอบแบบพหุมิติ และองค์ประกอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ดังนี้

แนวคิดพื้นฐานของลักษณะพหุมิติ

ลักษณะความเป็นพหุมิตินี้มีข้อตกลงเบื้องต้นที่ว่า คุณลักษณะที่จะทำการวัดมีความสัมพันธ์กันซึ่งมีอยู่ 2 ลักษณะ คือ 1) ความเป็นพหุมิติระหว่างข้อคำถาม (between-items multidimensionality) เป็นข้อคำถามวัดคุณลักษณะแฝงเดียว มีหลายคุณลักษณะแฝงทั้งนี้แต่ละคุณลักษณะแฝง มีความสัมพันธ์กัน และ 2) ความเป็นพหุมิติภายในข้อคำถาม (within-items multidimensionality) เป็นของข้อคำถามวัดหลายคุณลักษณะแฝง มีหลายคุณลักษณะแฝง ซึ่งเป็นแนวคิดพื้นฐานของลักษณะพหุมิติ (ชัยวิชิต เขียวชนะ. 2552, เมษา นวลศรี. 2559, สุกัญญา บุญศรี. 2560) โมเดลการตอบสนองข้อสอบแบบพหุมิติถือว่าคุณลักษณะแฝงของบุคคลมากกว่า 1 องค์ประกอบ ส่งผลต่อการตอบข้อสอบ ดังนั้น พารามิเตอร์ความสามารถของผู้สอบจึงมีตั้งแต่ 2 พารามิเตอร์ขึ้นไป ดังนั้น การพิจารณาความสามารถหลายมิติของบุคคลน่าจะช่วยให้โมเดลมีความสอดคล้องกับข้อมูลได้ดียิ่งขึ้น ทั้งนี้แต่ละคุณลักษณะแฝงมีความสัมพันธ์กัน เมื่อผู้สอบใช้ความสามารถหลายมิติที่แตกต่างกัน อย่างเป็นระบบในการตอบข้อสอบย่อมมีความสัมพันธ์กับระดับความยากและอำนาจจำแนกของข้อสอบรายข้อ และมีผลต่อความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง โดยมีสัญลักษณ์ของค่าต่าง ๆ ในโมเดล ดังนี้ (ศิริชัย กาญจนวาสี, 2555 หน้า 82-83)

X_{ip} = ผลการตอบข้อสอบข้อที่ i ของบุคคล p (0 =ผิด, 1 =ถูก)

θ_{pm} = ค่าระดับความสามารถหรือคุณลักษณะแฝงของผู้สอบ p ($m \geq 2$)

b_i = ค่าความยากของข้อสอบ ข้อที่ i

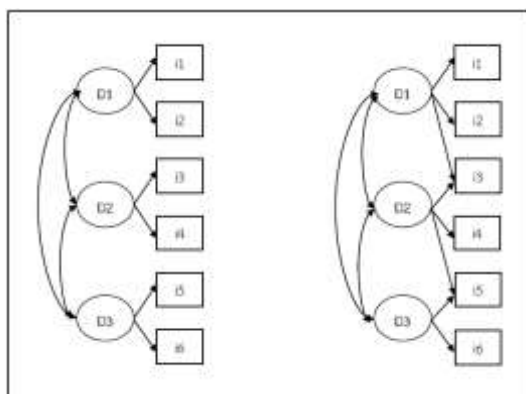
δ_i = ค่าจุดตัดความง่าย (easiness intercept) ของข้อสอบ ข้อที่ i

a_{im} = ค่าอำนาจจำแนกของข้อสอบข้อที่ i บนมิติ m

c_i = ค่าโอกาสการเดาข้อสอบได้ถูก (guessing) ของข้อสอบข้อที่ i

การวิเคราะห์โมเดลการตอบสนองข้อสอบแบบพหุมิติ

การวิเคราะห์โมเดลการตอบสนองข้อสอบแบบพหุมิติสามารถทำได้ 2 วิธี คือการวิเคราะห์โดยใช้โมเดลเชิงสำรวจ (exploratory model) และการวิเคราะห์โดยใช้โมเดลเชิงยืนยัน (confirmatory model) ซึ่งมีแนวคิดคล้ายกับการวิเคราะห์องค์ประกอบ (factor analysis) กล่าวได้ว่า Exploratory IRT Model (สุกัญญา บุญศรี, 2560) เป็นการประมาณค่าพารามิเตอร์ของข้อสอบและพารามิเตอร์ความสามารถของผู้สอบที่มีมากกว่า 1 มิติ เพื่อเพิ่มความสอดคล้องของโมเดลกับข้อมูลเชิงประจักษ์โดยจะไม่มีกำหนดจำนวนองค์ประกอบของคุณลักษณะแฝงในการวิเคราะห์ แต่จะหาจำนวนองค์ประกอบหรือมิติแฝงของบุคคลที่มีผลต่อการตอบข้อสอบ สำหรับ confirmatory IRT Model เป็นการประมาณค่าพารามิเตอร์ของข้อสอบและพารามิเตอร์ความสามารถของผู้สอบตามจำนวนองค์ประกอบของคุณลักษณะแฝงที่กำหนดไว้ล่วงหน้า (ศิริชัย กาญจนวาสี 2555, หน้า 83) ยกตัวอย่างเช่นแบบสอบที่สร้างขึ้นมุ่งวัดคุณลักษณะแฝงหรือความสามารถของบุคคล 3 ด้าน (3 มิติหรือ 3 องค์ประกอบ) ความสัมพันธ์ระหว่างคุณลักษณะของข้อสอบและความสามารถของผู้สอบที่ส่งผลต่อความน่าจะเป็นของการตอบข้อสอบได้ถูกต้อง สามารถเขียนโมเดลการวัดแบบพหุมิติได้ 3 ลักษณะ ดังนี้



A : Between-Items MIRT Model

B: Within-Items MIRT Model

(ที่มา : ชัยวิชิต เขียรชนะ, 2552)

โมเดลการตอบสนองข้อสอบแบบพหุมิติระหว่างข้อสอบ A
(Between-Items MIRT Model) และโมเดลการตอบสนองข้อสอบแบบพหุมิติ
ภายในข้อสอบ B (Within-Items MIRT Model)

จากโมเดล ในภาพ A เป็นโมเดลการตอบสนองข้อสอบแบบ 3 มิติ ซึ่งความสามารถทั้ง 3 ด้าน มีความเกี่ยวข้องสัมพันธ์กันโดยความสามารถแต่ละด้านถูกวัดระหว่างข้อสอบต่างชุดกัน (between-items) หรือคุณลักษณะแฝงแต่ละด้านถูกวัดโดยใช้ข้อสอบต่างชุดกัน ส่วนภาพ B เป็นโมเดล การตอบสนองข้อสอบแบบ 3 มิติ ซึ่งความสามารถทั้ง 3 ด้าน มีความเกี่ยวข้องสัมพันธ์กันโดยความสามารถแต่ละด้านถูกวัดภายในข้อสอบชุดเดียวกัน (within-items)

ลักษณะของโมเดลการตอบสนองข้อสอบแบบพหุมิติ

1) โมเดลการตอบสนองข้อสอบแบบพหุมิติที่มี 3 พารามิเตอร์ : โมเดลโลจิส

(Multidimensional Three-Parameter Logistic Model)

Embretson และ Reise (2000) อ้างถึงใน ศิริชัย กาญจนวาสี (2555) ได้ขยายโมเดลโลจิส 3 พารามิเตอร์แบบเอกมิติเป็นโมเดลโลจิส 3 พารามิเตอร์แบบพหุมิติ ดังนี้

$$P_i(\theta_p) = c_i + \frac{(1-c_i)}{1+e^{-(\sum_m a_{im}\theta_{pm}+\delta_i)}}$$

เมื่อ $P_i(\theta_p)$ น่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ_{pm} ($m \geq 2$) จะตอบข้อสอบได้ถูกต้อง

2) โมเดลการตอบสนองข้อสอบแบบพหุมิติที่มี 3 พารามิเตอร์: โมเดลโค้งปกติสะสม

(Multidimensional Three-Parameter Normal Ogive Model)

Bock, Gibbons และ Muraki (1998, อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) ได้นำแนวคิดของการวิเคราะห์องค์ประกอบชนิดใช้สารสนเทศทั้งหมด (factor analysis with full information) มาประยุกต์ โดยขยายโมเดลโค้งปกติสะสมพารามิเตอร์แบบเอกมิติเป็นโมเดลโค้งปกติสะสม 3 พารามิเตอร์แบบพหุมิติ

กำหนดให้ Z_{ip} เป็นศักยภาพของผู้สอบ p ในการตอบข้อสอบข้อ i โดย Z_{ip} เป็นค่าผลรวมถ่วงน้ำหนักของคุณลักษณะแฝง m องค์ประกอบตามคุณลักษณะของข้อสอบ i

$$Z_{ip} = \sum_m a_{im} \theta_{pm} + \delta_i$$

โมเดลโค้งปกติสะสม 3 พารามิเตอร์แบบพหุมิติมีสมการ เป็นดังนี้

$$P(\theta_i) = c_i + (1 - c_i) \int_{-zip}^{\infty} \frac{1}{2\pi} e^{-z^2/2} dz$$

โดย λ_{im} = น้ำหนักองค์ประกอบของข้อ i บนมิติ m

$$= \frac{a_{im}}{g_i}$$

$$b_i = \frac{\delta_i}{g_i}$$

$$\text{และ } g_i = \sqrt{1 + \sum m a_{im}^2}$$

สำหรับโมเดลโค้งปกติสะสม 3 พารามิเตอร์ค่า c_i สามารถกำหนดให้เป็นค่าอิสระหรือคงที่ของทุกข้อหรือคงที่สำหรับชุดของข้อสอบย่อยก็ได้

จากแนวคิดและความสำคัญของโมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model : MIRT) ทำให้นักการศึกษาหลายท่านเริ่มหันมาสนใจใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory) ในงานวิจัยมากขึ้น โดยได้นำวิธีการนี้มาผสมผสานเข้ากับการทดสอบปรับเหมาะแบบพหุมิติ (Multidimensional Adaptive Testing: MAT) และได้พัฒนาต่อเนื่องจนถึงปัจจุบันและเมื่อนำเทคโนโลยีคอมพิวเตอร์มาใช้ในการบวนการทดสอบ จึงเรียกว่า การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing: MCAT) ซึ่งวิธีการทดสอบดังกล่าวนี้ เป็นการทดสอบที่มีประสิทธิภาพในการวัดสูง ไม่ว่าจะเป็นความแม่นยำ ในการวัด การลดความยาวของข้อสอบ และเป็นการวัดที่ตรงกับสภาพจริงของลักษณะข้อสอบ

องค์ประกอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ เป็นการขยายมิติความสามารถของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติสู่การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ซึ่งการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิตินี้ประกอบด้วยขั้นตอน 5 ขั้นตอน ได้แก่ 1) การสร้างคลังข้อสอบ (Item pools) 2) การเริ่มต้นการทดสอบ (Initiation/Starting point) 3) การคัดเลือกข้อสอบข้อถัดไป (Item selection algorithm/Item selection) 4) การประมาณความสามารถ (Ability Estimation/Scoring algorithm) และ 5) เกณฑ์การยุติการทดสอบ (Termination criterion/Stopping rules) (Veldkamp & Matteucci, 2013; Thompson & Weiss, 2011) โดยแต่ละขั้นตอนมีรายละเอียด ดังนี้

1. คลังข้อสอบ (Item Pools)

คลังข้อสอบ (Item Pools, Item Bank) เป็นแหล่งรวบรวมข้อสอบหรือคำถามแบบทดสอบ การทดสอบด้วยแบบทดสอบดั้งเดิม ผู้สอบทั้งหมดจะได้รับข้อสอบชุดเดียวกัน แต่สำหรับการทดสอบแบบปรับเหมาะผู้สอบจะได้รับข้อสอบเหมาะสมกับความสามารถของผู้สอบแต่ละคน ดังนั้น ต้องใช้ข้อสอบจำนวนมากสำหรับสร้างคลังข้อสอบ ทำให้คลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะมีขนาดใหญ่กว่าการทดสอบดั้งเดิม (Thompson & Weiss, 2011; Babcock & Weiss, 2012, ชัยวิชิต เขียรชนะ, 2552)

2. การเริ่มต้นการทดสอบ (Starting the First Item)

การกำหนดจุดเริ่มต้นการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีหลายวิธี โดยทั่วไปจะใช้ข้อสอบข้อแรกที่มีค่าความยากปานกลาง จัดให้กับผู้สอบแต่ละคน เมื่อไม่ทราบค่าสารสนเทศเบื้องต้นหรือระดับความสามารถ

ของผู้สอบ จึงจำเป็นต้องพิจารณาว่าจะใช้เกณฑ์ใดในการเริ่มต้นการทดสอบซึ่งแนวทางในการกำหนดเกณฑ์เริ่มต้นการทดสอบมีหลากหลายแนวทาง ดังนี้ (Becker et al., 2008; Thompson & Weiss, 2011; Kirisci et al., 2012)

- 1) ใช้ข้อสอบที่มีค่าความยากปานกลางจัดให้กับผู้สอบแต่ละคนเมื่อไม่ทราบค่าสารสนเทศเบื้องต้นของผู้สอบ
- 2) ทำการทดสอบเบื้องต้นก่อนการทดสอบจริงและนำผลมากำหนดจุดเริ่มต้นการทดสอบ
- 3) หากไม่มีข้อมูลอื่นประกอบการพิจารณา นิยมใช้ค่าเฉลี่ยคุณลักษณะแฝงหรือระดับความสามารถของประชากร (Average Theta) ซึ่งมีค่าเท่ากับ 0.00 เป็นจุดเริ่มต้นในการเลือกข้อสอบข้อแรกให้เหมาะสมกับระดับความสามารถของผู้สอบดังกล่าว
- 4) ใช้วิธีการสุ่มเลือกข้อสอบด้วยโปรแกรมคอมพิวเตอร์
- 5) หากมีผลการทดสอบของมาตรวัดนั้น ๆ ในครั้งก่อนหน้า อาจเริ่มต้นด้วยค่าคุณลักษณะแฝงหรือระดับความสามารถของผู้สอบที่ได้จากครั้งก่อน
- 6) พิจารณาจากข้อมูลอื่น ๆ เช่น ผลการทดสอบจากมาตรวัดอื่น เกรดเฉลี่ย แรงจูงใจในการทดสอบ และฐานะทางสังคม เป็นต้น

3. การคัดเลือกข้อสอบ (Item Selection)

สำหรับการเลือกข้อสอบจากคลังข้อสอบเพื่อให้ผู้สอบตอบ โดยหลักการคัดเลือกข้อสอบมีเป้าหมายเพื่อลดความแปรปรวนของเวกเตอร์ความสามารถแบบพหุมิติชั่วคราว (Frey & Seitz, 2009) วิธีการคัดเลือกข้อสอบที่ใช้กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติอาศัยภาวะน่าจะเป็นสูงสุดหรือต่ำสุดจากการประมาณค่า (θ) และมีวิธีการคัดเลือกข้อสอบอยู่หลากหลายวิธี แต่ละวิธีสามารถใช้ร่วมกันกับวิธีการประมาณค่าทุกแบบและยังสามารถใช้วิธีการคัดเลือกข้อสอบหลายวิธีร่วมกันได้อีกด้วย (Reckase, 2010, p. 327) โดยมีวิธีการคัดเลือกที่สำคัญ ดังนี้

3.1 Maximize the Determinant of the Fisher Information Matrix

การคัดเลือกข้อสอบด้วยเมทริกซ์สารสนเทศฟิชเชอร์ ดำเนินการภายใต้เงื่อนไขการประมาณค่าเวกเตอร์ความสามารถที่กำหนดจากเวกเตอร์ค่าเฉลี่ยเท่ากับ 0 และเวกเตอร์ความสามารถจริง (True θ) มีการแจกแจงแบบปกติหลายตัวแปรด้วยเมทริกซ์ความแปรปรวนและความแปรปรวนร่วม (Variance-Covariance Matrix) ที่เกี่ยวข้องกับเมทริกซ์สารสนเทศของฟิชเชอร์ (Fisher Information Matrix) (Segall, 2010; Frey & Seitz, 2009; Reckase, 2010, p. 330)

3.2 Largest Decrement in the Volume of the Bayesian Credibility Ellipsoid

วิธีการประมาณค่าแบบเบย์ ที่นำเสนอโดย Segall (2010, p. 65- 74) นำไปสู่การคัดเลือกข้อสอบที่จะลดปริมาณของข้อสอบได้อย่างมาก (Frey & Seitz, 2009; Reckase, 2010, p. 331-332)

3.3 Maximize Kullback-Leibler Information

วิธีการ Posterior Expected Kullback-Leibler Information ถูกนำเสนอ ให้ใช้ในการคัดเลือกข้อสอบในการทดสอบแบบปรับเหมาะที่เป็นมิติเดียว Chang and Ying (1996) และต่อมา Veldkamp and van der Linden (2002) ได้เสนอทางเลือกให้ใช้ วิธีการ Posterior Expected Kullback-Leibler Information ในการคัดเลือกข้อสอบแบบปรับเหมาะแบบพหุมิติ (MCAT) โดยมีเกณฑ์การคัดเลือกข้อสอบจากระยะห่างของการกระจายของการตอบสนอง จากข้อสอบข้อที่จะถูกเลือกให้ใช้ทดสอบข้อต่อไป เป็นตำแหน่งที่ใช้ประมาณค่าความสามารถของผู้สอบกับความสามารถที่แท้จริงของผู้สอบ (θ) ด้วยการคาดหมายเอาจากการตอบข้อสอบ (Wang, Chang, and Boughton, 2012, Mulder, van der Linden, 2009, p. 85 ; Reckase, 2010, p. 334-335) กรณีข้อสอบ 1 ข้อ

4. การประมาณค่าความสามารถ (Ability Estimation)

การประมาณค่าความสามารถของโมเดลการตอบสนองข้อสอบแบบพหุมิติ (MIRT) มีเกณฑ์การประมาณค่าที่แตกต่างกันหลายแนวทาง เช่น Maximum Likelihood Criterion, Maximum a Posteriori Bayesian Criterion, Newton-Raphson Method และ Least Squares Criterion (Reckase, 2010, p. 139) แต่โดยทั่วไปการประมาณค่าความสามารถของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) ที่นิยมมีอยู่ 2 วิธี คือ วิธีการประมาณค่าแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) และวิธีการประมาณค่าแบบเบส์ (Bayesian Estimation) (Reckase, 2010, p. 314)

ในบทความนี้ผู้เขียนได้เลือกอธิบายวิธีการประมาณค่าความสามารถตามวิธีของเบส์ (Bayesian Estimation) ในการประมาณค่าความสามารถของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) ซึ่งเป็นที่นิยมในการประมาณค่าความสามารถสำหรับการทดสอบแบบปรับเหมาะที่มีคลังข้อสอบขนาดใหญ่ ซึ่งมีรายละเอียดดังนี้ (Hambleton & Swaminathan, 1991, p.76 – 97 อ้างถึงใน รังสรรค์ มณีเล็ก, 2540)

การประมาณค่าความสามารถของผู้สอบ ถ้าผู้สอบตอบข้อสอบถูกทุกข้อหรือผิดหมดทุกข้อการประมาณค่าความสามารถด้วยวิธีความเป็นไปได้สูงสุดก็จะไม่เหมาะสม เว้นแต่จะนำเอาผลการตอบของผู้สอบเหล่านั้นออกไปจากกลุ่มผู้สอบก่อน หากทราบค่าสารสนเทศกายแรก (Prior Information) การประมาณค่าความสามารถด้วยวิธีของเบส์ก็จะเกิดประสิทธิภาพมากกว่าค่าสารสนเทศกายแรกที่มีกัณยิมนใช้กันก็คือความสามารถของผู้สอบที่มีการแจกแจงเป็นรูปโค้งปกติมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 (Hambleton & Swaminathan, 1991, p. 92)

การกำหนดรูปร่างของการแจกแจงเป็นรูปโค้งปกตินี้ว่าค่อนข้างสะดวก เบอรันบวม (Birnbbaum) เชื่อว่าการแจกแจงกายแรกของความสามารถเป็นฟังก์ชันโลจิสติก (logistic density function) ซึ่งมีสมการ ดังนี้

$$f(\theta) = \frac{\exp.(\theta)}{[1 + \exp.(\theta)]^2}$$

หัวใจของวิธีการประมาณค่าความสามารถตามวิธีการของเบส์ คือ ทฤษฎีบทของเบส์ (Bayes theorem) ซึ่งมีความสัมพันธ์กับ condition probability และ marginal probability

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

ในการประมาณค่าความสามารถของผู้สอบ A อาจถูกกำหนดให้เป็น θ_a และ B เป็นชุดของคำตอบของข้อสอบ n ข้อ ได้คะแนน u แล้วสมการ $P(B|A)$ เขียนใหม่ได้เป็น

$$P(\theta_a|u) = \frac{P(u|\theta_a)P(\theta_a)}{P(u)}$$

เนื่องจาก θ_a เป็นตัวแปรต่อเนื่อง ดังนั้น จึงถูกแปลความว่าเป็นฟังก์ชันความหนาแน่น (density function) เพื่อเป็นการป้องกันการจำสับสนระหว่าง $P(\theta_a)$ กับ $P(\theta)$ ซึ่งเป็นฟังก์ชันการตอบสนองข้อสอบ (item response function) จึงใช้ $f(\theta_a)$ แทน $P(\theta)$ ดังนั้น จึงเขียนสมการใหม่ได้ ดังนี้

$$f(\theta_a|u) = \frac{f(u|\theta_a)f(\theta_a)}{f(u)}$$

เมื่อ $f(u)$ แทน ค่าคงที่

$f(\theta_a | u)$ แทน ความหนาแน่นภายหลัง (posterior density) ของ θ_a

$f(\theta_a)$ แทน ความหนาแน่นก่อน (prior density) ของ θ_a

แต่เนื่องจาก $f(u | \theta_a)$ เป็นฟังก์ชันความเป็นไปได้ (likelihood function) ของคะแนน

ดังนั้น จึงสามารถเขียนสมการ $f(u | \theta_a)$ ให้อยู่ในรูปของการแปรผันได้ ดังนี้

$$f(\theta_a | u) \propto L(u | \theta_a) f(\theta_a) \text{ หรือ Posterior Likelihood} \times \text{Prior}$$

วิธีการประมาณค่าความสามารถของผู้สอบตามวิธีการของเบส์นี้ ได้ถูกนำไปประยุกต์เป็นวิธีต่างๆ อีกหลายวิธี แต่วิธีที่ได้รับความนิยมมากที่สุด คือ วิธีของเบส์ที่ปรับปรุงใหม่ (Bayesian Updating) ซึ่งเสนอโดยโอเวน (Owen, 1975, p. 353 อ้างถึงใน รังสรรค์ มณีเล็ก, 2540; เกียรติศักดิ์ ส่องแสง, 2547, หน้า 47-48, ชัยมงคล ปินะสา, 2551, หน้า 60-61) ซึ่งมีวิธีการประมาณค่า ดังนี้

ในกรณีที่ตอบข้อสอบถูก

$$\theta_{m+1} = \theta_m + (1-c) \left(\frac{\sigma_m^2}{\sqrt{\frac{1}{a^2} + \sigma_m^2}} \right) \left(\frac{O(D)}{B} \right)$$

$$\sigma_{m+1}^2 = \sigma_m^2 + \left[1 - \left(\frac{1-c}{1 + \frac{1}{a^2 \sigma_m^2}} \right) \left(\frac{O(D)}{B} \right) \left(\frac{(1-c)O(D)}{B} \right) - D \right]$$

เมื่อ $D = \frac{(b-\theta)}{\sqrt{\frac{1}{a^2} + \sigma^2}}$ และ $B = c + (1-c)A(-D)$

เมื่อ θ_m แทน ค่าความสามารถของผู้สอบที่ประมาณได้ก่อนตอบข้อสอบข้อที่ $m+1$ ตามปกติแล้วถ้าไม่ทราบความสามารถเบื้องต้นของผู้สอบจะกำหนดให้ $\theta_m = 0.000$

θ_{m+1} แทน ความสามารถของผู้สอบที่ประมาณค่าได้หลังจากตอบข้อสอบข้อที่ $m+1$ แล้ว

σ_m^2 แทน ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบก่อนตอบข้อสอบข้อที่ $m+1$ ตามปกติแล้วถ้าไม่ทราบความแปรปรวนดังกล่าวมาก่อน จะกำหนดให้ $\sigma_m^2 = 1.000$

σ_{m+1}^2 แทน ค่าความแปรปรวนในการประมาณค่าความสามารถของผู้สอบเมื่อตอบข้อสอบข้อที่ $m+1$ แล้ว

a_i	แทน	ค่าอำนาจจำแนกของข้อสอบข้อที่ $m+1$
b_i	แทน	ค่าความยากของข้อสอบข้อที่ $m+1$
c_i	แทน	ค่าการเดาของข้อสอบข้อที่ $m+1$
D	แทน	จุดบนแกน x
$O(D)$	แทน	ค่าออร์ดิเนต (Ordinate) ของโค้งปกติที่จุด D
$A(D)$	แทน	พื้นที่ใต้โค้งปกติจากค่า D ต่ำสุดถึงจุด D

ในกรณีที่ตอบผิด

$$\theta_{m+1} = \theta_m - \left(\frac{\sigma_m^2}{\sqrt{\frac{1}{a^2} + \sigma_m^2}} \right) \left(\frac{O(D)}{A(D)} \right)$$

$$\sigma_{m+1}^2 = \sigma_m^2 + \left[1 - \frac{\left(\frac{O(D)}{1 + \frac{1}{a^2 \sigma_m^2}} \right) \left(\frac{O(D)}{A(D)} + D \right)}{A(D)} \right]$$

ตัวอย่างการประมาณค่าความสามารถและความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ เมื่อกำหนดให้ $\theta_m = 0.0000$ และ $\sigma_m^2 = 1.0000$ ข้อสอบทั้งหมดมีค่าอำนาจจำแนกเท่ากับ 1.00 ค่าการเดาเท่ากับ .20 (ในแต่ละขั้นของการทดสอบ การคัดเลือกข้อความยากของข้อสอบจะถูกตัดโดยแท้จริงกับระดับความสามารถของผู้สอบที่ประมาณค่าแรกเริ่มนั่นคือ $b_i - \theta_m = 0$) จากตัวอย่างการตอบของข้อสอบจำนวน 6 ข้อ มีผลการตอบเป็น ถูก, ผิด, ถูก, ถูก, ผิด, ผิด, สามารถนำมาประมาณค่าความสามารถและความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถได้จากการแทนค่าในสูตร ดังนี้

$$\theta_1 = 0 + (1 - 0.20) \left(\frac{1}{\sqrt{1+1}} \right) \left(\frac{0.3989}{0.20(0.80)(.50)} \right)$$

$$= 0.376$$

$$\sigma_1^2 = 1.0 \left\{ 1.0 - \left(\frac{1.0 - 0.20}{1.0 + \frac{1.0}{1.0}} \right) \left(\frac{0.3989}{0.20 + (0.80)(0.50)} \right) \left(\frac{(0.80)(0.3989)}{0.20 + (0.8)(0.5)} \right) \right\}$$

$$= 0.859$$

จากผลการคำนวณที่ผู้สอบตอบถูกในข้อแรก สามารถประมาณค่าความสามารถ ของผู้สอบได้เท่ากับ 0.376 ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถเท่ากับ $\sqrt{0.859}$ นั่นคือเท่ากับ 0.927 และเมื่อผู้สอบตอบข้อสอบข้อที่ 2 ผิด สามารถนำมาประมาณค่าความสามารถและความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ ได้ดังนี้

$$\theta_1 = 0.376 - \left(\frac{0.859}{\sqrt{1.0 + 0.859}} \right) \left(\frac{0.3989}{0.50} \right) \\ = -0.127$$

$$\sigma_1^2 = 0.859 \left\{ 1.0 - \left(\frac{0.3989}{1.0 + \frac{1.0}{0.859}} \right) \left(\frac{0.3989}{0.25} \right) \right\} \\ = 0.606$$

จากผลการประมาณค่าความสามารถและความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบ จำนวน 6 ข้อ แสดงผลการตอบและการประมาณค่าได้ ดังตารางที่ 1

ตารางที่ 1 ผลการตอบและการประมาณค่าความสามารถ

ข้อสอบ	ผลการตอบ	ค่าประมาณ ความสามารถ (θ)	ค่าความแปรปรวน (σ^2)	ค่าความคลาดเคลื่อน มาตรฐาน (σ)
1	ถูก	0.375	0.859	0.927
2	ผิด	-0.127	0.606	0.778
3	ถูก	0.127	0.541	0.736
4	ถูก	0.359	0.487	0.698
5	ผิด	0.040	0.385	0.620
6	ผิด	-0.221	0.317	0.563

ที่มา : Thondike, R.L (1982, p. 305)

จากตารางที่ 1 แสดงให้เห็นว่ากระบวนการในการคัดเลือกข้อสอบข้อใหม่ให้มีค่าความยากใกล้เคียงที่สุดกับความสามารถที่ประมาณค่าได้ใหม่ มีผลทำให้ค่าความสามารถที่ประมาณค่าได้ใหม่มีค่าเปลี่ยนแปลงไป กล่าวคือ ถ้าตอบข้อสอบถูกค่าความสามารถที่ประมาณได้ใหม่จะมีค่าสูงขึ้น และถ้าตอบผิดค่าความสามารถที่ประมาณได้ใหม่จะมีค่าลดลง ในขณะที่ความคลาดเคลื่อน มาตรฐานของการประมาณค่าจะมีค่าลดลงเรื่อย ๆ และค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถลดลงถึงจุดที่กำหนดไว้ คือ .30 กระบวนการทดสอบจะยุติ

การประมาณค่าความสามารถของผู้สอบด้วยวิธีเบย์ส์ที่ปรับเปลี่ยน (Bayesian Updating) ซึ่งเสนอโดยโอเวน (Owen) เป็นวิธีการที่ได้รับความนิยมมากจนถึงปัจจุบันนี้ วิธีการนี้ใช้ความรู้เกี่ยวกับโครงสร้างของความแปรปรวนร่วมระหว่างมิติของการวัดสำหรับการเลือกข้อสอบและการประมาณค่าความสามารถ ส่งผลให้จำนวนข้อสอบที่ใช้ลดลง ซึ่งเป็นคุณสมบัติที่จำเป็นของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ และการประมาณค่าความสามารถของผู้สอบที่ให้ค่าเที่ยงตรงเชิงสภาพสูง (ริงส์เวิร์ค มณีเล็ก, 2540) ดังนั้น ผู้เขียนบทความจึงได้นำเสนอแต่วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบย์ส์ที่ปรับเปลี่ยน ดังกล่าว

5. การยุติการทดสอบ (Stopping Rule)

กฎการยุติการทดสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์โดยทั่วไปใช้กฎการยุติการทดสอบมี 2 ลักษณะได้แก่ กฎความยาวคงที่ (Fixed-Length) และ จำนวนข้อสอบผันแปร (variable length) (Reckase, 2010, p. 335) และการพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard error of estimation: SE)

(Thissen, 1990) แต่การกำหนดเกณฑ์การยุติการทดสอบขึ้นอยู่กับวัตถุประสงค์ในการทดสอบ ลักษณะคลังข้อสอบ และการควบคุมการปฏิบัติงานต่างๆเกณฑ์ในการพิจารณาการยุติการทดสอบมีหลากหลายเกณฑ์ให้เลือกใช้ (Wainer et al., 2001, p. 113; Thompson & Weiss, 2011; Kuo et al., 2015) ดังนี้

5.1 ยุติการทดสอบเมื่อข้อสอบในคลังข้อสอบหมด

5.2 ยุติการทดสอบเมื่อการประมาณค่าคุณลักษณะแฝงมีการเปลี่ยนแปลงเล็กน้อยหรือไม่เปลี่ยนแปลงเลย

5.3 กฎความยาวคงที่ (Fixed-Length) และ จำนวนข้อสอบผันแปร (variable length) (Reckase, 2010, p. 335) สำหรับงานวิจัยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) ที่ผ่านมานั้น ส่วนใหญ่ จะใช้การจำกัดจำนวนข้อสอบที่ 10, 20, 30, 40, 50 และ 60 ข้อ (Chen, 2009; Diao & Reckase, 2010; Finch, 2010) ซึ่งงานวิจัยของ Diao and Reckase (2009) ปรากฏว่า การใช้ข้อสอบจำนวน 50 ข้อ ความลำเอียงจะน้อยกว่า การใช้ข้อสอบ จำนวน 20 ข้อ แต่ยังไม่ได้ว่าจำนวนข้อสอบเท่าใดจึงจะเหมาะสมสำหรับการทดสอบปรับเหมาะแบบ พหุมิติ

5.4 การพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard Error of estimation: SE) โดยทั่วไปจะใช้กฎการยุติการทดสอบจากค่าความคลาดเคลื่อนมาตรฐานกำหนดไว้ที่ 0.3 (Thissen, 1990) ส่วนงานวิจัยที่ผ่านมาจะกำหนด $SE \leq .55$, .50 และ 0.30 (Seo, 2011; Thissen, 1990; รังสรรค์ มณีเล็ก, 2540; สมประสงค์ เสนารัตน์, 2555; โสฬส สุขานนท์สวัสดิ์ เสรี ชัดแจ้ง และกฤษณะ ชินสาร, 2556; สุกัญญา บุญศรี, 2560)

ดังนั้น จุดสิ้นสุดการทดสอบ มีเกณฑ์ในการยุติการทดสอบที่นำเสนอหลากหลายเกณฑ์ เช่น หยุดเมื่อ จำนวนข้อสอบที่ทดสอบครบตามจำนวน หรือเมื่อการวัดมีความแม่นยำตามเกณฑ์ หรือใช้ทั้งสองเกณฑ์ร่วมกัน ในการศึกษาจากข้อมูลจำลองส่วนใหญ่ การทดสอบแบบปรับเหมาะแบบพหุมิติจะยุติการทดสอบภายหลังการสอบครบ ตามจำนวนข้อสอบที่กำหนดไว้ล่วงหน้า อย่างไรก็ตาม การกำหนดจำนวนข้อสอบที่ใช้ในการทดสอบเป็นเกณฑ์การยุติ การสอบจะส่งผลต่อความแม่นยำในการประมาณค่าความสามารถที่มีความแปรปรวนระหว่างผู้สอบสูงมาก ผู้สอบบางคน มีความคลาดเคลื่อนมาตรฐานของการวัดจากหนึ่งมิติหรือมากกว่าจะมีขนาดใหญ่ ในขณะที่คนอื่นจะมีขนาดเล็กกว่ามาก และความแม่นยำในการประมาณค่าความสามารถยังขึ้นอยู่กับคำตอบข้อสอบและคุณลักษณะของคลังข้อสอบ (Frey & Seitz, 2009: p. 91)

สรุป

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing: MCAT) มีเป้าหมายเพื่อที่จะดำเนินการสอบด้วยการคัดเลือกข้อสอบให้เหมาะสม (Tailoring) กับความสามารถ ของผู้สอบแต่ละคน โดยอาศัยทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model : MIRT) มาใช้เป็นพื้นฐานในการคำนวณค่าสารสนเทศของข้อสอบ ซึ่งจะมีการคัดเลือกข้อสอบที่สามารถให้ สารสนเทศอันเป็นประโยชน์เกี่ยวกับความสามารถของผู้สอบแต่ละคน ทำให้การทดสอบเป็นไปอย่างมีประสิทธิภาพ ใช้จำนวนข้อสอบที่น้อยลง ประหยัดเวลาของการสอบ ผลที่ได้มีความคลาดเคลื่อนต่ำ มีความแม่นยำของการวัด ความสามารถที่แท้จริงของผู้สอบได้อย่างมีประสิทธิภาพ และเหมาะที่จะนำไปใช้กับทดสอบกับผู้เรียนในยุคปัจจุบันนี้

เอกสารอ้างอิง

เกียรติศักดิ์ ส่องแสง. (2547). *ฟังก์ชันสารสนเทศของแบบทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วย คอมพิวเตอร์*. วิทยานิพนธ์ กศ.ด. กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.

- ชัยมงคล ปิ่นสา. (2551). การสร้างแบบทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบบนระบบเครือข่ายอินเทอร์เน็ตในวิชาวิทยาศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์ปริญญาโทบริหารการศึกษา. มหาสารคาม: มหาวิทยาลัยมหาสารคาม.
- ชัยวิชิต เขียวชนะ. (2552). "การวิเคราะห์พหุมิติ." วารสารศึกษาศาสตร์มหาวิทยาลัยขอนแก่น, 32(4), 13-21.
- เมษา นวลศรี. (2559). การพัฒนาและตรวจสอบโครงสร้างพหุมิติของความเป็นพลเมืองที่มีความรับผิดชอบของนักเรียนชั้นมัธยมศึกษาตอนต้น: การประยุกต์ใช้แนวคิดการสร้างแผนที่โครงสร้าง วิทยานิพนธ์ปริญญาโทบริหารการศึกษา. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- รังสรรค์ มณีเล็ก. (2540). ผลของตัวแปรบางตัวต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์. วิทยานิพนธ์ปริญญาโทบริหารการศึกษา. กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.
- โสฬส สุขานนท์สวัสดิ์, เสรี ชัดแจ้ง และกฤษณะ ชินสาร. (2556). การพัฒนาการคัดเลือกข้อสอบข้อถัดไปโดยใช้ทฤษฎีการตัดสินใจ ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์. วิทยาการวิจัยและวิทยาการปัญญา, 10(2), 71-85.
- สมประสงค์ เสนารัตน์. (2555). การพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อวินิจฉัยกระบวนการพุทธิปัญญาในการเรียนพีชคณิตของนักเรียนชั้นมัธยมศึกษาปีที่ 1 โดยประยุกต์ใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติ. วิทยานิพนธ์ปริญญาโทบริหารการศึกษา. มหาสารคาม: มหาวิทยาลัยมหาสารคาม.
- สุกัญญา บุญศรี (2560) การพัฒนาการทดสอบแบบปรับเหมาะแบบหลายขั้นตอนด้วยคอมพิวเตอร์เพื่อวัด สมรรถนะความรู้วิชาชีพพหุมิติของครูช่างอุตสาหกรรม. วิทยานิพนธ์ปริญญาโทบริหารการศึกษา. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2555). ทฤษฎีการทดสอบแนวใหม่ (ฉบับปรับปรุง) (พิมพ์ครั้งที่ 4). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- 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(12), E182-E192
- Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, 12(3), 261-280.
- Babcock, B., & Weiss, D. J. (2012). Termination criteria in computerized adaptive tests: Do variable-length CATs provide efficient and effective measurement?. *Journal of Computerized Adaptive Testing*, 1(1), 1-18.
- Chang, H. H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20(3), 213-229.
- Chang, H. H. (2012). Making computerized adaptive testing diagnostic tools for schools. In R. W. Lissitz & H. Jiao (Eds.), Computers and their impact on state assessments: Recent history and predictions for the future (pp. 195–226). Charlotte, North Carolina: Information Age.
- Chen J.Z. (2009). *Material flow and circular economy*. Wiley Online Library. Retrieved from <https://onlinelibrary.wiley.com/doi/epdf/10.1002/sres.968>.

- Diao, Q., & Reckase, M. (2009). *Comparison of ability estimation and item selection methods in multidimensional computerized adaptive testing*. Michigan State University. Department of Measurement and Quantitative Methods.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publications.
- Frey, A., & Seitz, N. N. (2009). Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation*, 35(2), 89-94.
- Finch, H. (2010). Item Parameter Estimation for the MIRT Model Bias and Precision of Confirmatory Factor Analysis—Based Models. *Applied Psychological Measurement*, 34(1), 10-26.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publications, Inc.
- Kuo, B. C., Daud, M., & Yang, C. W. (2015). Multidimensional Computerized Adaptive Testing for Indonesia Junior High School Biology. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(5), 1105-1118.
- 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 & Alcohol Dependence*, 123(1), S79-S86.
- Mulder, J., & van der Linden, W. J. (2009). *Multidimensional adaptive testing with Kullback–Leibler information item selection*. In W. J. van der Linden, & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 77–101). New York: Springer.
- Reckase, M. D. (2010). *Multidimensional item response theory*. New York: Springer.
- Seo, D. G. (2011). Application of the Bifactor Model to Computerized Adaptive Testing. *ProQuest*.
- Segall, D. O. (2010). *Principles of multidimensional adaptive testing,* in Wim J. Van Der Linden and Cees AW Glas (Eds.). *Elements of Adaptive Testing*. New York: Dordrecht Heidelberg London.
- Thompson, N.A., & Weiss, D.J. (2011). A Framework for the Development of computerized Adaptive Tests. *Practical Assessment, Research & Evaluation*, 16(1), 1-9.
- Thronrdike, R.L. (1982). *Applied Psychometrics*. Boston: Houghton Mifflin Company.
- Thissen, David. (1990). Reliability and Measurement Precision. In: Howard Wainer et al., editors. *Computerized Adaptive Testing: A Primer*. New Jersey: Lawrence Erlbaum.
- Veldkamp, B. P., & van der Linden, W. J. (2002). Multidimensional adaptive testing with constraints on test content. *Psychometrika*, 67(4), 575-588.
- Veldkamp, B. P., & Matteucci, M. (2013). Bayesian computerized adaptive testing. *Ensaio: Avaliação e Políticas Públicas em Educação*, 21(78), 57-82.
- Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B. F., Mislevy, R. J., & Steinberg, L. (2001). Computerized adaptive testing: A primer. *Qual Life Res*, 10(8), 733-734.