

## การประยุกต์ใช้การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ Application of Multidimensional Computerized Adaptive Testing

ศักดิ์ชัย จันทะแสง<sup>1\*</sup>  
Sakchai Jantasang<sup>1\*</sup>

### บทคัดย่อ

บทความนี้มีวัตถุประสงค์เพื่อให้เกิดความเข้าใจเกี่ยวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ที่เป็นรูปแบบการทดสอบที่เหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งเป็นวิธีการที่ผสมผสานระหว่างทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ และการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติพัฒนาขึ้นมาเพื่อวัดความสามารถของผู้สอบมากกว่า 1 มิติ ผ่านการทดสอบเพียงครั้งเดียว มีประสิทธิภาพสูง มีความแม่นยำ และลดจำนวนข้อสอบ ซึ่งการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติมีองค์ประกอบที่สำคัญ 5 องค์ประกอบ ได้แก่ 1) การสร้างคลังข้อสอบ 2) การเริ่มต้นการทดสอบ 3) การคัดเลือกข้อสอบ 4) การประมาณค่าความสามารถของผู้สอบ และ 5) เกณฑ์ยุติการทดสอบ

**คำสำคัญ:** ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

### Abstract

The purposes of this article was to understand the multidimensional computerized adaptive testing. The multidimensional computerized adaptive testing (MCAT), which is a combination of the multidimensional item response theory (MIRT) and computerized adaptive testing (CAT), is a form of testing that adapts to examinee's proficiency level. MCAT was employed for measurement of several latent traits by a single examination which was highly efficient, precise and reduced the number of exam. MCAT consists of 5 components including: 1) development of item bank 2) start of the first item 3) selection of the next item 4) estimation of examinees' ability and 5) termination criterion.

**Keywords:** multidimensional item response theory, multidimensional computerized adaptive testing

### บทนำ

การทดสอบแบบปรับเหมาะ (Adaptive Testing) ในระยะแรกเป็นการทดสอบที่ใช้กระดาษกับดินสอ (Paper and Pencil Test) ในระหว่างกระบวนการสอบนั้น ต้องมีการปรับเหมาะโดยการคัดเลือกข้อสอบข้อถัดไป เพื่อให้เหมาะสมกับความสามารถของผู้สอบ ซึ่งการคัดเลือกข้อสอบข้อถัดไปถือว่าเป็นเรื่องที่ยุ่งยาก ต่อมาจึงได้มีการนำคอมพิวเตอร์มาประยุกต์ใช้กับการทดสอบแบบปรับเหมาะ โดยนำทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) กับเทคโนโลยีซิลิคอนชิป (Silicon Chip Technology) เข้าด้วยกันเพื่อประโยชน์ของการทดสอบ เรียกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) โดยใช้คอมพิวเตอร์ในการประมาณค่าความสามารถของผู้สอบ และคัดเลือกข้อสอบที่เหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งสามารถลดจำนวนข้อสอบลงครึ่งหนึ่งโดยไม่สูญเสียความแม่นยำในการวัด ดังนั้น การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จึงช่วยลดภาระงาน เวลา และค่าใช้จ่ายลงอย่างมาก (Frey and Seitz, 2009)

<sup>1</sup> สำนักงานประกันคุณภาพการศึกษา มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ กรุงเทพมหานคร 10120

<sup>1</sup> Office of Educational Quality Assurance, Rajamangala University of Technology Krungthep, Bangkok, 10120

\*Corresponding author: e-mail: amchai.j@gmail.com

Received: July 31, 2020, Accepted: September 19, 2020, Published: October 22, 2020



ระยะแรกของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ โดยใช้โมเดลการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional Item Response Theory Model: UIRT) ซึ่งเชื่อว่าข้อสอบที่สร้างขึ้นนั้นวัดมิติเดียว ซึ่งการวัดผลทางการศึกษาและจิตวิทยาในปัจจุบันเชื่อว่าข้อสอบที่สร้างขึ้นมามีความสามารถที่ซับซ้อนมากกว่าหนึ่งมิติ การวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติจะทำให้มีการฝ่าฝืนข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ ที่กำหนดว่าแบบทดสอบที่ใช้วัดความสามารถจะต้องวัดความสามารถด้านเดียวหรือเป็นเอกพันธ์ (Homogeneous) หรือมีลักษณะเป็นเอกมิติ (Unidimensional) การศึกษาจึงได้ขยายแนวคิดของทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ สู่โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory: MIRT) โดยถือว่าคุณลักษณะแฝงหรือความสามารถของบุคคลมากกว่า 1 องค์ประกอบ ส่งผลต่อการตอบข้อสอบ พารามิเตอร์ความสามารถของผู้สอบจึงมี 2 พารามิเตอร์ขึ้นไป จากแนวคิดดังกล่าวจึงได้ขยายรูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ (Unidimensional Computerized Adaptive Testing: UCAT) สู่รูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing: MCAT) ซึ่งเป็นรูปแบบใหม่สำหรับการประมาณความสามารถที่แท้จริงของผู้สอบ (Frey and Carstensen, 2009) เนื่องจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติมีอัตราการวัดที่แม่นยำกว่าเมื่อใช้ข้อสอบเท่ากัน สามารถลดจำนวนข้อสอบลงได้มากกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ 30-50% และมีประสิทธิภาพสูงกว่าถึง 1.3 เท่าและสามารถลดจำนวนข้อสอบลงได้มากกว่าการทดสอบแบบดั้งเดิมที่กำหนดจำนวนข้อสอบไว้ 70% โดยไม่สูญเสียความแม่นยำการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ จึงมีประสิทธิภาพการทดสอบสูงกว่าการทดสอบแบบปรับเหมาะคอมพิวเตอร์แบบเอกมิติ (Frey and Seitz, 2009)

จากจุดเด่นของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ที่สามารถคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบมีประสิทธิภาพสูง มีความแม่นยำในการประมาณค่าความสามารถของผู้สอบโดยใช้ข้อสอบและเวลาทดสอบที่ลดลงครึ่งหนึ่งของการทดสอบแบบดั้งเดิม จึงเป็นเรื่องที่น่าศึกษาและนำมาประยุกต์ใช้ในการดำเนินการสอบ

## ความรู้เบื้องต้นเกี่ยวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

### 1. โมเดลการตอบสนองข้อสอบแบบพหุมิติ

โดยทั่วไปในการระบุความน่าจะเป็นของโมเดลการตอบสนองข้อสอบแบบพหุมิตินั้น การตอบสนองที่ถูกต้องของข้อสอบที่  $i$  ( $U_i = 1$ ) ขึ้นอยู่กับความสามารถแฝงจำนวน  $p$  ซึ่งแทนด้วย  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  และคุณลักษณะของข้อสอบที่ครอบคลุมด้วยพารามิเตอร์ข้อสอบหนึ่งพารามิเตอร์หรือมากกว่า (Frey and Seitz, 2009) โดย Segall (1996, 2010) สำหรับการนำเสนอโมเดลการระบุความน่าจะเป็นในการตอบที่ถูกต้องของโมเดลโลจิสติกแบบพหุมิติ  $P_i(\theta)$  คำนวณได้จากสมการที่ 1

$$P_i(\theta) \equiv P(U_i = 1|\theta) = c_i + \frac{1 - c_i}{1 + \exp[-Da'_i(\theta - b_i1)]} \dots\dots\dots(1)$$

เมื่อ  $P_i(\theta)$  แทน ฟังก์ชันการตอบสนองข้อสอบข้อที่  $i$

$U_i$  แทนตัวแปรสุ่มการตอบสนองข้อสอบข้อที่  $i$  ( $U_i = 1, 0$  เมื่อตอบข้อที่  $i$  ถูกต้องและไม่ถูกต้องตามลำดับ)

$c_i$  แทน พารามิเตอร์โอกาสการเดาของข้อสอบข้อที่  $i$  ได้ถูกต้อง

$b_i$  แทน พารามิเตอร์ความยากของข้อสอบข้อที่  $i$

$a'_i$  แทน เวกเตอร์  $1 \times p$  ของพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่  $i$

1 แทน เวกเตอร์  $p \times 1$  ของ 1's และ D แทน ค่าคงที่ เท่ากับ 1.702

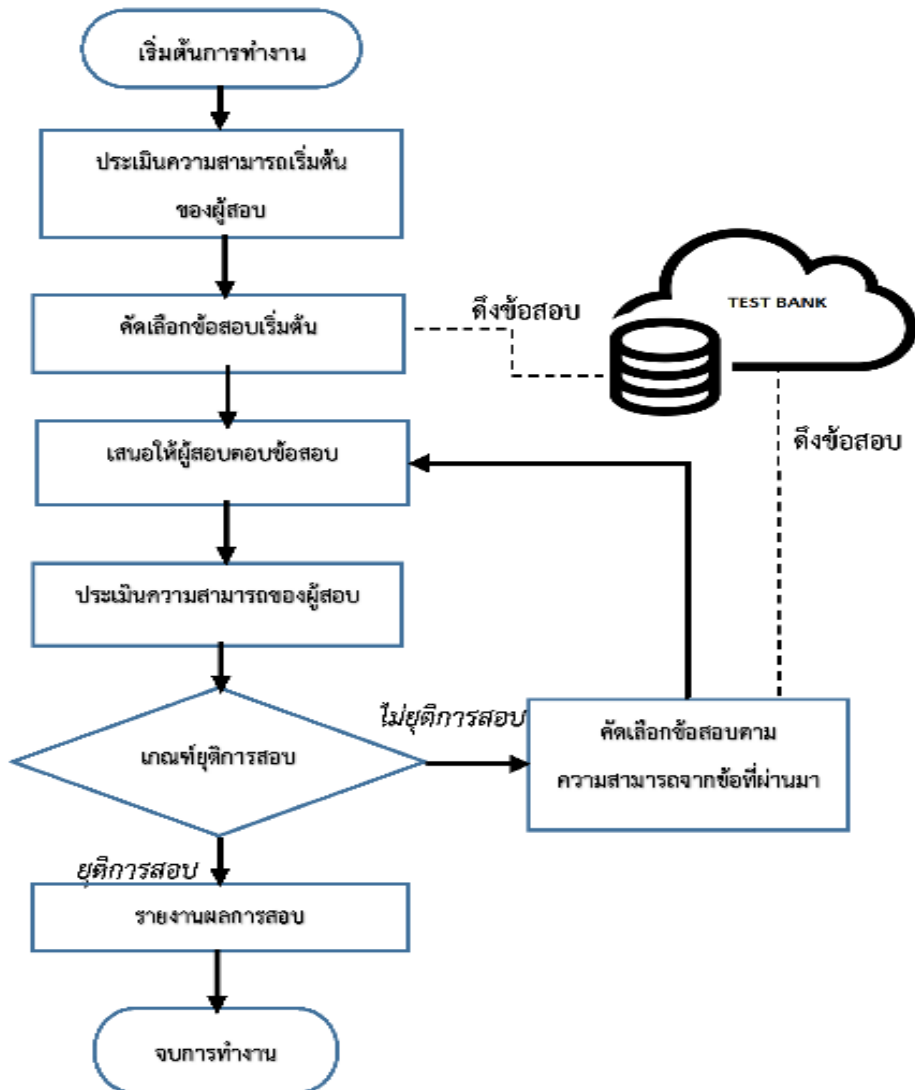
โดยที่เลขชี้กำลังของตัวหารในสมการที่ (1) สามารถอธิบายในรูปของสเกลดังนี้

$$-Da'_i(\theta - b_i1) = -D \sum_{k=1}^p a_{ki} (\theta_k - b_i) \dots\dots\dots(2)$$



## 2. องค์ประกอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ประกอบด้วยขั้นตอน 5 ขั้นตอน ดังภาพที่ 1 (Veldkamp and Matteucci, 2013; Thompson and Weiss, 2011)



ภาพที่ 1 กระบวนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

กระบวนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีหลักการคัดเลือกข้อสอบที่เหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งตั้งอยู่บนฐานการตอบข้อสอบในข้อที่ผ่านมาของผู้สอบ กล่าวคือ เมื่อผู้สอบทำข้อสอบข้อเริ่มต้นหรือชุดเริ่มต้น (ขึ้นอยู่กับการออกแบบ) และจะนำผลการตอบข้อสอบนั้นมาวิเคราะห์หรือประมาณค่าความสามารถของผู้สอบ เพื่อที่จะเลือกข้อสอบที่เหมาะสมกับข้อสอบในข้อถัดไป โดยอาศัยหลักการคัดเลือกข้อสอบว่า ถ้าทำข้อสอบข้อที่ผ่านมาถูกต้อง ข้อสอบข้อถัดไปจะยากมากขึ้นหากทำข้อสอบข้อที่ผ่านมาผิด ข้อสอบข้อถัดไปจะง่ายลง กระบวนการนี้จะดำเนินการไปจนกว่าการประมาณค่าความสามารถของผู้สอบเป็นไปตามเกณฑ์การยุติการทดสอบที่กำหนด (ขึ้นอยู่กับการออกแบบ) การทดสอบจึงยุติลง ในปัจจุบันมีการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติมากขึ้น โดยผู้เขียนได้สรุปรายละเอียดของตัวอย่างงานวิจัย พร้อมจำแนกประเด็นสำคัญต่าง ๆ ที่เกี่ยวข้อง ดังตารางที่ 1

**ตารางที่ 1** สรุปตัวอย่างงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ผู้วิจัย	Item Bank	ประเด็นที่ศึกษา		
		Item Selection	Ability Estimate	Termination Criteria
Kuo <i>et al.</i> (2015)	300 ข้อ	ML	MLE, MAP	SEE $\leq$ 0.30
Mao <i>et al.</i> (2016)	450 ข้อ	D-Optimal, Posterior Expectation, KL*	MPI	SEE $\leq$ 0.30 and fixed number of items (k=30)
Sahin and Gelbal (2020)	200 ข้อ	D-rule, KL*, W-rule, weighted W-tule, T-rule, และ weighted T-rule	EAP, MAP	SEE=0.40, $\theta$ convergence ( $\Delta\theta < 0.05$ ) and fixed number of items (k=20)
นฤมล (2558)	269 ข้อ	Maximize the Determinant of the Fisher Information Matrix	Bayesian*	การจำกัดจำนวนข้อ (Fixed Length) และใช้เงื่อนไขการทวนคำตอบที่มีการจำกัดช่วง (Block Review)
สุชาติ และคณะ (2559)	100, 150, 200 ข้อ	Maximize Kullback-Leibler Information*	Bayesian*	จำนวนข้อสอบ 20 ข้อ, SEE $\leq$ 0.30 และ SEE $\leq$ 0.43
พัชราภรณ์ และ ไพรัตน์ (2561)	163 ข้อ	Bayesian Volume Decrease	Bayesian*	จำนวนข้อสอบคงที่ 50 ข้อ
ธีรวัช และคณะ (2562)	162 ข้อ	MI	Bayesian*	SEE $\leq$ 0.30 หรือ หมดคลังข้อสอบ
ศักดิ์ชัย (2562)	193 ข้อ	Maximize Kullback-Leibler Information*	Bayesian*	SEE $\leq$ 0.30 หรือ 16 ข้อ

ML คือ Maximum Likelihood, KL คือ Kullback-Leibler Information, EAP คือ Expected a Posteriori, MAP คือ Maximum a Posteriori, MLE คือ Maximum Likelihood Estimation, MPI คือ Maximum Priority Index, SEE คือ Standard Error of Estimation,

\* คือ วิธีการที่นิยมใช้

## 2.1 การสร้างคลังข้อสอบ (Development of Item bank)

คลังข้อสอบ (Item Pools, Item Bank) เป็นแหล่งจัดเก็บข้อมูลต่าง ๆ ที่เกี่ยวข้องกับข้อสอบ (โจทย์และตัวเลือก) รวมทั้งค่าพารามิเตอร์ต่าง ๆ ของข้อสอบที่ผ่านการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT หรือ MIRT) คลังข้อสอบควรมีข้อสอบมากเพียงพอและครอบคลุมเนื้อหาทั้งหมด ซึ่งในการพัฒนาคลังข้อสอบจะเริ่มต้นจากการนำข้อสอบที่สร้างขึ้น ไปพัฒนาผ่านระบบการสอบออนไลน์ที่ทันสมัย เพื่อให้การเก็บข้อมูลผู้สอบเป็นไปอย่างรวดเร็วและลดข้อผิดพลาด อันเนื่องมาจากการป้อนข้อมูลในรูปแบบเดิมที่ได้จากรูปแบบการทดสอบที่ใช้กระดาษกับดินสอ (Paper and Pencil Test) และผลการทดสอบที่เก็บรวบรวมได้นี้ พร้อมทั้งจะนำไปใช้งานต่อด้านต่าง ๆ โดยเฉพาะการวิเคราะห์คุณภาพข้อสอบสำหรับกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์คุณภาพข้อสอบนั้นจะต้องมีจำนวนมากพอและขึ้นอยู่กับโมเดลการตอบสนองข้อสอบ (IRT หรือ MIRT) ที่ใช้วิเคราะห์ด้วย (Thompson and Weiss, 2011; Wainer *et al.*, 2001) การวิเคราะห์จะช่วยให้สามารถตัดสินใจเลือกข้อสอบที่มีความเหมาะสมเข้าสู่คลังข้อสอบได้ โดยวิธีการตรวจสอบคุณภาพของข้อสอบสามารถใช้วิธีการทางสถิติของทฤษฎีการทดสอบแบบดั้งเดิม เช่น สัดส่วนของผู้ที่ตอบถูก (Proportion Correct) หรือ สหสัมพันธ์แบบไบเซเรียล (Biserial Correlation) ร่วมกับทฤษฎีการทดสอบแนวใหม่ เช่น พารามิเตอร์ของข้อสอบ (Item Parameters) และค่าสารสนเทศของข้อสอบ (Item Information) (Wainer, 1989) และสิ่งที่สำคัญอีกประการหนึ่ง คือ คุณลักษณะของข้อสอบ (Item Characteristic) เพื่อตรวจสอบว่าข้อสอบแต่ละข้อเหมาะสมกับโมเดลการวัด ดังนั้น คลังข้อสอบเป็นปัจจัยสำคัญที่ทำให้การทดสอบดำเนินไปได้อย่างมีประสิทธิภาพ การเลือกข้อสอบที่มีคุณภาพเข้าสู่คลังข้อสอบอาจมีรูปแบบข้อสอบที่แตกต่างกันอยู่ในคลังเดียวกันได้ เช่น ข้อสอบแบบหลายตัวเลือก (Multiple Choice Questions) โดยปกติแล้วหากคลังข้อสอบมีการตรวจให้คะแนนแบบ 2 ค่า (เช่นตอบถูกให้ 1 หรือ ตอบผิดให้ 0) ควรมีข้อสอบอยู่ในคลังข้อสอบไม่ต่ำกว่า 100 ข้อ (Embretson and Reise, 2000)

## 2.2 การคัดเลือกข้อสอบข้อเริ่มต้น (Starting of the First Item)

การกำหนดจุดเริ่มต้นการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีหลายวิธี โดยทั่วไปจะใช้ข้อสอบข้อแรกมีค่าความยากปานกลาง จัดให้กับผู้สอบแต่ละคน เมื่อไม่ทราบค่าสารสนเทศเบื้องต้นหรือระดับความสามารถของผู้สอบ จึงจำเป็นต้องพิจารณาว่าจะใช้เกณฑ์ใดในการเริ่มต้นการทดสอบซึ่งแนวทางกำหนดเกณฑ์เริ่มต้นการทดสอบมีดังนี้ (Becker *et al.*, 2008; Thompson and Weiss, 2011; kirisci *et al.*, 2012) 1) ใช้ข้อสอบมีค่าความยากปานกลางจัดให้ผู้สอบแต่ละคนเมื่อไม่ทราบค่าสารสนเทศเบื้องต้นของผู้สอบ 2) ทำการทดสอบเบื้องต้นก่อนการทดสอบจริงและนำผลมากำหนดจุดเริ่มต้นการทดสอบ 3) หากไม่มีข้อมูลอื่นประกอบการพิจารณา นิยมใช้ค่าเฉลี่ยคุณลักษณะแฝง (Latent Trait) หรือระดับความสามารถของประชากร (Average Theta) ซึ่งมีค่าเท่ากับ 0.00 เป็นจุดเริ่มต้นในการเลือกข้อสอบข้อแรกให้เหมาะสมกับระดับความสามารถของผู้สอบดังกล่าว 4) ใช้วิธีการสุ่มเลือกข้อสอบด้วยโปรแกรมคอมพิวเตอร์ 5) หากมีผลการทดสอบของมาตรวัดนั้น ในครั้งก่อนหน้า อาจเริ่มต้นด้วยค่าคุณลักษณะแฝงหรือระดับความสามารถของผู้สอบที่ได้จากครั้งก่อน และ 6) พิจารณาจากข้อมูลอื่น เช่น ผลการทดสอบจากมาตรวัดอื่น เกรดเฉลี่ย แรงจูงใจในการทดสอบ และเศรษฐกิจฐานะทางสังคม เป็นต้น

## 2.3 การคัดเลือกข้อสอบข้อถัดไป (Selecting of the Next Item)

สำหรับการเลือกข้อสอบจากคลังข้อสอบเพื่อให้ผู้สอบตอบ โดยหลักการคัดเลือกข้อสอบมีเป้าหมายเพื่อลดความแปรปรวนของเวกเตอร์ความสามารถแบบพหุมิติชั่วคราว (Frey and Seitz, 2009) และวิธีการคัดเลือกข้อสอบที่ใช้กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติอาศัยภาวะน่าจะเป็นสูงสุดหรือต่ำสุดจากการประมาณค่าความสามารถ ( $\theta$ ) และมีวิธีการคัดเลือกข้อสอบอยู่หลากหลายวิธี แต่ละวิธีสามารถใช้ร่วมกันกับวิธีการประมาณค่าทุกแบบและยังสามารถใช้วิธีการคัดเลือกข้อสอบหลายวิธีร่วมกันได้อีกด้วย (Reckase, 2009) โดยมีวิธีการคัดเลือกข้อสอบแบบพหุมิติ คือ วิธี Maximum Likelihood และวิธี Bayesian (Segall, 1996, 2002, 2010) แต่วิธี Bayesian ได้รับความนิยมมากกว่า วิธี Maximum Likelihood (Segall, 2002; van der Linden, 2005) ในเวลาต่อมามีผู้พัฒนาเกณฑ์การเลือกข้อสอบขึ้นมาอย่างหลากหลายโดยอาศัยทฤษฎีของ Bayesian เช่น วิธี Maximize the Determinant of the Fisher Information Matrix (Segall, 1996; Frey and Seitz, 2009; Reckase, 2009) และ วิธี Maximize Kullback-Leibler Information (Chang and Ying, 1996; Veldkamp and van der Linden, 2002) แต่อย่างไรก็ตามในที่นี้จะอธิบายรายละเอียดเฉพาะ วิธี Maximize Kullback-Leibler Information เท่านั้น ผลจากตารางที่ 1 เป็นวิธีการคัดเลือกข้อสอบที่มีนักวิจัยสนใจศึกษา

**วิธี Maximize Kullback-Leibler Information** เป็นวิธีการ Posterior Expected Kullback-Leibler Information ถูกนำเสนอ ให้ใช้ในการคัดเลือกข้อสอบในการทดสอบแบบปรับเหมาะที่เป็นมิติเดียว โดย Chang and Ying (1996) และต่อมา Veldkamp and van der Linden (2002) ได้เสนอทางเลือกให้ใช้วิธีการ Posterior Expected Kullback-Leibler Information ในการคัดเลือกข้อสอบแบบปรับเหมาะแบบพหุมิติ (MCAT) โดยมีเกณฑ์การคัดเลือกข้อสอบจากระยะห่างของการกระจายของการตอบสนองจากข้อสอบ ข้อที่ถูกเลือกให้ใช้ทดสอบข้อต่อไป เป็นตำแหน่งที่ใช้ประมาณค่าความสามารถของผู้สอบ ( $\hat{\theta}$ ) กับความสามารถที่แท้จริงของผู้สอบ ( $\theta$ ) ด้วยการคาดหมายเอาจากการตอบข้อสอบ (Wang *et al.*, 2011, Mulder and van der Linden, 2009; cheng, 2009; Reckase, 2009) กรณีข้อสอบ 1 ข้อ คำนวณได้จากสมการที่ 3

$$K_i(\hat{\theta}, \theta) = -E \left[ \log \frac{f(u_i | \hat{\theta})}{f(u_i | \theta)} \right] \dots \dots \dots (3)$$

เมื่อ  $\int(u_i | \hat{\theta})$  และ  $\int(u_i | \theta)$  เป็นการแจกแจงความน่าจะเป็น 2 แพล่ง โดยทั่วไป  $\int(u_i | \hat{\theta})$  แทนการแจกแจงจริงของข้อมูล จากการสังเกตหรือความแม่นยำในการแจกแจงทางทฤษฎีการวัด  $\int(u_i | \theta)$  มักจะแสดงให้เห็นถึงทฤษฎี โมเดล การอธิบาย หรือประมาณของ  $\int(u_i | \hat{\theta})$  ซึ่งสารสนเทศ KL ไม่เคร่งครัดในการวัด

ระยะห่าง เนื่องจากไม่มีสมมาตร นั่นคือ  $\int (u_i | \hat{\theta}) \neq \int (u_i | \theta)$  โดยจะกล่าวถึงระยะห่างของ KL เพียงในกรณีทีสะท้อนถึงวิธีการลู่ออกหรือวิธีการที่ห่างกันของความน่าจะเป็นทั้ง 2 แหล่ง (cheng, 2009) และ Chang and Ying (1996) ได้แสดงฟังก์ชัน KL ไว้ คำนวณได้ดังสมการที่ 4

$$K_i(\hat{\theta}, \theta) = p_i(\hat{\theta}) \log \frac{P_i(\hat{\theta})}{P_i(\theta)} + Q_i(\hat{\theta}) \log \frac{Q_i(\hat{\theta})}{Q_i(\theta)} \dots\dots\dots(4)$$

ให้  $P_i(\theta)$  แทน ฟังก์ชันการตอบสนองข้อสอบสำหรับข้อที่ i และ  $Q_i(\theta) = 1 - P_i(\theta)$  ซึ่ง  $U_i(\theta)$  แทนผลการตอบข้อสอบข้อที่ i ถ้ากรณีข้อสอบ n ข้อ คำนวณได้ดังสมการที่ 5

$$K_n(\hat{\theta}; \theta) = \sum_{i=1}^n K_i \dots\dots\dots(5)$$

ในส่วน  $\theta$  และ  $\theta_n$  แทน 2 ระดับคุณลักษณะ โดย Chang and Ying สังเกตว่ามีคุณลักษณะที่สำคัญหลายประการของฟังก์ชัน KL ได้แก่ 1)  $KL_i(\theta \| \theta_0) \neq KL_i(\theta_0 \| \theta)$  2)  $KL_i(\theta \| \theta_0) \geq 0$  และ  $KL_i(\theta_0 \| \theta) \geq 0$  และ 3) คล้ายกับผลรวมของค่าฟังก์ชันสารสนเทศของฟิชเชอร์ กล่าวคือ สารสนเทศระดับคะแนนแบบสอบเป็นผลรวมของสารสนเทศระดับข้อสอบ คำนวณได้ดังสมการที่ 6

$$KL^n(\theta \| \theta_0) = \sum_{i=1}^n KL_i(\theta \| \theta_0) \dots\dots\dots(6)$$

ในขณะที่ GWIC คือ ฟังก์ชันถ่วงน้ำหนักที่สามารถประยุกต์ใช้กับ  $KL_i(\theta \| \theta_0)$  เพื่อที่จะหาเกณฑ์การคัดเลือกข้อสอบ ซึ่งฟังก์ชันถ่วงน้ำหนักทั่วไป คำนวณได้จากสมการที่ 7

$$w(\theta, \theta_0) = \begin{cases} 1, & \theta \in (\hat{\theta} - \delta_1, \hat{\theta} + \delta_2) \\ 0, & \text{otherwise} \end{cases} \quad \text{and } \theta_0 \in (\hat{\theta} - \delta_3, \hat{\theta} + \delta_4) \dots\dots\dots(7)$$

ซึ่งการประมาณค่าความสามารถ ( $\hat{\theta}$ ) เป็นความสามารถที่แท้จริง ( $\theta$ ) ตามข้อสอบ n ข้อ และพื้นที่ที่ใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบ ข้อสอบที่มีพื้นที่สูงสุดซึ่งเท่ากับสารสนเทศ  $KL_i(\theta \| \theta_0)$  ที่เฉลี่ยสูงสุดจะถูกคัดเลือกสามารถ คำนวณได้จากสมการที่ 8

$$KL_i(\hat{\theta}) = \int_{\theta_1}^{\theta_2} KL_i(\theta \| \hat{\theta}) d\theta \dots\dots\dots(8)$$

เมื่อ  $(\hat{\theta}_1, \hat{\theta}_2) = \left( \hat{\theta} - \frac{z}{\sqrt{n}}, \hat{\theta} + \frac{z}{\sqrt{n}} \right)$  และ  $\frac{z}{\sqrt{n}}$  เป็นลิมิตความเชื่อมั่น (Confidence Limit)

แต่อย่างไรก็ตามวิธีการนี้ค่อนข้างจะมีความซับซ้อนและจากการทดลองใช้งานของ Reckase โดยใช้โปรแกรม MATLAB ด้วยคุณลักษณะแฝง 3 คุณลักษณะ พบว่า การคำนวณทำได้ช้ามาก (Reckase, 2009)

#### 2.4) การประมาณค่าความสามารถ (Ability Estimation)

การประมาณค่าความสามารถของโมเดลการตอบสนองข้อสอบแบบพหุมิติ (MIRT) มีเกณฑ์แตกต่างกันหลายแนวทาง เช่น Maximum Likelihood Criterion, Maximum a Posteriori Bayesian Criterion, Newton-Raphson Method และ Least Squares Criterion (Reckase, 2009) แต่โดยทั่วไปการประมาณค่าความสามารถของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) ที่นิยมมีอยู่ 2 วิธี คือ วิธีการประมาณค่าแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) และวิธีการประมาณค่าแบบเบย์ (Bayesian Estimation) (Reckase, 2009) และจากข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระของข้อสอบและผู้สอบในทฤษฎีการตอบสนองข้อสอบ ถ้าตัวแปรความสามารถมีค่าคงที่การตอบสนองข้อสอบของผู้สอบต่อข้อสอบแต่ละข้อของแบบทดสอบ ถูกคาดหวังว่าเป็นอิสระจากกันในเชิงสถิติ

การประมาณค่าความสามารถด้วยเทคนิคการประมาณค่าแบบภาวะน่าจะเป็นสูงสุดก่อให้เกิดความไม่ชัดเจนในการประมาณค่าของเวกเตอร์ความสามารถ ณ จุดเริ่มต้นการทดสอบ แต่วิธีการประมาณค่าแบบเบย์จะหลีกเลี่ยงข้อบกพร่องนี้ แต่ก็ยังไม่มีข้อสรุปเกี่ยวกับปัญหาในการประมาณค่าความสามารถ ในทางปฏิบัติจึงมีการใช้วิธีการประมาณค่าทั้งแบบภาวะน่าจะเป็นสูงสุดและแบบเบย์ (Frey and Seitz, 2009) แต่อย่างไรก็ตามในที่นี้จะอธิบายรายละเอียดเฉพาะ วิธีการประมาณค่าแบบเบย์ (Bayesian Estimation) เท่านั้น

**วิธีการประมาณค่าแบบเบย์ (Bayesian Estimation)** วิธีการนี้คำนวณได้ดังสมการที่ 9 (Segall, 2010; Reckase, 2009)

$$f(\theta|u) = \frac{L(u|\theta)f(\theta)}{f(u)} \dots\dots\dots(9)$$

เมื่อ  $f(\theta|u)$  แทน ความน่าจะเป็นที่เกิดขึ้นหลังของ  $\theta$  ที่ทำให้เกิดอนุกรมคะแนนรายข้อ

$L(u|\theta)$  แทน ฟังก์ชันความน่าจะเป็น (Likelihood Function)

$f(\theta)$  แทน ความน่าจะเป็นที่เกิดขึ้นก่อนของ  $\theta$

$f(u)$  แทน ความน่าจะเป็นภายในขอบเขตของ  $u$  ซึ่ง  $f(u)$  คำนวณได้จากสมการที่ 10 และ 11

$$f(u) = \int_{-\infty}^{\infty} L(u|\theta)f(\theta)d\theta \dots\dots\dots(10)$$

$$f(\theta|u) = \frac{L(u|\theta)f(\theta)}{\int_{-\infty}^{\infty} L(u|\theta)f(\theta)d\theta} \dots\dots\dots(11)$$

โดยฟังก์ชันความน่าจะเป็น (Likelihood Function) ซึ่งคำนวณได้จากสมการที่ 12

$$L(u|\theta) \equiv L(u_{v_1}, u_{v_2}, \dots | \theta = \prod_{i \in v} P_i(\theta)^{u_i} Q_i(\theta)^{1-u_i}) \dots\dots\dots(12)$$

เมื่อ  $P_i(\theta)$  แทน ฟังก์ชันการตอบสนองของข้อสอบข้อที่  $i$  และ  $Q_i(\theta) = 1 - P_i(\theta)$

$v$  แทน เวกเตอร์ที่บรรจุข้อสอบแบบปรับเหมาะ

การประมาณค่าด้วยวิธี ML เป็นสมการที่ดำเนินการพร้อมกัน  $p$  มิติ ซึ่งคำนวณได้จากสมการที่ 13 (Segall, 1996)

$$\frac{\partial}{\partial \theta} \ln f(\theta|u) = 0 \dots\dots\dots(13)$$

อนุพันธ์แยกส่วนลำดับที่ 1 ซึ่งคำนวณได้จากสมการที่ 14 (Segall, 2010)

$$\frac{\partial}{\partial \theta} \ln f(\theta|u) = D \sum_{i \in s} v_i a_i - \phi^{-1}(\theta - \mu) \dots\dots\dots(14)$$

เมื่อผลรวมรายข้ออยู่ภายใน  $S$  และ  $v_i$  คำนวณได้จากสมการที่ 15

$$v_i = \frac{(p_i(\theta) - c_i)(u_i - p_i(\theta))}{(1 - c_i)p_i(\theta)} \dots\dots\dots(15)$$



วิธีการประมาณค่าแบบเบย์ มีข้อตกลงเบื้องต้นว่ามีการแจกแจงเป็นแบบปกติหลายตัวแปร (Multivariate Normal Distribution) ที่มีค่าเฉลี่ยเท่ากับ 0 และมีเมทริกซ์ความแปรปรวนร่วมเท่ากับ  $\phi$  คำนวณได้จากสมการที่ 16 (Segall, 2010)

$$f(\theta) = (2\pi)^{-\frac{H}{2}} |\phi|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\theta - \mu)' \phi^{-1} (\theta - \mu)\right] \dots\dots\dots(16)$$

เมื่อเวกเตอร์  $\mu = \{\mu_1, \mu_2, \dots, \mu_H\}$  และเมทริกซ์ความแปรปรวนร่วม  $H \times H$  ของ  $\phi$  การประมาณค่าที่ดี คือ ทำให้อยู่ในรูปแบบทั่วไป (Segall, 2010) คำนวณได้จากสมการที่ 17 และ 18

$$\theta^{(m+1)} = \theta^{(m)} - \delta^{(m)} \dots\dots\dots(17)$$

เมื่อ  $\theta^{(m)}$  และ  $\theta^{(m+1)}$  แทน ค่าความสามารถข้อที่ m และ m+1 ตามลำดับ

$$\delta^{(m)} = [M(\theta^{(m)})]^{-1} \times \frac{\partial}{\partial \theta} \ln f(\theta^{(m)} | u) \dots\dots\dots(18)$$

โดย  $f(\theta^{(m)} | u)$  แทน ความน่าจะเป็นที่เกิดขึ้นหลังของ  $\theta$  ที่ทำให้เกิดอนุกรมคะแนนรายชื่อ  $M(\theta^{(m)})$  แทน เมทริกซ์ทั้งเมทริกซ์ของอนุพันธ์แยกส่วนลำดับที่ 2 ของ  $J(\theta)$   
 คำนวณได้จากสมการที่ 19 และ 20

$$J_s(\theta) \equiv \frac{\partial^2}{\partial \theta \partial \theta'} \ln f(\theta | u) = D^2 \sum_{i \in S} a_i a_i' w_i - \phi^{-1} \dots\dots\dots(19)$$

เมื่อ D แทน ค่าคงที่ เท่ากับ 1.702

$a_i$  แทน เวกเตอร์พารามิเตอร์อำนาจจำแนกของข้อที่ i

$a_i'$  แทน เวกเตอร์  $1 \times p$  ของพารามิเตอร์อำนาจจำแนกข้อที่ i

$$w_i = \frac{Q_i(\theta)[P_i(\theta) - c_i][c_i u_i - P_i^2(\theta)]}{P_i^2(\theta)(1 - c_i)^2} \dots\dots\dots(20)$$

โดย  $c_i$  แทน พารามิเตอร์โอกาสการเดาข้อสอบได้ถูกต้องในการทำข้อสอบข้อที่ i

$u_i$  แทน ผลการตอบข้อสอบข้อที่ i ซึ่ง  $u_i = 1$  เมื่อตอบถูก และ  $u_i = 0$  เมื่อตอบผิด

$P_i(\theta)$  แทน ฟังก์ชันการตอบสนองข้อสอบข้อที่ i ได้ถูก และ  $Q_i(\theta) = 1 - P_i(\theta)$

ทั้งนี้เกณฑ์การประเมินระดับความสามารถของผู้สอบมีรายละเอียด ดังนี้ 1) ค่าประมาณความสามารถของผู้สอบ มากกว่า 2.0000 (ระดับความสามารถสูงมาก) 2) ค่าประมาณความสามารถของผู้สอบ ตั้งแต่ 1.0001 ถึง 2.0000 (ระดับความสามารถสูง) 3) ค่าประมาณความสามารถของผู้สอบ ตั้งแต่ 0.5001 ถึง 1.0000 (ระดับความสามารถค่อนข้างสูง) 4) ค่าประมาณความสามารถของผู้สอบ ตั้งแต่ -0.4999 ถึง 0.5000 (ระดับความสามารถปกติ) 5) ค่าประมาณความสามารถของผู้สอบ ตั้งแต่ -1.4999 ถึง -0.5000 (ระดับความสามารถค่อนข้างต่ำ) 6) ค่าประมาณความสามารถของผู้สอบ ตั้งแต่ -2.0001 ถึง -1.5000 (ระดับความสามารถต่ำ) และ 7) ค่าประมาณความสามารถของผู้สอบ ต่ำกว่า -2.0000 (ระดับความสามารถต่ำมาก)





จากการศึกษา ของ Diao and Reckase (2009) โดยใช้ข้อสอบจำนวน 50 ข้อ พบว่า การประมาณค่าด้วยวิธีนี้ประสบความสำเร็จในการประมาณค่าด้วยเวลาที่รวดเร็ว และจากตารางที่ 1 ก็พบว่าวิธีการประมาณค่าแบบเบย์ (Bayesian Estimation) เป็นวิธีการที่ได้รับความนิยมมากในการประมาณค่าความสามารถของผู้สอบ

## 2.5) เกณฑ์การยุติการทดสอบ (Termination Criterion)

การกำหนดเกณฑ์การยุติการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ขึ้นอยู่กับวัตถุประสงค์ของการทดสอบ ลักษณะคลังข้อสอบ และการควบคุมการปฏิบัติงาน เกณฑ์ในการพิจารณาการยุติการทดสอบมีหลายเกณฑ์ให้เลือกใช้ (Wainer *et al.*, 2001; Thompson and Weiss, 2011; Kuo *et al.*, 2015) ดังนี้ 1) ยุติการทดสอบเมื่อข้อสอบในคลังข้อสอบหมด 2) ยุติการทดสอบเมื่อการประมาณค่าคุณลักษณะแฝงมีการเปลี่ยนแปลงเล็กน้อยหรือไม่เปลี่ยนแปลงเลย 3) กฎความยาวคงที่ (Fixed-Length) และ จำนวนข้อสอบผันแปร (Variable Length) และ 4) การพิจารณาจากค่าความคลาดเคลื่อนมาตรฐาน (Standard Error of estimation: SEE) โดยทั่วไปจะใช้กฎการยุติการทดสอบจากค่าความคลาดเคลื่อนมาตรฐานกำหนดไว้ที่ 0.30

### ประโยชน์และข้อจำกัดของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ประโยชน์ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Latu and Chapman, 2002) สรุปได้ดังนี้ 1) ช่วยเพิ่มประสิทธิภาพในการทดสอบโดยเฉพาะภาระงานทดสอบต่าง ๆ ได้แก่ การดำเนินการทดสอบ การกำหนดเกณฑ์และการให้คะแนน 2) แบบทดสอบที่เก็บไว้ในคลังข้อสอบในคอมพิวเตอร์จะมีความปลอดภัย (Security) ต่อการคัดลอกมากกว่า 3) ไม่มีข้อจำกัดเรื่องเวลา ผู้สอบสามารถทำแบบทดสอบในเวลาที่แตกต่างกันตามความพร้อมของแต่ละบุคคลภายในช่วงเวลาที่กำหนด 4) ผู้สอบไม่เกิดความท้อแท้ในการทำข้อสอบเนื่องจากการเป็นการคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของแต่ละบุคคล 5) การตรวจข้อสอบมีความชัดเจนขึ้น เนื่องจากไม่มีปัญหาที่เกิดจากกระดาษคำตอบ เช่น การทำเครื่องหมายไม่ชัดเจน ความไม่ชัดเจนในข้อที่เลือกเนื่องจากการลบเมื่อต้องการเปลี่ยนคำตอบ เป็นต้น 6) ข้อสอบสามารถตรวจให้คะแนนและให้ผลย้อนกลับ (Feedback) กับผู้สอบได้ทันที 7) สามารถคัดเลือกข้อสอบจากการตอบข้อสอบเบื้องต้นได้อย่างรวดเร็ว 8) ผู้สอบได้ทำข้อสอบที่มีความยาวเหมาะสมกับระดับความสามารถ และ 9) มีการเสนอข้อสอบให้ผู้สอบตอบในรูปแบบที่หลากหลายทั้งภาพนิ่ง ภาพเคลื่อนไหวและเสียงประกอบ

ข้อจำกัดของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Latu and Chapman, 2002) ดังนี้ 1) ความปลอดภัยของแบบทดสอบ (Test security) แม้ว่าจะมีนักวิจัยหลาย ๆ คน แสดงให้เห็นว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีความปลอดภัยของแบบทดสอบมากกว่าการทำแบบทดสอบบนโต๊ะ (Desk Drawer) แต่ก็มีนักวิจัยบางท่านโต้แย้งว่า ปัจจุบันมีวิธีการขโมยข้อสอบได้หลายวิธี เช่น Pager กล้องขนาดเล็ก (Miniature Cameras) กล้องวิดีโอ (Video Transmitters) เครื่องบันทึกวิดีโอขนาดเล็ก (Micro Video Recorders) นอกจากนี้ ผู้สอบสามารถคัดลอกข้อสอบจากเครื่องคอมพิวเตอร์ได้ง่ายและรวดเร็วกว่าคัดลอกข้อสอบจากกระดาษคำตอบอีกด้วย 2) ปัญหาเกี่ยวกับผู้สอบ (Examinee Issues) แม้การทดสอบแบบปรับเหมาะจะมีประโยชน์ต่อผู้สอบมาก แต่ยังคงมีข้อจำกัดเกี่ยวกับผู้สอบ เช่น ผู้สอบหลายคนไม่เคยชินกับการใช้คอมพิวเตอร์ ทำให้เกิดความวิตกกังวล ผู้สอบบางคนแสดงให้เห็นว่า การนำเสนอโดยใช้คอมพิวเตอร์เป็นสิ่งที่ยุ่งยาก (Difficult or Fatiguing) โดยเฉพาะการตอบโดยใช้เมาส์หรือคีย์บอร์ด และ 3) ค่าใช้จ่ายในการดำเนินการ (Financial realities) การทดสอบแบบปรับเหมาะมีค่าใช้จ่ายในการดำเนินการสร้างแบบทดสอบเตรียมการและการดำเนินการสอบที่ค่อนข้างสูง

เนื่องจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติมีประโยชน์และข้อจำกัดเช่นเดียวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติที่กล่าวไว้ข้างต้น แต่ประโยชน์ที่เหนือกว่า คือ มีอัตราการวัดที่แม่นยำกว่าเมื่อใช้ข้อสอบเท่ากัน สามารถลดจำนวนข้อสอบลงได้มากกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ 30-50% และมีประสิทธิภาพสูงกว่าถึง 1.3 เท่าและสามารถลดจำนวนข้อสอบลงได้มากกว่าการทดสอบแบบดั้งเดิมที่กำหนดจำนวนข้อสอบไว้ 70% โดยไม่สูญเสียความแม่นยำ (Frey and Seitz, 2009)

## ทิศทางในอนาคตของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติเป็นการทดสอบที่มีความซับซ้อนและต้องอาศัยความรู้และความสามารถจากผู้เชี่ยวชาญในหลากหลายสาขา ซึ่งการพัฒนานั้นไม่ใช่เรื่องง่ายในต่างประเทศ มีการศึกษาและพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในการวัดมิติด้านพุทธิพิสัยมานานทำให้เกิดองค์ความรู้และความเข้าใจในการพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการวัดความสามารถ ซึ่ง Reckase (2009) กล่าวว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ยังเป็นเรื่องใหม่สำหรับประเด็นที่น่าสนใจในอนาคต เนื่องจากงานวิจัยที่ดีพิมพ์ออกมานั้นจะศึกษาเฉพาะกรณีที่เป็น 2 มิติ ซึ่งอาจจะไม่ตรงกับสภาพความเป็นจริงที่จะนำไปใช้ทางปฏิบัติได้ ดังนั้นจึงควรจะศึกษากรณีที่มี 3 มิติ หรือมากกว่า เพื่อจะหาข้อสรุปที่มีความเหมาะสมกับการทดสอบที่มีความหลากหลายมากขึ้น นอกจากนี้ งานวิจัยที่ศึกษาในปัจจุบันจะใช้การจำกัดจำนวนข้อสอบในการยุติการทดสอบ ซึ่งอาจจะได้สารสนเทศน้อยหรือหากจะใช้วิธีดังกล่าวควรจะมีการศึกษาจำนวนความยาวของข้อสอบที่เหมาะสมสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติและควรมีการพัฒนาแนวทางสำหรับการสร้างคลังข้อสอบให้มีความเหมาะสม โดยการศึกษาจำนวนข้อสอบ และลักษณะข้อสอบ โดยคำนึงถึงจำนวนมิติของการทดสอบว่ามีผลต่อจำนวนข้อสอบในคลังข้อสอบอย่างไร รวมถึงศึกษาความเหมาะสมค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบ

ในส่วนของประเทศไทยนั้นการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติเป็นประเด็นหนึ่งที่ได้รับ ความสนใจจากนักวัดผลการศึกษาโดยมีการศึกษาทั้งในเชิงเทคนิควิธีการ เช่น การพัฒนาวิธีการจัดคลังข้อสอบ (ศักดิ์ชัย และคณะ, 2561; พ็ชรภรณ์ และไพรัตน์, 2561) การเลือกข้อสอบ (ทัศนศิริจันทร์, 2554) การประเมินประสิทธิภาพการทดสอบ (สุชาติ และคณะ, 2559) การศึกษาวิธีการประมาณความสามารถ วิธีการคัดเลือกข้อสอบและเกณฑ์ยุติการทดสอบ (พงษ์พิชิต และคณะ, 2555) รวมทั้งการศึกษาในเชิงการพัฒนาแบบทดสอบจากรูปแบบกระดาษดินสอไปสู่การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ เพื่อวัดความสามารถด้านต่าง ๆ เช่น การทดสอบความถนัดทางการเรียน (นฤมล, 2558) การทดสอบ Pre O-NET กลุ่มสาระการเรียนรู้ภาษาไทย ระดับชั้นมัธยมศึกษาปีที่ 3 (พัชรภรณ์ และไพรัตน์, 2561) การทดสอบความรู้ทางไอซีที (ธีรวันซ์ และคณะ, 2562) และ การทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 (ศักดิ์ชัย, 2562) อย่างไรก็ตาม จะเห็นได้ชัดว่าในประเทศไทยนั้นได้ศึกษาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติเพื่อวัดความสามารถของผู้สอบ ซึ่งเป็นข้อสอบที่เป็น การตรวจให้คะแนนแบบ 2 ค่า (ตอบถูกได้ 1 ตอบผิดได้ 0) และมีการนำวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติไปใช้กับมาตรวัดด้านเจตพิสัยที่เป็นการตรวจให้คะแนนแบบมากกว่า 2 ค่า เช่น การพัฒนามาตรวัดสมรรถนะความรู้วิชาซีพพหุมิติของนักศึกษาครูช่างอุตสาหกรรม (สุกัญญา และคณะ, 2560) และการพัฒนามาตรวัดเจตคติต่อวิชาชีพครู (ชญารัตน์ และคณะ, 2561) ซึ่งผู้ที่ศึกษาส่วนใหญ่จะเป็นนักวัดผลการศึกษาทางการศึกษาซึ่งมีความรู้ความสามารถทางด้านกรวัดและประเมินผลทางการศึกษาเป็น อย่างดี ประเด็นที่น่าสนใจในที่นี่เป็นเพียงบางส่วนของความสนใจจากนักวัดผลการศึกษาทางการศึกษาที่ ทำการศึกษาและพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติในการดำเนินการทดสอบใน สถานการณ์จริงในยุค 4.0 ที่สังคมไทยเข้าสู่สังคมดิจิทัล และทันต่อความก้าวหน้าทางวิทยาการและวิจัยทำให้องค์ความรู้ในด้านนี้ซัดเจนมากขึ้นเรื่อย ๆ นักวิจัยต่างพยายามหาทางลดข้อจำกัดต่าง ๆ และพัฒนาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติให้สมบูรณ์

## บทสรุป

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) เป็นการผสมผสานแนวคิดระหว่างทฤษฎีการตอบสนองข้อสอบแบบพหุมิติกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) มีองค์ประกอบที่ คล้ายกันกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ซึ่งมีองค์ประกอบสำคัญ 5 องค์ประกอบ ได้แก่ 1) การสร้างคลังข้อสอบ 2) การเริ่มต้นการทดสอบ 3) การคัดเลือกข้อสอบข้อถัดไป 4) การประมาณค่าความสามารถของผู้สอบ และ 5) เกณฑ์ยุติการทดสอบ แต่มีข้อแตกต่างที่ข้อสอบแต่ละข้อของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติสามารถวัดความสามารถของผู้สอบมากกว่า

1 มิติ แต่การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ข้อสอบแต่ละข้อวัดความสามารถของผู้สอบเพียงมิติเดียว จึงทำให้การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติให้ค่าสารสนเทศได้มากกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ซึ่งเป็นการทดสอบที่มีประสิทธิภาพสูง มีความแม่นยำ และลดจำนวนข้อสอบ

### เอกสารอ้างอิง

- ชญารัตน์ บุญพุกกร เสรี ชัดเข้ม และ ปิยะทิพย์ ประดุงพรม. 2561. การพัฒนามาตรวัดเจตคติต่อวิชาชีพครูตามโมเดลการตอบสนองข้อสอบแบบพหุมิติ: การทดสอบแบบพหุมิติด้วยคอมพิวเตอร์. วารสารวิชาการสถาบันเทคโนโลยีแห่งสุวรรณภูมิ. 5(1): 187-204.
- ทัศน์ศิริรินทร์ สว่างบุญ. 2554. การเปรียบเทียบวิธีการคัดเลือกข้อสอบขั้นแรกและลำดับข้อสอบที่มีคุณภาพของการทดสอบปรับเหมาะแบบพหุมิติด้วยคอมพิวเตอร์. วิทยานิพนธ์ ค.ศ. จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพมหานคร.
- ธีรวัช สุขวิสัยศิริภู เสรี ชัดเข้ม และ ปิยะทิพย์ ประดุงพรม. 2562. การพัฒนาโปรแกรมการทดสอบปรับเหมาะแบบพหุมิติด้วยคอมพิวเตอร์ สำหรับทดสอบความรู้ไอซีที. วารสารวิชาการครุศาสตร์อุตสาหกรรม พระจอมเกล้าพระนครเหนือ. 10(1): 98-109.
- นฤมล ขุนไกร. 2558. การพัฒนาโปรแกรมการทดสอบความถนัดทางการเรียนแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภายใต้เงื่อนไขการทวนคำตอบที่แตกต่างกัน. วิทยานิพนธ์ปรัชญาดุษฎีบัณฑิต สาขาวิชาวิจัยและประเมินผลการศึกษา. มหาวิทยาลัยมหาสารคาม. มหาสารคาม.
- พงษ์พิชิต พรหมสิทธิ์ ไพศาล สุวรรณน้อย และ รังสรรค์ มณีเล็ก. 2555. การศึกษาวิธีการประมาณความสามารถ วิธีการคัดเลือกข้อสอบและเกณฑ์ยุติการทดสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ. วารสารศึกษาศาสตร์, 35(1): 38-47.
- พัชราภรณ์ ไวกุณฐ์วิวรรณ และ ไพรัตน์ วงษ์นาม. 2561. การประยุกต์ใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติในการพัฒนาคลังข้อสอบ Pre O-NET กลุ่มสาระการเรียนรู้ภาษาไทย ระดับชั้นมัธยมศึกษาปีที่ 3. วารสารบัณฑิตศึกษา มหาวิทยาลัยราชภัฏเชียงราย. 11(2): 17-27.
- ศักดิ์ชัย จันทะแสง เสรีชัดเข้ม และ ปิยะทิพย์. 2561. การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา. วิทยาการวิจัยและวิทยาการปัญญา. 16(2): 109-125.
- ศักดิ์ชัย จันทะแสง. 2562. การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 6. วารสารแม่โจ้เทคโนโลยีสารสนเทศและนวัตกรรม. 5(2): 50-69.
- สุกัญญา บุญศรีภมลวรรณ ตั้งธนกานนท์ และ ศิริชัย กาญจนวาสี. 2560. การพัฒนาคลังข้อสอบแบบหลายขั้นตอนเพื่อวัดสมรรถนะด้านความรู้วิชาชีพพหุมิติของนักศึกษาครูช่างอุตสาหกรรม. วารสารวิทยาลัยดุสิตธานี. 12(3): 600-615.
- สุชาติ หอมจันทร์ สมบัติ ท้ายเรือคำ และ บังอร กุมพล. 2559. ประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์โดยประยุกต์ใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติ. วารสารวิชาการและสังคมศาสตร์. 11(31): 73-88.
- Becker, J., Fliege, H., Kocalevent, R.D., Bjorner, J.B., Rose, M., Walter, O.B. and B.F. Klapp. 2008. Functioning and validity of A Computerized Adaptive test to measure anxiety (A-CAT). *Depression and Anxiety*. 25(12): E182-E192.
- Chang, H.H. and Z. Ying. 1996. A global information approach to computerized adaptive testing. *Applied Psychological Measurement*. 20(3): 213-229.
- Chen, P.H. 2009. Comparison of adaptive Bayesian estimation and weighted Bayesian estimation in multidimensional computerized adaptive testing. In: Weiss, D.J. (Ed.). *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. [Online]. Available: <https://www.psych.umn.edu/psylabs/CATCentral/>. (Retrieved June, 2020).
- Diao, Q. and M. Reckase. 2009. Comparison of Ability Estimation and Item Selection Methods in Multidimensional Computerized Adaptive Testing. Department of Measurement and Quantitative Methods. Michigan State University.
- Embretson, S.E. and S.P. Reise. 2000. *Item Response Theory for Psychologists*. Mahwah. Lawrence Erlbaum Associates Publications. New Jersey
- Frey, A. and N.N. Seitz. 2009. Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation*. 35(2): 89-94.
- Frey, A. and C.H. Carstensen. 2009. Diagnostic classification models and multidimensional adaptive testing: a commentary on rupp and templin. *Measurement: Interdisciplinary Research & Perspective*. 7(1): 58-61.
- Kirisci, L., Tarter, R., Reynolds, M., Ridenour, T., Stone, C. and M. Vanyukov. 2012. Computer adaptive testing of liability to addiction: Identifying individuals at risk. *Drug & Alcohol Dependence*. 123(1): S79-S86.

- Kuo, B.C., Daud, M. and C.W. Yang. 2015. Multidimensional computerized adaptive testing for indonesia junior high school biology. *Eurasia Journal of Mathematics, Science and Technology Education*. 11(5): 1105-1118.
- Latu, E. and E. Chapman. 2002. Computerised adaptive testing. *British journal of Educational Technology*. 33(5): 619-622.
- Mao, X., Ozdemir, B., Wang, Y. and T. Xiu. 2016. A comparison study of item exposure control strategies in MCAT. ERIC Institute of Education Resources (Online Submission).
- Mulder, J. and W.J. van der Linden. 2009. Multidimensional adaptive testing with Kullback–Leibler information item selection. p. 77–101. *In: W.J. van der Linden and C.A.W. Glas (Eds.). Elements of adaptive testing*. Springer. New York.
- Şahin, M.D. and S. Gelbal. 2020. Development of a multidimensional computerized adaptive test based on the bifactor model. *International Journal of Assessment Tools in Education*. 7(3): 323-342.
- Segall, D.O. 1996. Multidimensional adaptive testing. *Psychometrika*. 61(2): 331-354.
- Segall, D.O. 2002. Principles of multidimensional adaptive testing. p. 53–74. *In: W.J. van der Linden and C.A.W. Glas (Eds.). Computerized Adaptive Testing: Theory and Practice*. Kluwer. Netherlands.
- Segall, D.O. 2010. Principles of multidimensional adaptive testing. p. 57-75. *In: W.J. van der Linden and C.A.W. Glas (Eds.). Elements of Adaptive Testing*. New York: Dordrecht Heidelberg London.
- Reckase, M.D. 2009. *Multidimensional Item Response Theory*. Springer. New York.
- Thompson, N.A. and D.J. Weiss. 2011. A Framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation*. 16(1): 1-9.
- Van der Linden, W.J. 2005. Comparison of item-selection methods for adaptive tests with content constraints. *Journal of Educational Measurement*. 42(3): 283-302.
- Veldkamp, B.P. and W.J. van der Linden. 2002. Multidimensional adaptive testing with constraints on test content. *Psychometrika*. 67(4): 575-588.
- Veldkamp, B.P. and M. Matteucci. 2013. Bayesian computerized adaptive testing. *Ensaio: Avaliação e Políticas Publicasem Educação*. 21(78): 57-82.
- Wainer, H. 1989. The future of item analysis. *Journal of Educational Measurement*. 26(2). 191-208.
- Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B.F., Mislevy, R.J. and L. Steinberg. 2001. *Computerized adaptive testing: Aprimer*. *Qual Life Res*. 10(8): 733-734.
- Wang, C., Chang, H.H. and K.A. Boughton. 2011. Kullback–Leibler information and its applications in multi-dimensional adaptive testing. *Psychometrika*. 76(1): 13-39.