

## การวิเคราะห์ข้อสอบปรนัย: ทฤษฎีการตอบสนองข้อสอบ Item Analysis: Item Response Theory

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

### บทคัดย่อ

บทความนี้มีวัตถุประสงค์เพื่อให้เกิดความเข้าใจเกี่ยวกับการวิเคราะห์ข้อสอบปรนัยตามทฤษฎีการตอบสนองข้อสอบ ซึ่งเป็นกระบวนการวิเคราะห์คุณภาพข้อสอบ เพื่อนำข้อมูลการวิเคราะห์มาปรับปรุงข้อสอบเพื่อใช้ในโอกาสต่อไป การวิเคราะห์ข้อสอบสามารถวิเคราะห์เชิงปริมาณเป็นตัวเลข และการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ ได้แก่ ค่าอำนาจจำแนก ค่าความยาก และค่าโอกาสการเดา ข้อสอบที่ดีมีคุณภาพจะส่งผลให้การวัดผลมีความน่าเชื่อถือ ทำให้การตัดสินผลการสอบมีความยุติธรรมมากขึ้น และยังสามารถคัดเลือกข้อสอบที่ตรงตามเกณฑ์ที่กำหนดไว้เพื่อไปจัดทำคลังข้อสอบ

**คำสำคัญ:** การวิเคราะห์ข้อสอบ ข้อสอบปรนัย ทฤษฎีการตอบสนองข้อสอบ

### Abstract

The purpose of this article was to enable readers to understand the analysis of the multiple-choice exam based on item response theory which is the procedure of the quality analysis of exam in order to improve the exam for the next use. The exam analysis can be both analyzed of the analyzed quantitatively of number and analysis of exam quality according to the item response theory uses 3 parameters which are the difficulty, discrimination and guessing opportunity. A good quality exam results in a reliable educational measurement, making judgments fairer and can also select the exam that meets the specified criteria in order to prepare the item bank.

**Keywords:** item analysis, multiple choice exam, item response theory

### บทนำ

การวัดผลการศึกษาเป็นการตรวจสอบผู้เรียนว่า มีความรู้หรือคุณลักษณะที่ต้องการวัดอยู่ในระดับใด ซึ่งผลที่ได้จากการวัดมีความสำคัญต่อการพัฒนาการศึกษา การทดสอบเป็นวิธีการวัดผลการศึกษาวิธีหนึ่งที่นิยมใช้กันมาก เครื่องมือที่ใช้ในการทดสอบที่สำคัญก็คือ ข้อสอบชนิดต่าง ๆ ข้อสอบเป็นเครื่องมือหลักที่ใช้ในการวัดและประเมินผลทางการศึกษาเพื่อตรวจสอบว่า ผู้เข้าสอบมีคุณลักษณะแฝงหรือความสามารถอยู่ในระดับใด ซึ่งการสร้างข้อสอบที่ดีต้องคำนึงถึงความเที่ยงตรง (Validity) ความเที่ยง (Reliability) ความยาก (Difficulty) อำนาจจำแนก (Discrimination) และความเป็นปรนัย (Objectivity) เป็นสิ่งสำคัญ ทั้งนี้เพราะว่าความตรงเป็นคุณสมบัติของแบบทดสอบที่แสดงถึงความสามารถในการวัดได้ถูกต้องแม่นยำ (ศิริชัย, 2555)

กระบวนการดำเนินการวิเคราะห์ข้อสอบมีขั้นตอนและระยะเวลาในการดำเนินการ คือ เริ่มต้นด้วยการออกข้อสอบ จัดพิมพ์ (กระดาษ หรือการบันทึกข้อมูลอัตโนมัติ (Automated Data Collection: ADC)) ตรวจสอบความถูกต้อง จัดพิมพ์เป็นชุด แล้วนำมาทดสอบกับนักเรียนหรือกลุ่มตัวอย่าง นำกระดาษคำตอบมาตรวจให้คะแนน และนำเอาคะแนนที่ได้มาวิเคราะห์ข้อสอบก่อนนำข้อสอบไปใช้หรือจัดเก็บข้อสอบต่อไป โดยส่วนใหญ่แต่ละโรงเรียน อาจารย์ผู้สอนหรืออาจารย์ประจำวิชาเป็นผู้ออกข้อสอบเอง หลังจากพิมพ์ข้อสอบแล้ว

<sup>1</sup> สำนักงานประกันคุณภาพการศึกษา มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ กรุงเทพฯ 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: October 6, 2020, Accepted: January 10, 2020, Published: January 27, 2021



ทดสอบนักเรียนโดยไม่ได้ผ่านการวิเคราะห์ข้อสอบก่อนและหลังการสอบ และไม่ได้การจัดเก็บข้อสอบอย่างเป็นระบบ ทำให้เมื่อถึงเวลาที่ต้องทำการวัดและประเมินผู้เรียนอีกครั้งผู้สอนต้องออกข้อสอบใหม่ หรือบางวิชาอาจนำข้อสอบเดิมที่บันทึกไว้แล้วนำมาสอบใหม่โดยที่ไม่ได้ตัดข้อสอบที่ไม่มีคุณภาพออก เมื่อนำไปใช้ก็อาจเกิดผลเสียหลายได้ นอกจากนี้ในกระบวนการต่าง ๆ ยังทำให้เกิดการสิ้นเปลืองทรัพยากรจำนวนมากในการดำเนินการสอบแต่ละครั้ง

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่แท้จริงของบุคคลกับพฤติกรรมการตอบสนองข้อสอบในแต่ละข้อว่ามีความน่าจะเป็นในการตอบถูกได้มากน้อยเพียงใดโดยใช้โค้งลักษณะของข้อสอบ (Item Characteristic Curves: ICCs) ซึ่งมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ เรียกว่า ฟังก์ชันโลจิสติก (Logistic Function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal Ogive Function) ซึ่งทฤษฎีการตอบสนองข้อสอบ (IRT) ทำให้สามารถประมาณค่าตามระดับความรู้ของผู้สอบได้ยืดหยุ่นกว่าเดิม แต่การนำข้อสอบไปเก็บในฐานข้อมูลหรือคลังข้อสอบได้นั้นจะต้องทำการวิเคราะห์ข้อสอบ (Item Analysis) (โชติกา, 2555) เป็นการพิจารณาตรวจสอบความเหมาะสม ความถูกต้องของข้อสอบที่สร้างขึ้นเป็นรายข้อ เพื่อพิจารณาว่าข้อสอบแต่ละข้อมีความยาก (b) และค่าอำนาจจำแนก (a) เพียงใด รวมทั้งพิจารณาถึงประสิทธิภาพของตัวลงในข้อเลือกตอบของข้อสอบข้อนั้นด้วย ผลการวิเคราะห์จะทำให้ทราบว่าข้อสอบแต่ละข้อมีความเหมาะสมมากน้อยเพียงใด ข้อสอบที่มีคุณภาพจะสามารถนำไปวัดและประเมินผลได้อย่างเที่ยงตรง การวิเคราะห์ข้อสอบด้านคุณภาพโดยรวมจะใช้วิธีการให้ผู้รู้หรือผู้เชี่ยวชาญทางด้านการวัดผลเป็นผู้วินิจฉัยว่าคำถามนั้นมีความเที่ยงตรงตามเนื้อหาที่จะวัดหรือไม่ สอดคล้องกับ บุญเชิด (2526) การพิจารณาและตรวจสอบคุณภาพของข้อสอบ จำแนกเป็น 2 ลักษณะ คือ การวิเคราะห์ข้อสอบรายข้อและการวิเคราะห์ข้อสอบทั้งฉบับ การวิเคราะห์ข้อสอบรายข้อพิจารณาว่า ข้อสอบมีความยากง่ายและอำนาจจำแนกตามเกณฑ์หรือไม่ ส่วนการวิเคราะห์ข้อสอบทั้งฉบับเพื่อดูว่าข้อสอบมีความตรง และความเชื่อมั่นมากน้อยเพียงใด ข้อสอบใดที่มีอำนาจจำแนกรายข้อสูงเป็นส่วนมากแล้วคุณภาพทั้งฉบับก็จะมีแนวโน้มสูงตามด้วย และ บุญชม และคณะ (2524) กล่าวว่า การวิเคราะห์ข้อสอบเป็นกรรมวิธีการตรวจสอบคำตอบของนักเรียนในแต่ละข้อ เพื่อพิจารณาว่าข้อสอบแต่ละข้อมีระดับความยากง่ายและค่าอำนาจจำแนกความสามารถของนักเรียนได้มากน้อยเพียงใด ส่วนการวิเคราะห์ข้อสอบหรือการหาคุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) นั้น สามารถพิจารณาจากค่าพารามิเตอร์ของผู้สอบ ได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) ค่าโอกาสการเดา (c) และค่าความสามารถของผู้สอบ ( $\theta$ ) ที่ใช้อธิบายถึงลักษณะหรือคุณภาพของข้อสอบและความสามารถของผู้สอบ

ในบทความนี้ผู้เขียนขอแนะนำเสนอการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อคัดเลือกข้อสอบเข้าสู่คลังข้อสอบและสามารถพัฒนาคลังข้อสอบในรูปแบบอิเล็กทรอนิกส์เพื่อรองรับการทดสอบด้วยระบบคอมพิวเตอร์ต่อไป

## 1. ความรู้เบื้องต้นเกี่ยวกับข้อสอบแบบปรนัย

ข้อสอบแบบปรนัยนิยมนำมาใช้เป็นเครื่องมือทดสอบผู้เรียน โดยมีองค์ประกอบที่สำคัญ 2 ส่วน คือ โจทย์หรือข้อคำถามและตัวเลือก รายละเอียด ดังนี้ (Mullis and Martin, 2013; Stone, 2001; ศิริชัย, 2555)

1. โจทย์ (Stem) หรือ คำถาม (Lead-in) เป็นส่วนที่ให้ข้อมูลเพื่อใช้ในการตอบคำถาม อาจมี รูปภาพ กราฟหรือแผนภูมิประกอบ

2. ตัวเลือก (Options หรือ Alternative) เป็นคำ หรือ ข้อความที่ผู้ออกข้อสอบนำเสนอตามหลังจากโจทย์เพื่อให้ผู้สอบเลือกไปใช้ตอบคำถาม หรือเติมลงในช่องว่างในโจทย์ โดยทั่วไปอาจมี 4 ตัวเลือก หรือ 5 ตัวเลือก การมีจำนวนตัวเลือกน้อยเกินไป จะมีโอกาสเดาได้ถูกมากขึ้น ตัวเลือก 5 ข้อ จะมีความเที่ยง (Reliability) มาก อย่างไรก็ตามในบางกรณี การหาตัวเลือกที่ 5 เป็นเรื่องที่ยากมากหรือไม่มีโอกาสเป็นไปได้ก็อาจใช้เพียง 4 ตัวเลือก ส่วนการใช้ตัวเลือกมากกว่า 5 ข้อ ก็ไม่มีความจำเป็นและไม่ทำให้อำนาจจำแนกสูงขึ้น สอดคล้องกับงานวิจัยของ สิริพร (2545) ได้ทำการศึกษาวิจัยด้านจำนวนตัวเลือกที่ใช้ในกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 3 พบว่า ข้อสอบที่มีจำนวนตัวเลือกแตกต่างกัน มีผลต่อความยากมาตรฐานเฉลี่ยของข้อสอบอย่างมีนัยสำคัญ ข้อสอบแบบ 3 ตัวเลือกง่ายที่สุด อำนาจจำแนกต่ำที่สุดและมีความเที่ยงต่ำค่ากว่า

ข้อสอบแบบ 5 ตัวเลือก ขณะที่ข้อสอบแบบ 5 ตัวเลือกยากที่สุดแต่มีอำนาจจำแนกและความเที่ยงสูงกว่าข้อสอบแบบ 3 ตัวเลือก เมื่อเปรียบเทียบระหว่างข้อสอบแบบ 4 ตัวเลือก และ 5 ตัวเลือก พบว่า ค่าอำนาจจำแนกและความเที่ยงไม่แตกต่างกัน จึงได้ให้แนะนำในการนำไปใช้ว่า ในสถานการณ์การทดสอบทั่วไป อาจใช้ข้อสอบแบบ 4 ตัวเลือก หรือ 5 ตัวเลือก (รณิตา, 2551)

2.1 ตัวเลือกที่ถูกต้อง (Correct Answer หรือ Key) เป็นตัวเลือกที่ถูกต้องที่สุด มีเพียง 1 ตัวเลือกเท่านั้นในข้อสอบแต่ละข้อ

2.2 ตัวลวง (Distractors) เป็นตัวเลือกที่เป็นคำตอบที่ไม่ถูกต้อง มีไว้เพื่อลวงให้ผู้สอบที่ไม่มีความรู้หรือมีความเข้าใจไม่ถูกต้องเกี่ยวกับเนื้อหาที่นำมาออกข้อสอบเลือกตอบ การสร้างตัวลวงที่ดีจะทำให้ความเที่ยงของข้อสอบสูงขึ้นด้วย

## 2. ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)

ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ภายในบุคคล (Latent Trait or Ability) กับผลการตอบข้อสอบโดยใช้โค้งลักษณะข้อสอบ (Item Characteristic Curves: ICCs) ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก ( $b$ ) อำนาจจำแนก ( $a$ ) และโอกาสการเดาข้อสอบถูก ( $c$ ) โดยทฤษฎีการตอบสนองข้อสอบจึงอยู่บนพื้นฐานที่สำคัญ 2 ประการ คือ 1) ผลการตอบข้อสอบของผู้ตอบสามารถอธิบายได้ด้วยความสามารถที่มีอยู่ภายในตัวของผู้สอบ และ 2) ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายในตัวผู้สอบสามารถอธิบายได้ด้วยโค้งลักษณะข้อสอบ (ICCs) ซึ่งมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ เรียกว่า ฟังก์ชันโลจิสติก (Logistic Function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal Ogive Function) ที่สามารถสะท้อนความสัมพันธ์ระหว่างความสามารถของผู้สอบกับลักษณะของข้อสอบและผลการตอบข้อสอบของผู้สอบได้ซึ่งทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีการทดสอบแนวใหม่ ซึ่งจำแนกได้ 3 ประเภท คือ 1) ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (Binary IRT หรือ Dichotomous IRT) 2) ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT) และ 3) ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนแบบต่อเนื่อง (Continuous IRT) (ศิริชัย, 2555; De Ayala, 2009; Wainer *et al.*, 2000; Hambleton *et al.*, 1991) โดยในที่นี้จะกล่าวถึงเฉพาะทฤษฎีการตอบสนองข้อสอบแบบให้คะแนน 2 ค่า เท่านั้น

### 2.1 ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (Dichotomous IRT)

โมเดลการตอบสนองข้อสอบรูปแบบนี้เหมาะกับข้อมูลเมื่อตรวจให้คะแนน (Scored) แล้วมีคะแนนที่เป็นไปได้ 2 ค่า (Binary Data) คือ 0 กับ 1 (ตอบผิดได้ 0 ตอบถูกได้ 1) ลักษณะของการตรวจให้คะแนน 2 ค่า อาจได้มาจากการตอบข้อคำถามหลากหลายรูปแบบ เช่น ในแบบทดสอบวัดความสามารถที่มีตัวเลือก “ถูกหรือผิด” แบบวัดผลสัมฤทธิ์ทางการเรียนที่มีตัวเลือกแบบหลายตัวเลือก “4 ตัวเลือก หรือ 5 ตัวเลือก” แบบประเมินตนเองด้านบุคลิกภาพที่มีตัวเลือกว่า “เป็นจริงหรือไม่เป็นจริง” ในแบบทดสอบเจตคติที่มีตัวเลือกว่า “เห็นด้วยหรือไม่เห็นด้วย” และในมาตราประมาณค่าทางด้านพฤติกรรมที่มีตัวเลือกว่า “ใช่หรือไม่ใช่” เป็นต้น โดยทั่วไปการนำโมเดลการตอบสนองข้อสอบไปใช้ในการศึกษา ต้องมีการตรวจสอบข้อตกลงเบื้องต้นของโมเดลการตอบสนองข้อสอบ มี 4 ข้อ ดังนี้

1) **ความเป็นเอกมิติ (Unidimensionality)** หมายถึง ข้อคำถามหรือข้อสอบทุกข้อในเครื่องมือ (แบบทดสอบ) มุ่งวัดคุณลักษณะภายในของผู้สอบเพียงด้านใดด้านหนึ่ง การตรวจสอบความเป็นเอกมิติของแบบทดสอบ สามารถดำเนินการโดยใช้เทคนิคการวิเคราะห์ทางสถิติ ได้แก่ การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis) เพื่อคำนวณหาค่าไอแกน (Eigen Value) แล้วคำนวณอัตราส่วนระหว่างค่าไอแกนขององค์ประกอบแรกกับองค์ประกอบถัดไป ถ้าอัตราส่วนที่คำนวณได้มีค่าสูงแสดงว่า เครื่องมือหรือแบบทดสอบวัดคุณลักษณะเด่นเดียว (Single Dominant Factor) หรือใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) เพื่อตรวจสอบยืนยันว่า เครื่องมือหรือแบบทดสอบมุ่งวัดเพียงคุณลักษณะเดียว

**2) ความเป็นอิสระ (Local Independence) มี 2 ลักษณะ ดังนี้**

2.1) ความเป็นอิสระระหว่างข้อสอบ หมายถึง เมื่อมีการควบคุมความสามารถของผู้สอบให้คงที่แล้ว ผลการตอบข้อสอบแต่ละข้อของผู้สอบ จะต้องเป็นอิสระต่อกัน นั่นคือ ถ้าสุ่มผู้สอบที่มีความสามารถขึ้นมา 1 คน ซึ่งได้สอบข้อสอบ  $k$  ข้อ และให้เป็นผลการตอบข้อสอบข้อที่  $j$  เมื่อ  $j = 1, 2, \dots, k$  เมื่อควบคุมค่าความสามารถของผู้สอบ ( $\theta$ ) แล้วผลการตอบข้อสอบในแต่ละคู่ต้องไม่มีความสัมพันธ์กัน (อิสระต่อกัน) สามารถคำนวณได้จากสมการที่ 1 ดังนี้

$$P(U_1, U_2, \dots, U_k | \theta) = P(U_1 | \theta) P(U_2 | \theta) \dots P(U_k | \theta) \dots \dots \dots (1)$$

$$= \sum_{j=1}^k P(U_j | \theta)$$

โดยที่  $U_1, U_2, \dots, U_k$  คือ ผลการตอบข้อสอบข้อที่ 1, 2, ...,  $k$  เมื่อควบคุมค่าความสามารถของผู้สอบ ( $\theta$ )

ถ้าผลการตอบข้อสอบของผู้สอบคนหนึ่งเป็นอิสระต่อกัน แล้วค่าความน่าจะเป็นของรูปแบบในการตอบข้อสอบทุกข้อ จะเท่ากับผลคูณระหว่างค่าความน่าจะเป็นของผลการตอบข้อสอบแต่ละข้อ

2.2) ความเป็นอิสระระหว่างผู้สอบ หมายถึง เมื่อมีการควบคุมความสามารถของผู้สอบแต่ละคนให้คงที่แล้ว ผลการตอบข้อสอบระหว่างผู้สอบแต่ละคน จะต้องเป็นอิสระต่อกัน นั่นคือ ถ้าสุ่มข้อสอบขึ้นมา 1 ข้อ ในการตอบข้อสอบของผู้สอบ  $n$  คน และให้เป็นผลการตอบของข้อสอบของคนที่  $i$  เมื่อ  $i = 1, 2, \dots, n$  เมื่อควบคุมค่าความสามารถของผู้สอบ ( $\theta$ ) แล้วผลการตอบข้อสอบของผู้สอบแต่ละคู่ตั้งไม่มีความสัมพันธ์กัน (อิสระต่อกัน) สามารถคำนวณได้จากสมการที่ 2 ดังนี้

$$P(U_1, U_2, \dots, U_n | \theta) = P(U_1 | \theta) P(U_2 | \theta) \dots P(U_n | \theta) \dots \dots \dots (2)$$

$$= \sum_{i=1}^n P(U_i | \theta)$$

โดยที่  $U_1, U_2, \dots, U_n$  คือ ผลการตอบข้อสอบของคนที่ 1, 2, ...,  $n$  เมื่อควบคุมค่าความสามารถของผู้สอบ ( $\theta$ )

ถ้าผลการตอบข้อสอบข้อเดียวกันของผู้สอบแต่ละคนมีความเป็นอิสระต่อกัน แล้วความน่าจะเป็นของรูปแบบในการตอบข้อสอบข้อนั้นของผู้สอบทุกคน จะเท่ากับผลคูณระหว่างค่าความน่าจะเป็นของผลการตอบข้อสอบข้อนั้นของผู้สอบแต่ละคน

**3) โมเดลการตอบสนองข้อสอบ (Item Response Models) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างความน่าจะเป็นของการตอบข้อสอบถูกต้องกับความสามารถของผู้สอบ ซึ่งเป็นไปตามเส้นโค้งคุณลักษณะของข้อสอบโดยจะกล่าวถึงในหัวข้อถัดไป**

**4) การทดสอบที่ไม่แข่งขันด้านเวลา (Nonspeeded Test Administration) กล่าวคือ ทฤษฎีการตอบสนองข้อสอบถือว่า ความสามารถของผู้สอบเป็นปัจจัยที่สำคัญต่อผลการทดสอบ ความเร็วในการตอบจะต้องไม่มีอิทธิพลต่อผลการตอบ การจัดการทดสอบจึงไม่อยู่ในสถานการณ์ที่มีการแข่งขันด้วยเวลา ผู้สอบที่มีความสามารถจะต้องมีเวลาเพียงพอในการทำข้อสอบ (Power Test Administration) การตรวจสอบความเหมาะสมในด้านเวลา สำหรับการดำเนินการสอบ สามารถทำได้โดยพิจารณาจากร้อยละของจำนวนผู้สอบที่ทำข้อสอบได้ครบทุกข้อต่อผู้สอบทั้งหมด ถ้าเวลาในการดำเนินการสอบมีความเหมาะสม ผู้สอบส่วนใหญ่ร้อยละ 80 สามารถสอบข้อสอบได้ครบหรือเกือบครบทุกข้อ พิจารณาได้จากการเปรียบเทียบระหว่างความแปรปรวนของจำนวนข้อสอบที่เว้นกับความแปรปรวนของจำนวนข้อสอบที่ตอบผิด ถ้าอัตราส่วนความแปรปรวนมีค่าเข้าใกล้ 0 แสดงว่า เวลาในการสอบมีความเหมาะสม**

## 2.2 พารามิเตอร์ (Parameter) ในทฤษฎีการตอบสนองข้อสอบ (ศิริชัย, 2555)

### 1) พารามิเตอร์ของผู้สอบ (Person Parameter)

$\theta$  คือ ระดับความสามารถของผู้สอบ ซึ่งประมาณจากโมเดลการตอบสนองข้อสอบ นิยมปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 และค่า  $\theta$  มีพิสัยอยู่ระหว่าง  $(-\infty$  ถึง  $+\infty)$  แต่ผลการวิเคราะห์ส่วนใหญ่มักให้ค่าอยู่ในช่วง  $-3.00$  ถึง  $+3.00$  ค่าที่เป็นลบแสดงว่ามีความสามารถหรือระดับคุณลักษณะนั้นต่ำ และค่าที่เป็นบวก แสดงว่า มีความสามารถหรือระดับคุณลักษณะนั้นสูง โดยค่า  $P_i(\theta)$  = ค่าความน่าจะเป็นที่ผู้มีความสามารถ  $\theta$  จะตอบข้อสอบข้อที่  $i$  ถูก

### 2) พารามิเตอร์ของข้อสอบ (Item Parameter)

$b_i$  คือ ค่าพารามิเตอร์ความยากของข้อสอบข้อที่  $i$  (Difficulty Parameter) ซึ่งตำแหน่งของโค้งบนสเกลของความสามารถ ( $\theta$ ) ทำให้มีโอกาสตอบข้อสอบได้ถูกเท่ากับ  $\frac{1+c_i}{2}$  สำหรับโมเดล 1 พารามิเตอร์ และ

2 พารามิเตอร์  $P_i(\theta) = 0.50$  ส่วนโมเดล 3 พารามิเตอร์  $P_i(\theta) = \frac{1+c_i}{2}$  ในทางทฤษฎีการตอบสนองข้อสอบ (IRT) มีค่าระหว่าง  $-\infty$  ถึง  $+\infty$  แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่า  $b_i$  อยู่ระหว่าง  $-2.50$  ถึง  $+2.50$  ค่า  $b_i$  ที่อยู่ใกล้  $-2.50$  แสดงว่าเป็นข้อสอบที่ง่าย ส่วนค่า  $b_i$  ที่อยู่ใกล้  $+2.50$  แสดงว่าเป็นข้อสอบที่ยาก

$a_i$  คือ ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่  $i$  (Discrimination Parameter) การจำแนกค่าความต่าง  $P_i(\theta)$  ระหว่างผู้ที่มีความสามารถน้อยกว่าหรือเท่ากับ  $\theta$  กับผู้ที่มีความสามารถมากกว่า  $\theta$  มีค่าเป็นสัดส่วนโดยตรงของค่าความชันของ ICCs ที่ตำแหน่ง  $b_i$  ถ้าค่า  $a_i$  ที่สูง แสดงถึงการจำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี ในทางทฤษฎีการตอบสนองข้อสอบ (IRT) มีค่าระหว่าง  $-\infty$  ถึง  $+\infty$  แต่ในทางปฏิบัติมีค่าที่เป็นบวก อยู่ระหว่าง  $+0.50$  ถึง  $+2.50$  โดยปกติมีค่าไม่เกิน  $+2.50$

$c_i$  คือ ค่าพารามิเตอร์โอกาสการเดา (Guessing Parameter) ซึ่งโอกาสการตอบข้อสอบได้ถูกต้องของผู้สอบที่มีความสามารถต่ำจะเป็นค่ากำกับต่ำสุด (Lower Asymptote) ของ ICCs ในทางทฤษฎีการตอบสนองข้อสอบ (IRT) มีค่าอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปนิยมใช้ข้อสอบที่มีค่า  $c_i$  ไม่เกิน 0.30 ตามปกติควรมีค่าต่ำกว่าโอกาสในการตอบถูกโดยโอกาสการเดาตามทฤษฎีการทดสอบแบบดั้งเดิม (CCT)

### 3) ค่าคงที่ (Constant)

ค่าคงที่ ประกอบด้วย  $\exp$  หรือ  $e$  เป็นค่าคงที่ของลอการิทึมธรรมชาติ (Natural log) มีค่าเท่ากับ 2.71828 หรือ  $\approx 2.72$  และ  $D$  คือ ค่าองค์ประกอบของการปรับสเกล (Scaling factor) กำหนดให้เป็น 1 หรือ 1.70 เพื่อให้ฟังก์ชันโลจิสติก (Logistic Function) มีค่าใกล้เคียงกับฟังก์ชันปกติสะสม (Normal Ogive Function) หรือมีค่าต่างกันไม่เกิน 0.01

## 2.3 โมเดลการตอบสนองข้อสอบ (Item Response Models)

โมเดลการตอบสนองข้อสอบ เป็นโมเดลทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่าง โอกาสตอบข้อสอบถูก ( $P_i$ ) กับความสามารถที่มีอยู่ภายในของผู้สอบ ( $\theta$ ) ซึ่งอยู่ในรูปฟังก์ชันทางคณิตศาสตร์ 2 แบบ คือ ฟังก์ชันโลจิสติก (Logistic Function) และฟังก์ชันปกติสะสม (Normal Ogive Function) โดยโมเดลทั้ง 2 ฟังก์ชันนี้ให้ผลลัพธ์ของการประมาณค่าใกล้เคียงกันมากแต่ฟังก์ชันโลจิสติกจะได้รับความนิยมมากกว่า เนื่องจากมีลักษณะสูตรทางคณิตศาสตร์ ที่คำนวณง่ายและ สะดวกกว่าฟังก์ชันปกติสะสม โมเดลโลจิสติกมีหลายลักษณะขึ้นอยู่กับพารามิเตอร์ที่ปรากฏอยู่ในโมเดล ซึ่งมีตั้งแต่ 1 - 3 พารามิเตอร์ ดังแสดงในตารางที่ 1 และภาพที่ 1 โดยแต่ละโมเดลมีรายละเอียดดังนี้

### 1) โมเดลโลจิสติก 1 พารามิเตอร์ (One-Parameter Logistic Model: 1PL Model)

โมเดลโลจิสติก 1 พารามิเตอร์ หรือ ราสช์ โมเดล (Rasch Model) เป็นโมเดลที่แสดงความน่าจะเป็นของการตอบข้อสอบถูก สำหรับผู้สอบที่มีค่าความสามารถเท่ากับ  $\theta$  โดยมีข้อตกลงเบื้องต้นว่า ข้อสอบแต่ละข้อมีค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ ( $a$ ) เท่ากับ 1 มีค่าพารามิเตอร์โอกาสการเดาของข้อสอบ ( $c$ ) เท่ากับ 0 แต่มีความแตกต่างเฉพาะค่าพารามิเตอร์ความยากของข้อสอบ ( $b$ ) เท่านั้น โมเดลนี้

จะเหมาะสำหรับข้อสอบแบบอิงเกณฑ์ที่มีสลับซับซ้อน สามารถเขียนฟังก์ชันโลจิสติกและสามารถคำนวณได้จากสมการที่แสดงในตารางที่ 1

จากภาพที่ 1(ก) แสดงโค้งคุณลักษณะของข้อสอบ (ICCs) จากโมเดลโลจิสติก 1 พารามิเตอร์ ซึ่งเปรียบเทียบข้อสอบจำนวน 3 ข้อ ที่มีค่าความยากของข้อสอบ (b) แตกต่างกัน คือ -1, 0 และ 1 ตามลำดับ จากโค้งคุณลักษณะแสดงให้เห็นว่า 1) ค่าความน่าจะเป็นในการตอบข้อสอบถูกจะค่อย ๆ เพิ่มขึ้นตามค่าความสามารถของผู้สอบ 2) โค้งคุณลักษณะของข้อสอบทั้ง 3 เส้น มีความชันไม่แตกต่างกัน ถึงแม้มีค่าความยากของข้อสอบต่างกัน และ 3) จุดเปลี่ยนโค้งของเส้นโค้งคุณลักษณะของข้อสอบทั้ง 3 เส้น จะอยู่ตรงที่ค่าความน่าจะเป็นในการตอบถูกเท่ากับ 0.5 เสมอ

## 2) โมเดลโลจิสติก 2 พารามิเตอร์ (Two-Parameter Logistic Model: 2PL Model)

โมเดลโลจิสติก 2 พารามิเตอร์ เป็นโมเดลที่แสดงความน่าจะเป็นของการตอบข้อสอบถูกสำหรับผู้สอบที่มีค่าความสามารถเท่ากับ  $\theta$  โดยมีข้อตกลงเบื้องต้นว่า ข้อสอบแต่ละข้อมีพารามิเตอร์โอกาสการเดาของข้อสอบ (c) เท่ากับ 0 และมีความแตกต่างกันของพารามิเตอร์ความยากของข้อสอบ (b) กับพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) สามารถเขียนฟังก์ชันโลจิสติกและสามารถคำนวณได้จากสมการที่แสดงในตารางที่ 1

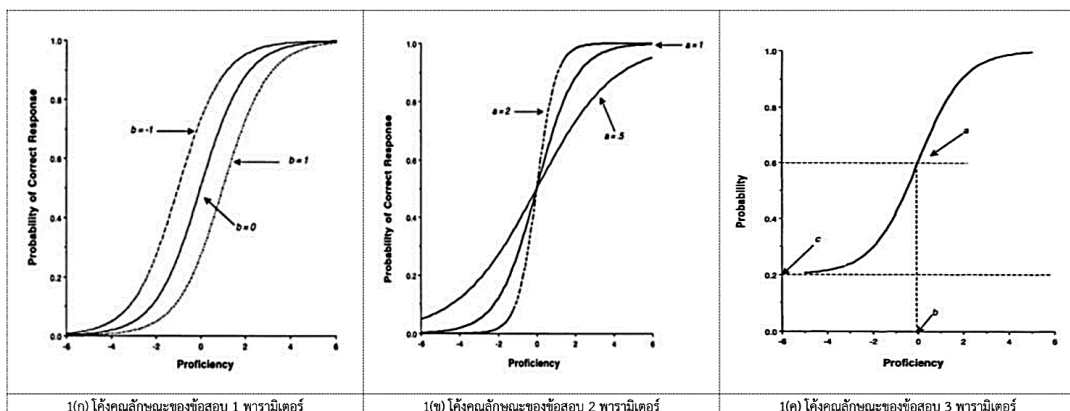
ตารางที่ 1 ตัวแบบทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ทั้ง 3 ตัวแบบ

Model	Logistic Function	Condition
1-Parameter 1PL (Rasch) Model (1)	$P_i(\theta) = \frac{1}{1 + e^{-(\theta - b_i)}}; i = 1, 2, \dots, n$	ข้อสอบแต่ละข้อมีค่า $c = 0$ และ ค่า $a = 1$ แต่มีความแตกต่างกันเฉพาะค่า $b$ เท่านั้น
2-Parameter 2PL Model (2)	$P_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}}; i = 1, 2, \dots, n$	ข้อสอบแต่ละข้อมีค่า $c = 0$ มีความแตกต่างกันของค่า $a$ และค่า $b$
3-Parameter 3PL Model (3)	$P_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta - b_i)}}; i = 1, 2, \dots, n$	ข้อสอบแต่ละข้อมีค่าพารามิเตอร์แตกต่างกันได้ทั้ง 3 ค่า คือ ค่า $a$ ค่า $b$ และค่า $c$

โดยมีนิยามเกี่ยวกับพารามิเตอร์และตัวแปรต่าง ๆ ดังนี้

- เมื่อ  $P_i(\theta)$  คือ ความน่าจะเป็นของผู้เข้าสอบที่มีความสามารถเท่ากับ  $\theta$  จะตอบข้อสอบข้อที่  $i$  ได้ถูกต้อง
- $\theta$  คือ ความสามารถของผู้เข้าสอบ  $a_i$  คือ ค่าอำนาจจำแนกของข้อสอบข้อที่  $i$
- $b_i$  คือ ค่าความยากของข้อสอบข้อที่  $i$   $c_i$  คือ ค่าการเดาของข้อสอบข้อที่  $i$
- D คือ ค่าองค์ประกอบของการปรับสเกล (Scaling Factor)

สำหรับทั้ง 3 ตัวแบบ นั้นสามารถเขียนเป็นกราฟ เพื่อแสดงให้เห็นการเปลี่ยนแปลงในแต่ละพารามิเตอร์ ดังแสดงในภาพที่ 1



ภาพที่ 1 โค้งคุณลักษณะของข้อสอบ (Wainer et al, 2000)

จากภาพที่ 1(ข) แสดงโค้งคุณลักษณะของข้อสอบ (ICCs) จากโมเดลโลจิสติก 2 พารามิเตอร์ ซึ่งเปรียบเทียบข้อสอบจำนวน 3 ข้อ ที่มีค่าความยากของข้อสอบ (b) ไม่แตกต่างกัน แต่มีค่าอำนาจจำแนกของข้อสอบ (a) แตกต่างกัน คือ 0.5, 1 และ 2 ตามลำดับ จากโค้งคุณลักษณะแสดงให้เห็นว่า ความชันของเส้นโค้งจะเปลี่ยนไปตามค่าอำนาจจำแนกของข้อสอบ กล่าวคือ ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง แล้วเส้นโค้งคุณลักษณะของข้อสอบจะมีความชันมาก และจุดเปลี่ยนโค้งของเส้นโค้งคุณลักษณะทั้ง 3 เส้น จะอยู่ตรงกับค่าความน่าจะเป็นในการตอบถูก เท่ากับ 0.5 เช่นเดียวกับ โค้งคุณลักษณะข้อสอบ แบบ 1 พารามิเตอร์

### 3) โมเดลโลจิสติก 3 พารามิเตอร์ (Three-Parameter Logistic Model: 3PL Model)

โมเดลโลจิสติก 3 พารามิเตอร์ เป็นโมเดลที่แสดงความน่าจะเป็นของการตอบข้อสอบถูก สำหรับผู้สอบที่มีความสามารถ  $\theta$  โดยมีข้อตกลงเบื้องต้นว่า ข้อสอบแต่ละข้อมีพารามิเตอร์ของข้อสอบที่แตกต่างกันทั้ง 3 พารามิเตอร์ ได้แก่ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ และค่าโอกาสการเดาของข้อสอบ โมเดลนี้จึงมีความเหมาะสมสำหรับข้อสอบที่มีรายการคำตอบแบบหลายตัวเลือก เนื่องจากผู้สอบสามารถเดาคำตอบได้ สามารถเขียนฟังก์ชันโลจิสติกและสามารถคำนวณได้จากสมการที่แสดงในตารางที่ 1

จากภาพที่ 1(ค) แสดงโค้งคุณลักษณะของข้อสอบ (ICCs) จากโมเดลโลจิสติก 3 พารามิเตอร์ของข้อสอบ 1 ข้อ จากโค้งคุณลักษณะจะเห็นว่าเส้นโค้งไม่ได้เริ่มจากศูนย์ แต่เริ่มจากค่าโอกาสการเดาของข้อสอบ นั้นแสดงให้เห็นว่าผู้สอบมีโอกาสตอบข้อสอบถูก ทั้งที่ไม่มีความสามารถในเรื่องที่ต้องการวัดการตอบข้อสอบถูกนั้นเกิดจากการเดาของผู้สอบ และจะสังเกตเห็นว่าจุดเปลี่ยนโค้งของเส้นโค้งคุณลักษณะจะไม่อยู่ที่ค่าความน่าจะเป็นในการตอบถูก เท่ากับ 0.5 เหมือนกับ 2 โมเดลแรก เนื่องจากโค้งคุณลักษณะของข้อสอบไม่ได้เริ่มจากศูนย์ ส่งผลให้โอกาสในการตอบข้อสอบถูก มีค่าสูงขึ้น

#### 2.4 คุณสมบัติของความไม่แปรเปลี่ยนของค่าพารามิเตอร์ (Invariance)

เมื่อโมเดลการตอบสนองข้อสอบมีความสอดคล้องกับข้อมูล (Model-Data Fit) จะทำให้เกิดคุณสมบัติที่สำคัญของทฤษฎีการตอบสนองข้อสอบ คือ คุณสมบัติความไม่แปรเปลี่ยนมี 2 ลักษณะ

1) ความไม่แปรเปลี่ยนของพารามิเตอร์ของข้อสอบ (Item Parameter) นั่นคือ ค่าพารามิเตอร์ของข้อสอบจะไม่เปลี่ยนแปลงไปตามกลุ่มผู้สอบ (โค้งคุณลักษณะข้อสอบ ICCs จะมี ลักษณะเดียวกัน ทุกกลุ่มความสามารถของผู้สอบ)

2) ความไม่แปรเปลี่ยนของพารามิเตอร์ความสามารถของผู้สอบ (Ability Parameter) นั่นคือ ค่าพารามิเตอร์ความสามารถของผู้สอบจะไม่เปลี่ยนแปลงไปตามชุดของข้อสอบ

### 3. การวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ

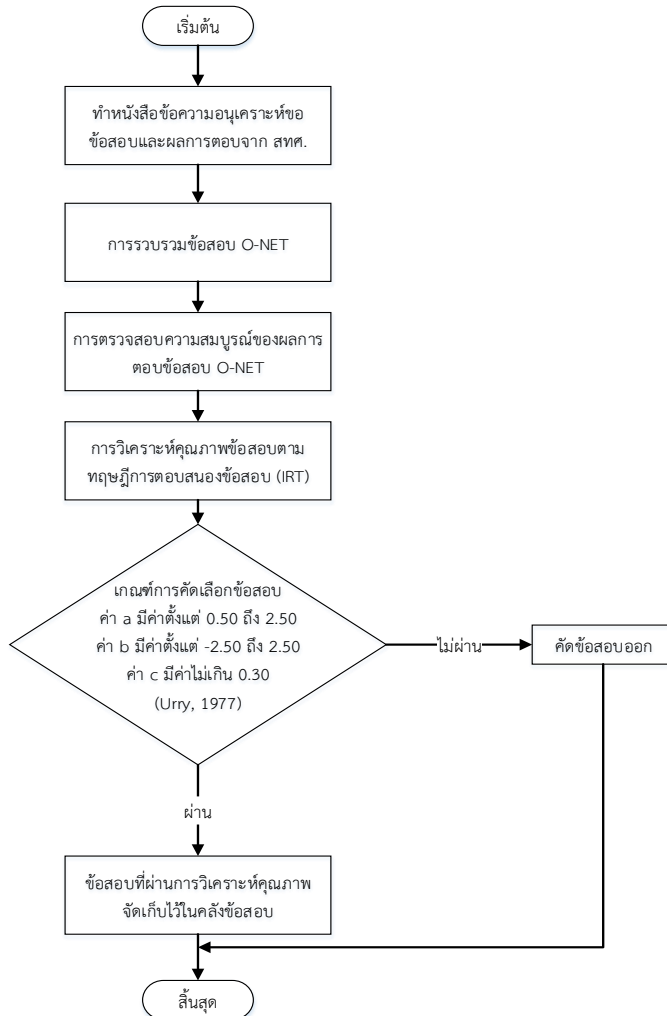
การวิเคราะห์หาคุณภาพข้อสอบเป็นตัวบ่งชี้ที่สำคัญในการสะท้อนความสามารถของข้อสอบในการวัดในสิ่งที่มุ่งวัด ซึ่งก็คือ ความรู้หรือความสามารถของผู้สอบได้อย่างถูกต้อง แม่นยำ น่าเชื่อถือ ในส่วนนี้ผู้เขียนได้นำเสนอขั้นตอนการวิเคราะห์คุณภาพข้อสอบ O-NET เพื่อการจัดทำคลังข้อสอบ มีขั้นตอนแสดงตามภาพที่ 2

จากภาพที่ 2 แสดงขั้นตอนการวิเคราะห์คุณภาพข้อสอบ O-NET เพื่อจัดทำคลังข้อสอบ ซึ่งมีรายละเอียดในการดำเนินการดังนี้

3.1 ทำหนังสือขอความอนุเคราะห์ขอข้อมูลเพื่อการวิจัยจากหน่วยงานต้นสังกัดถึงผู้อำนวยการสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) โดยมีความประสงค์ขอความอนุเคราะห์ในการขอข้อสอบ O-NET ในระดับชั้นประถมศึกษาปีที่ 6 ระดับชั้นมัธยมศึกษาปีที่ 3 และระดับชั้นมัธยมศึกษาปีที่ 6 ซึ่งในแต่ละระดับชั้นจะมีข้อสอบ 8 กลุ่มสาระการเรียนรู้ ประกอบด้วย 1) ภาษาไทย 2) คณิตศาสตร์ 3) วิทยาศาสตร์ 4) ภาษาต่างประเทศ 5) ศิลปะ 6) สุขศึกษาและพลศึกษา 7) สังคมศึกษา ศาสนาและวัฒนธรรม และ 8) การงานอาชีพและเทคโนโลยี พร้อมผลการตอบและเฉลย

3.2 การรวบรวมข้อสอบ O-NET ที่ได้จากสถาบันทดสอบการศึกษาแห่งชาติ (องค์การมหาชน) ระหว่าง ปี พ.ศ. ที่ผู้วิจัยต้องการศึกษา

3.3 ตรวจสอบความสมบูรณ์ของข้อสอบ O-NET ทั้งข้อคำถาม ตัวเลือก และเฉลย เสร็จแล้วเตรียมข้อมูลจากผลการตอบข้อสอบเพื่อทำการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์



ภาพที่ 2 ขั้นตอนการวิเคราะห์คุณภาพข้อสอบ O-NET เพื่อจัดทำคลังข้อสอบ

3.4 การวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) มีโปรแกรมสำเร็จรูปสำหรับการวิเคราะห์ข้อสอบที่พัฒนาขึ้นเพื่อใช้สำหรับการวิเคราะห์ข้อมูลที่ได้มาจากแบบทดสอบ แบบประเมินแบบสอบถาม หรือแบบสำรวจต่าง ๆ ซึ่งสร้างขึ้นเพื่อใช้เก็บรวบรวมข้อมูลตัวแปรที่แฝงที่สนใจศึกษา ข้อมูลทางการวัดบางส่วนสามารถใช้โปรแกรมสำเร็จรูปสำหรับการวิเคราะห์ข้อมูลทางสถิติทั่วไป (เช่น SPSS) ในการวิเคราะห์ได้ ขณะที่การวิเคราะห์ข้อสอบ/ข้อคำถามส่วนใหญ่ต้องใช้โปรแกรมที่มีความเฉพาะเจาะจง การวิเคราะห์ข้อมูลการวัดแปรตามโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า มีหลากหลายโปรแกรม แต่ละโปรแกรมมีจุดเด่นและข้อจำกัดแตกต่างกันไปขึ้นอยู่กับการใช้งาน ในที่นี้ขอเสนอโปรแกรมสำเร็จรูปที่ใช้ในการวิเคราะห์ ได้แก่ Xcaliber, IRTPRO และ BILOG MG

1) Xcaliber (Assessment Systems Corporation, 2018) เป็นโปรแกรมสำเร็จรูปสำหรับการวิเคราะห์ข้อสอบตามแนวคิดของโมเดลการตอบสนองข้อสอบ ซึ่งสามารถวิเคราะห์โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (Dichotomous IRT Model) ได้แก่ One-parameter



Logistic Model (1PL), Two-parameter Logistic Model (2PL), Three-parameter Logistic Model (3PL) และ โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT Model) ได้แก่ Rating Scale Model (RSM), Partial credit Model (PCM), Generalized Rating Scale Model (GRSM), Generalized Partial credit Model (GPCM) และ Graded Response Model (GRM) รวมทั้งสามารถวิเคราะห์ข้อสอบในกรณีที่เป็นวัด/มาตรฐานวัดมีการผสมผสานรูปแบบของข้อสอบ/ข้อคำถามระหว่างการตรวจให้คะแนนแบบ 2 ค่า กับแบบมากกว่า 2 ค่า (Mixed-format Test)

2) IRTPRO (Cai et al., 2018) เป็นโปรแกรมสำเร็จรูปที่พัฒนาขึ้นมาใหม่ล่าสุด ซึ่งสามารถวิเคราะห์ข้อสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบได้อย่างครอบคลุม ไม่ว่าตัวแปรนั้นจะมีโครงสร้างแบบเอกมิติ (Unidimensional) หรือพหุมิติ (Multidimensional) ตรวจให้คะแนนแบบ 2 ค่า (ได้แก่ 1PL, 2PL, 3PL) หรือมากกว่า 2 ค่า (ได้แก่ GRM, GPCM, NRM) รวมทั้งสามารถวิเคราะห์ข้อสอบที่มีการผสมผสานรูปแบบการตรวจให้คะแนนทั้ง 2 แบบ ด้วย

3) BILOG-MG (Du Toit, 2003) เป็นโปรแกรมสำเร็จรูปที่พัฒนาขึ้นเพื่อใช้สำหรับการวิเคราะห์ข้อสอบตามแนวคิดทฤษฎีการตอบสนองข้อสอบ โดยส่วนขยายของโปรแกรม BILOG-MG ที่ออกแบบมาเพื่อการวิเคราะห์ข้อสอบแบบ การวิเคราะห์ข้อสอบแบบตรวจให้คะแนน (Scored) แล้วมีคะแนนที่เป็นไปได้ 2 ค่า (Binary Data) คือ 0 กับ 1 (ตอบผิดได้ 0 ตอบถูกได้) ซึ่งโปรแกรม BILOG-MG มีข้อจำกัดคือสามารถวิเคราะห์ค่าพารามิเตอร์ได้ 2 ค่า คือ พารามิเตอร์ค่าความยาก ( $b_i$ ) และค่าพารามิเตอร์ค่าอำนาจจำแนก ( $a_i$ ) ถ้าจะหาค่าพารามิเตอร์โอกาสการเดา ก็จะนิยมใช้โปรแกรม TESTFACT

4) IRT-BAY (ศิริชัย และ คณิต, 2535) เป็นโปรแกรมที่เขียนโดยคนไทย โปรแกรมที่ใช้กับระบบปฏิบัติการ DOS ด้วยเนื้อที่เพียง 270 MB ใช้กับข้อสอบที่ตรวจให้คะแนนแบบผิดให้ 0 คะแนน ถูกให้ 1 คะแนน วิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) สามารถประมาณค่าความสามารถของผู้สอบ (q) ค่าอำนาจจำแนกของข้อสอบ (a) 2) ค่าความยากของข้อสอบ (b) และ 3) ค่าโอกาสการเดาของข้อสอบ (c) โดยใช้วิธีการประมาณค่าแบบ Bayesian และสามารถวิเคราะห์ข้อสอบรายข้อตามทฤษฎีดั้งเดิม (Classical Test Theory: CCT) โดยแสดงค่าความยาก (p) ค่าอำนาจจำแนก (Rbis, Rpbis) ความเชื่อมั่น (KR-20) และแสดงโค้งการสนเทศของข้อสอบ

ในการเลือกใช้โปรแกรมการวิเคราะห์คุณภาพข้อสอบนั้น การเตรียมข้อมูลในการวิเคราะห์ข้อสอบไม่แตกต่างกันเท่าใด แต่ที่ต่างกัน คือ การได้มาซึ่งโปรแกรมที่ใช้สำหรับการวิเคราะห์คุณภาพข้อสอบ ถ้าคำนึงถึงโปรแกรมที่มีหลักการทำงานไม่ซับซ้อนเกินไป ผู้เขียนขอแนะนำโปรแกรม Xcaliber ที่ใช้งานได้ง่ายและให้ค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ (IRT)

3.5 การคัดเลือกข้อสอบ คัดเลือกข้อสอบจากผลการวิเคราะห์คุณภาพข้อสอบโดยคัดเลือกข้อสอบที่ตรงตามเงื่อนไขที่กำหนดไว้ ซึ่งผู้เขียนขอแนะนำเกณฑ์การคัดเลือกของข้อสอบของ Urry (1977) ดังนี้ 1) ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง 2.50 2) ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -2.50 ถึง 2.50 และ 3) ค่าโอกาสการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.30

สำหรับความหมายของระดับค่าความยากเฉลี่ยของข้อสอบ มีเกณฑ์ดังนี้

มากกว่า 2.0000	หมายถึง	ข้อสอบยากมาก
1.0001 ถึง 2.0000	หมายถึง	ข้อสอบยาก
0.5001 ถึง 1.0000	หมายถึง	ข้อสอบค่อนข้างยาก
-0.4999 ถึง 0.5000	หมายถึง	ข้อสอบปานกลาง
-0.9999 ถึง -0.5000	หมายถึง	ข้อสอบค่อนข้างง่าย
-2.0000 ถึง -1.0000	หมายถึง	ข้อสอบง่าย
น้อยกว่า -2.0000	หมายถึง	ข้อสอบง่ายมาก

3.6 การจัดทำคลังข้อสอบ O-NET ได้นำข้อคำถาม ตัวเลือก เฉลย ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และ 3) ค่าโอกาสการเดาของข้อสอบ (c) บันทึกลงในคลังข้อสอบที่พัฒนาขึ้นโดยใช้ระบบฐานข้อมูล MySQL ซึ่งเป็นฐานข้อมูลที่ไม่เสียค่าใช้จ่ายใด ๆ ในการใช้งาน

## ตารางที่ 2 สรุปตัวอย่างงานวิจัยที่วิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ

แหล่งข้อมูล	ผู้วิจัย								
	สุนันทา และคณะ	เบญจมาภรณ์ และคณะ	สุชาดา และคณะ	จารุจิตร และคณะ	ณัฐรณม์ และศิริเดช	กนกกร และคณะ	ปรีชา และคณะ	สุรเดช และคณะ	
ปี พ.ศ.	2556	2559	2559	2559	2561	2562	2562	2563	
โปรแกรม	IRT-BAY	IRT-BAY	Xcalibre	Xcalibre	IRTPRO	BILOG-MG	BILOG-MG	IRTPRO	
วิเคราะห์ข้อมูล									
ลักษณะข้อสอบ	ข้อสอบสาระเทคโนโลยี สารสนเทศ และการสื่อสาร	ข้อสอบประเมินความรู้ทางการศึกษา	ข้อสอบ O-NET 8 กลุ่มสาระ การเรียนรู้	ข้อสอบ O-NET 8 กลุ่มสาระ การเรียนรู้	วิชาเคมี เรื่องพันธะ โคเวเลนต์	วิชาคณิตศาสตร์ เรื่อง การประยุกต์ ของสมการเชิงเส้น	วิชาคณิตศาสตร์ เรื่อง ระบบจำนวนจริง	วิชาเคมี เรื่อง พันธะเคมี	
แหล่งที่มา	พัฒนาขึ้นเอง	พัฒนาขึ้นเอง	สทศ.	สทศ.	พัฒนาขึ้นเอง	พัฒนาขึ้นเอง	พัฒนาขึ้นเอง	พัฒนาขึ้นเอง	
ระดับชั้น	ม.3	ป.ตรี	ป.6, ม.3, ม.6	ม.6	ม.4	ม.ต้น	ม.ต้น	ม.4	
จำนวนข้อที่สร้าง	140 ข้อ	519 ข้อ	357, 542 และ 1,335 ข้อ	542 ข้อ	20 ข้อ	75 ข้อ	70 ข้อ	96 ข้อ	
เกณฑ์การคัดเลือกข้อสอบ	2	1	1	1	2	3	3	2	
หมายเหตุ: เกณฑ์การคัดเลือกข้อสอบ	1. a มีค่าตั้งแต่ 0.50-2.50 b มีค่าตั้งแต่-2.50-2.50 c มีค่าไม่เกิน .30								
	2. a มีค่าตั้งแต่ 0.50-2.50 b มีค่าตั้งแต่-3.00-3.00 c มีค่าไม่เกิน .30		3. a มีค่าตั้งแต่ 0.50-2.50 b มีค่าตั้งแต่-3.50-3.50 c มีค่าไม่เกิน .30						

จากการทบทวนงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบในประเทศไทย ตามตารางที่ 2 พบว่า การวิเคราะห์คุณภาพข้อสอบส่วนใหญ่จะนำข้อสอบที่ผ่านเกณฑ์การคัดเลือกข้อสอบแล้วนำไปพัฒนาเป็นคลังข้อสอบเพื่อการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เป็นส่วนใหญ่

## 5. ประโยชน์ของการวิเคราะห์ข้อสอบ

การวิเคราะห์ข้อสอบหลังการสอบเป็นขั้นตอนที่สำคัญของกระบวนการวัดและประเมินผล ซึ่งจะเป็นประโยชน์ต่อตัวผู้สอนและผู้เรียน ประโยชน์ของการวิเคราะห์ข้อสอบ ได้แก่

1. ทำให้ทราบข้อมูลพื้นฐานของตัวข้อสอบ ( เช่น ค่าพารามิเตอร์ ค่าความยาก อำนาจจำแนก และโอกาสการเดา) และคำตอบ รวมถึงข้อสอบแต่ละข้อได้ทำหน้าที่วัดผลสัมฤทธิ์อย่างมีคุณภาพเพียงใด
2. ชี้ให้เห็นถึงจุดแข็งและจุดอ่อนที่ผู้สอนหรือผู้เรียนต้องปรับปรุงแก้ไขเพื่อพัฒนาความสามารถและทักษะของผู้เรียนให้เป็นไปตามคาดหวัง
3. เป็นพื้นฐานสำคัญในการสะท้อนถึงประสิทธิภาพของการจัดการเรียนการสอน ครอบคลุมเนื้อหาที่สำคัญตามวัตถุประสงค์หรือไม่ ผู้เรียนเข้าใจประเด็นใดยังไม่ถูกต้องหรือคลาดเคลื่อน
4. ใช้เป็นแนวทางในการพัฒนาและปรับปรุงการสร้างข้อสอบ ตลอดจนการเขียนข้อสอบในลักษณะใดจึงจะได้ข้อสรุปที่ดีในอนาคต
5. ทำให้สามารถคัดเลือกข้อสอบที่มีคุณภาพ เพื่อเก็บข้อสอบไว้เป็นฐานข้อมูลคลังข้อสอบที่สามารถนำไปพัฒนาเป็นคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในอนาคต

## บทสรุป

การวิเคราะห์ข้อสอบ ทั้งในเชิงปริมาณและคุณภาพเป็นกระบวนการสำคัญและควรกำหนดเป็นขั้นตอนหนึ่งอย่างมีระบบของกระบวนการวัดผลและประเมินผลที่มีคุณภาพ เพราะเป็นกระบวนการที่จะได้ข้อสอบที่ดี เมื่อทำการวิเคราะห์ข้อสอบหลังจากนำข้อสอบไปใช้แล้ว ควรนำข้อมูลที่ได้มาปรับปรุงข้อสอบข้อนั้น ให้ดียิ่งขึ้นเพื่อใช้ในโอกาสต่อไป เมื่อได้ข้อสอบที่ดีมีคุณภาพก็ส่งผลให้การวัดผลน่าเชื่อถือ ทำให้การตัดสินใจผลการสอบมีความยุติธรรมมากขึ้น และยังสามารถคัดเลือกข้อสอบที่ตรงตามเกณฑ์ที่กำหนดไว้เพื่อนำไปจัดทำคลังข้อสอบและพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ต่อไป

## เอกสารอ้างอิง

- กนกกร พวงสมบัติ. 2562. การพัฒนาคลังข้อสอบ สำหรับโปรแกรมทดสอบ และวินิจฉัยทักษะการแก้โจทย์ปัญหา. วารสารบัณฑิตศึกษา มหาวิทยาลัยราชภัฏสกลนคร. 16(74): 27-39.
- จารุจิตร สิทธิประยูร ปิยะทิพย์ ดินวร และโสฬส สุขานนท์สวัสดิ์. 2559. การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการจัดสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ 3. วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม. 22(1): 47-62.
- โชติกา ภาชีผล. 2555. คู่มือการออกข้อสอบและการวิเคราะห์ข้อสอบ สำหรับการสอบขึ้นทะเบียนเป็นผู้ประกอบวิชาชีพกายภาพบำบัด. สภากายภาพบำบัด. กรุงเทพมหานคร.
- ณัฐภรณ์ เลขะวัฒนพงษ์ และศิริเดช สุชีวะ. 2561. การเปรียบเทียบคุณภาพของวิธีการให้คะแนนความรู้บางส่วนระหว่างวิธีประยุกต์ของคู่มือ วิธีตัดตัวลง-เลือก ตัวถูก และ วิธีประยุกต์ตัดตัว ลง-เลือก ตัวถูก. วารสารอิเล็กทรอนิกส์ทางการศึกษา. 13(1): 1-17.
- บุญชม ศรีสะอาด มนตรี อนันตรักษ์ และ นิภา ศรีโพธิ์โรจน์. 2524. การวัดและการประเมินผลทางการศึกษา. จินตทัศน์. ภาพลื่นจู้.
- บุญเชิด ภัฏญอนันตพงษ์. 2526. การสอนแบบอิงเกณฑ์: แนวคิดและวิธีการ. กรุงเทพฯ: ภาควิชาพื้นฐานการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ. กรุงเทพมหานคร.
- เบญจมาภรณ์ เสนารัตน์ บุญชม ศรีสะอาดและ จริยา ภัคตราจันทร์. 2559. การพัฒนาโปรแกรมการประเมินความรู้ทางการวิจัย การศึกษาแบบปรับเหมาะด้วยคอมพิวเตอร์ออนไลน์ สำหรับนักศึกษาคณะศึกษาศาสตร์มหาวิทยาลัยราชภัฏในภาคตะวันออกเฉียงเหนือ. วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม. 22(1): 164-182.
- ปรีชา บุญจิตร เอกราช ดินาง และ พงษ์ศักดิ์ ศรีจันทร์. 2562. การพัฒนาคลังข้อสอบวินิจฉัย วิชาคณิตศาสตร์ เรื่อง ระบบจำนวนจริงของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 20. วารสารครูศึกษา มหาวิทยาลัยราชภัฏอุดรธานี. 1(1): 23-34.
- รณิดา เขยขุ่ม. 2551. การเปรียบเทียบคุณภาพข้อสอบและแบบสอบหลายตัวเลือกที่มีรูปแบบตัวเลือกต่างกัน วิทยานิพนธ์ดุขุฎิบัณฑิต. สาขาการวัดและประเมินผลทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพมหานคร.
- ศิริชัย กาญจนวาสิ. 2555. ทฤษฎีการทดสอบแนวใหม่. พิมพ์ครั้งที่ 4. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพมหานคร.
- ศิริชัย กาญจนวาสิ และ คณิต ไชมงคล. 2535การพัฒนโปรแกรมคอมพิวเตอร์สำหรับกรมคอมพิวเตอร์ สำหรับวิเคราะห์ข้อสอบและประมาณค่าความสามารถของผู้สอบตามทฤษฎีตอบสนองข้อสอบด้วยวิธีของเบย์. จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพมหานคร.
- สุชาติา กรเพชรปานี ปิยะทิพย์ ดินวร และโสฬส สุขานนท์สวัสดิ์. 2559. การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการจัดสอบ O-NET. วิทยาการวิจัยและวิทยาการปัญญา. 14(1): 14-31.
- สุนันทา ศิริเบญจา ไชยรัตน์ ปราณีและดวงใจ สีเขียว. 2556. การพัฒนาแบบทดสอบปรับเหมาะโดยใช้คอมพิวเตอร์ สาระเทคโนโลยีสารสนเทศ และการสื่อสารโดยประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบย์ สำหรับนักเรียนชั้นมัธยมศึกษา ปีที่ 3. วารสารวิชาการและวิจัยสังคมศาสตร์. 8(22): 87-102.
- สุรเดช อนันตสวัสดิ์ กมลวรรณ ตั้งธนกานนท์ และโชติกา ภาชีผล. 2563. การพัฒนาแบบสอบวินิจฉัยสามระดับ เรื่อง พันธะเคมี สำหรับนักเรียนชั้นมัธยมศึกษา ปีที่ 4. วารสารครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. 48(2): 280-300.
- สิริพร ทิพย์คง. 2545. หลักสูตรและการสอนคณิตศาสตร์. สถาบันพัฒนาคุณภาพวิชาการ (พว). กรุงเทพมหานคร.
- Assessment Systems Corporation. 2018. Xcaliber: Bringing you the Benefits of Item Response Theory (IRT). [Online]. Available: <http://assess.com/Xcaliber>. (Retrieved 18, November 2020).
- Cai, L., Thissen, D., and S. Du Toit. 2018. IRTPRO 2.1 for Windows. [Online]. Available: [http://www.ssicentral.com/irt/IRTPRO\\_by\\_SSI.pdf](http://www.ssicentral.com/irt/IRTPRO_by_SSI.pdf). (Retrieved 20 July, 2020).
- De Ayala, R.J. 2009. The Theory and Practice of Item Response Theory. The Guilford Press. New York.
- Du Toit, M. (ed.). 2003. IRT from SSI: Bilog-MG, Multilog, Parscale, Testfact. Scientific Software International. Lincolnwood, Illinois, United States.
- Hambleton, R.K., Swaminathan, H. and H.J. Rogers 1991. Fundamentals of Item Response Theory. Sage Publications. New Delhi, India.
- Mullis, I. and M.O. Martin. 2013. Item Writing Guidelines. [Online]. Available: [https://timssandpirls.bc.edu/publications/timss/T15\\_item\\_writing\\_guide\\_lines.pdf](https://timssandpirls.bc.edu/publications/timss/T15_item_writing_guide_lines.pdf). (Retrieved 18, November 2020).
- Stone, W. 2001. Measuring Social Capital: towards a theoretically informed measurement framework for researching social capital in family and community life., Research Paper No. 24, Australian Institute of Family Studies. Melbourne, Victoria, Australia.
- Urry, V.W. 1977. Tailored testing: A successful application of latent trait theory. Journal of Educational Measurement. 14(2): 181-196.
- Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., and R.J. Mislevy. 2000. Computerized Adaptive Testing: A Primer (2<sup>nd</sup> ed.). Lawrence Erlbaum Associates. Mahwah, New Jersey, United States.