

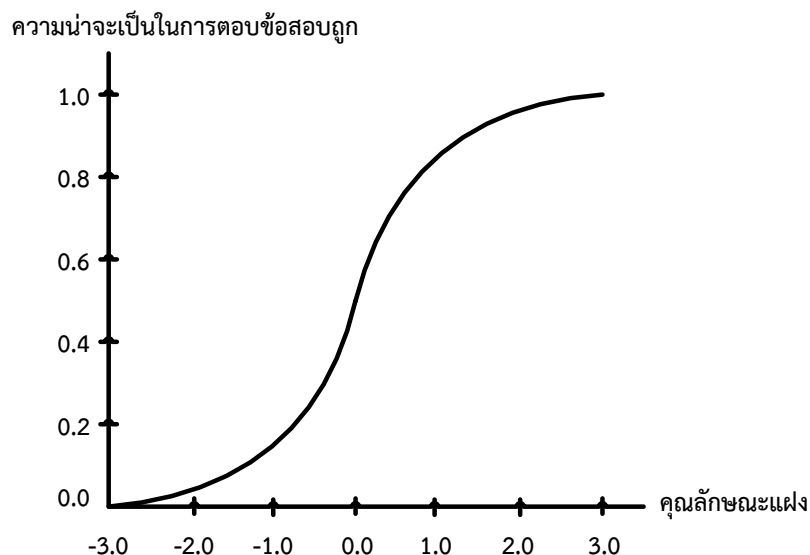
โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า สำหรับตัวแปรด้านเจตพิสัย

POLYTOMOUS ITEM RESPONSE THEORY MODELS FOR AFFECTIVE VARIABLES

อาจารย์ ดร.สุชาติ สกลกิจรุ่งโรจน์

ตัวแปรด้านเจตพิสัยเป็นตัวแปรที่เกี่ยวข้องกับค่านิยม ความรู้สึก ความซาบซึ้ง เจตคติ ความเชื่อ ความสนใจ และคุณธรรม ซึ่งเป็นตัวแปรที่แสดงถึงคุณลักษณะภายในของบุคคลและไม่สามารถสังเกตได้โดยตรง แต่มีลักษณะเป็นตัวแปรแฝง (Latent Variables) ที่มีอิทธิพลอยู่เบื้องหลังพฤติกรรมที่แสดงออก การวัดตัวแปรด้านเจตพิสัยจึงต้องวัดผ่านตัวแปรสังเกตได้ (Observable Variables) หรือพฤติกรรมที่บุคคลแสดงออกผ่านสถานการณ์หรือสิ่งเร้าที่กำหนดให้ เช่น งาน (Tasks) ข้อคำถาม (Items) ข้อสอบ (Test items) โดยนำทฤษฎีการวัด (Measurement Theory) มาเป็นแนวทางในการแสดงความสัมพันธ์ระหว่างพฤติกรรมกับตัวแปรที่ต้องการวัดอย่างเป็นเหตุเป็นผล (Rationale) ซึ่งโดยทั่วไปทฤษฎีการวัดจะแสดงความสัมพันธ์ระหว่างพฤติกรรมกับตัวแปรที่ต้องการวัดด้วยโมเดลเชิงพฤติกรรม (Model of Behavior) ทฤษฎีการวัดที่เป็นที่รู้จักโดยทั่วไป ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการ (Generalizability Theory) และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)

ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นทฤษฎีการทดสอบแนวใหม่ซึ่งได้รับความนิยมและมีการนำไปใช้เพื่อพัฒนาคุณภาพของแบบวัดอย่างแพร่หลาย ทฤษฎีนี้มุ่งศึกษาคุณลักษณะแฝง (คุณลักษณะต่าง ๆ ที่ซ่อนอยู่ในตัวบุคคล เช่น ความสามารถทางคณิตศาสตร์ ความสามารถทางภาษา ความสุข ความเศร้า เจตคติ เป็นต้น) ด้วยโมเดลคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างความน่าจะเป็น (Probability) ในการเลือกคำตอบถูกต้องกับคุณลักษณะแฝงที่มีอิทธิพลอยู่เบื้องหลังการเลือกคำตอบถูกต้องนั้น ทฤษฎีการตอบสนองข้อสอบ มีแนวคิดที่สำคัญสองประการ คือ 1) พฤติกรรมการตอบข้อสอบของผู้สอบ (Examinee) สามารถทำนายหรืออธิบายได้ด้วยชุดขององค์ประกอบ ที่เรียกว่า คุณลักษณะ (Traits) คุณลักษณะแฝง (Latent Traits) หรือความสามารถ (Abilities) และ 2) ความสัมพันธ์ระหว่างพฤติกรรมการตอบข้อสอบของผู้สอบกับชุดของคุณลักษณะแฝงซึ่งมีอิทธิพลอยู่เบื้องหลังพฤติกรรมการตอบข้อสอบนั้น สามารถอธิบายได้ด้วยฟังก์ชันคณิตศาสตร์ ที่เรียกว่า ฟังก์ชันลักษณะข้อสอบ (Item Characteristic Function) หรือ โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีลักษณะเป็นเส้นโค้งรูปตัวเอส (S Curve) ดังภาพที่ 1 (Hambleton, Swaminathan, & Rogers, 1991; Crocker & Algina, 2008)



ภาพที่ 1 โค้งลักษณะข้อสอบ (ปรับจาก Embetson & Reise, 2000, p.46)

จากภาพที่ 1 จะเห็นได้ว่าแกนนอนแสดงถึงระดับคุณลักษณะแฝงของบุคคล ซึ่งมีค่าที่เป็นไปได้ตั้งแต่ $-\infty$ ถึง $+\infty$ แต่โดยปกติแล้วผลการวิเคราะห์ส่วนใหญ่จะให้ค่าตั้งแต่ -3 ถึง +3 ขณะที่แกนตั้งแสดงถึงค่าความน่าจะเป็นในการตอบข้อสอบถูกซึ่งมีค่าตั้งแต่ 0.00 ถึง 1.00 จากกราฟจะเห็นได้ว่าเมื่อระดับของคุณลักษณะแฝงเพิ่มขึ้น ความน่าจะเป็นในการตอบข้อสอบถูก (Correct Response) จะเพิ่มขึ้นด้วย

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่มีลักษณะเป็นกลุ่มโมเดลทางคณิตศาสตร์ แต่ละโมเดลจะมีรายละเอียดที่ต่างกันไป การเลือกใช้โมเดลการตอบสนองข้อสอบจึงต้องพิจารณาความเหมาะสมของโมเดลกับลักษณะอื่นๆ ประกอบด้วย เช่น ลักษณะของข้อมูล มิติของตัวแปร จำนวนผู้สอบ จำนวนข้อสอบ/ข้อคำถาม เป็นต้น หากจำแนกโมเดลการตอบสนองข้อสอบตามลักษณะของข้อมูลสามารถจำแนกได้ 2 ประเภท (ศิริชัย กาญจนวาสี, 2555; De Ayala, 2009; Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991) ได้แก่ 1) โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนสองค่า (Binary IRT Model หรือ Dichotomous IRT Model) สำหรับข้อมูลที่เมื่อตรวจให้คะแนน (Score) แล้วมีคะแนนที่เป็นไปได้สองค่า (Binary Data) เป็น 1 กับ 0 เช่น ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน ซึ่งอาจได้มาจากข้อสอบ/ข้อคำถามหลากหลายรูปแบบ เช่น แบบทดสอบแบบหลายตัวเลือกแต่มีคำตอบที่ถูกเพียงตัวเลือกเดียว แบบทดสอบที่ให้เลือกตอบแบบถูกผิด แบบประเมินบุคลิกภาพที่มีตัวเลือกเป็นจริงหรือไม่เป็นจริง แบบทดสอบเจตคติที่มีตัวเลือกเห็นด้วยหรือไม่เห็นด้วย มาตราประมาณค่าทางด้านพฤติกรรมที่มีตัวเลือกใช่หรือไม่ใช่ และ 2) โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า (Polytomous IRT Model) ในบทความนี้จะนำเสนอเฉพาะโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า ซึ่งมีการจำแนกเนื้อหาเป็น 3 ส่วน โดยส่วนแรกจะกล่าวถึงโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่าที่นิยมใช้ ส่วนที่สองกล่าวถึงโปรแกรมสำเร็จรูปที่ใช้ในการวิเคราะห์ และส่วนที่สามกล่าวถึง

ตัวอย่างการนำโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่าไปใช้ในการวิเคราะห์ตัวแปรด้านเจตพิสัย

1. โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า

โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า (Polytomous IRT Models) เป็นโมเดลการตอบสนองข้อสอบที่เหมาะสมกับแบบวัดตัวแปรด้านเจตพิสัยซึ่งข้อคำถามในแบบวัดมักมีลักษณะเป็นรายการคำตอบแบบเรียงลำดับหลายตัวเลือก (Ordered Categories Responses) และมีการตรวจให้คะแนนมากกว่าสองค่า ตัวอย่างเช่น แบบประเมินความสุขซึ่งมีลักษณะเป็นมาตราประมาณค่า 5 ระดับ มีตัวเลือกและการตรวจให้คะแนน คือ หากเลือก “น้อยที่สุด” ให้ 1 คะแนน “น้อย” ให้ 2 คะแนน “ปานกลาง” ให้ 3 คะแนน “มาก” ให้ 4 คะแนน “มากที่สุด” ให้ 5 คะแนน เป็นต้น โมเดลการตอบสนองข้อสอบในกลุ่มนี้มีหลากหลายโมเดล และมีผู้พัฒนาโมเดลใหม่ ๆ ขึ้นมาเป็นระยะ แต่โมเดลที่คุ้นเคยและรู้จักกันโดยทั่วไปมี 6 โมเดลหลัก ได้แก่ Graded Response Model (GRM), Modified Graded Response Model (M-GRM), Partial Credit Model (PCM), Generalized Partial Credit Model (G-PCM), Rating Scale Model (RSM), Nominal Response Model (NRM; Embretson & Reise, 2000)

1.1 Graded Response Model (GRM)

โมเดลนี้พัฒนาโดย Samejima ในปี ค.ศ. 1969 เป็นโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า ที่เหมาะกับข้อมูลที่มีตัวเลือกรายการคำตอบแบบเรียงลำดับ (Ordered Categories Responses) เช่น มาตราประมาณค่าของลิเคิร์ต (Likert Rating Scale) โมเดล GRM เป็นโมเดลที่ขยายมาจากโมเดลโลจิสแบบสองพารามิเตอร์ (Two-Parameter Logistic Model: 2PL) ซึ่งเป็นโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนสองค่าที่มีแนวคิดที่ข้อสอบแต่ละข้อมีพารามิเตอร์ที่แตกต่างกัน 2 พารามิเตอร์¹ ซึ่งส่งผลต่อโอกาสหรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบ คุณสมบัติดังกล่าว คือ ความยากและอำนาจจำแนก ในการวิเคราะห์ข้อสอบจึงมีการนำค่าพารามิเตอร์ความยาก (Difficulty Parameter) และค่าพารามิเตอร์อำนาจจำแนก (Discrimination Parameter) มาใช้ในการคำนวณความน่าจะเป็นในการตอบข้อสอบถูกต้อง เนื่องจากเชื่อว่าโอกาสที่ผู้สอบจะตอบข้อสอบข้อใดข้อหนึ่งถูกต้องหรือไม่ นั้น มิได้ขึ้นอยู่กับระดับความสามารถหรือคุณลักษณะแฝงของผู้สอบเพียงประการเดียวแต่ยังเกี่ยวข้องกับพารามิเตอร์ของข้อสอบแต่ละข้อด้วย

โมเดล GRM ซึ่งเป็นโมเดลที่มีการขยายมาจากโมเดล 2PL จึงเป็นโมเดลที่มีพารามิเตอร์ที่เกี่ยวข้อง 2 พารามิเตอร์ โดยมีการเรียกชื่อพารามิเตอร์แตกต่างไปจากโมเดล 2PL คือ พารามิเตอร์ความชัน หรือ α_i (เทียบเท่ากับพารามิเตอร์อำนาจจำแนก) และพารามิเตอร์เทรชโฮลด์ หรือ β_{ij} (เทียบเท่ากับพารามิเตอร์ความยาก) การวิเคราะห์โมเดล GRM นั้น De Ayala (2009) ได้เสนอว่าขนาดกลุ่มตัวอย่างที่ใช้ในการประมาณค่าพารามิเตอร์ในโมเดล GRM ควรมีไม่ต่ำกว่า 500 คน เพื่อให้สามารถวิเคราะห์ความสอดคล้องของโมเดล

¹ ทฤษฎีการตอบสนองข้อสอบจะมีการประมาณค่าพารามิเตอร์ของข้อสอบซึ่งเป็นค่าที่แสดงถึงคุณสมบัติทางกรวัด หากเทียบกับทฤษฎีการทดสอบแบบดั้งเดิมจะคล้ายกับค่าความยากและอำนาจจำแนกของข้อสอบ ที่บ่งบอกคุณสมบัติของข้อสอบว่ามีความยากง่ายระดับใด สามารถจำแนกกลุ่มผู้สอบได้มากน้อยเพียงใด อย่างไรก็ดี แนวคิดและวิธีการคำนวณค่าพารามิเตอร์ของทฤษฎีทั้งสองนี้มีความแตกต่างกันอย่างสิ้นเชิง

ประเมินความเป็นเอกมิติของคุณลักษณะแฝงที่ต้องการวัดได้² ทำให้สามารถประมาณค่าพารามิเตอร์ความชันได้ไม่ยากเกินไป และเพื่อให้มั่นใจได้ว่ามีข้อมูลการตอบในทุกตัวเลือกการรายการคำตอบ อย่างไรก็ตาม ทัศนคติที่ไม่ใช่กฎตายตัว ในทางปฏิบัติการกำหนดขนาดกลุ่มตัวอย่างที่เหมาะสมควรพิจารณาจากปัจจัยหลากหลายด้าน เช่น วัตถุประสงค์ในการนำค่าพารามิเตอร์ที่ประมาณค่าได้ไปใช้ วิธีการประมาณค่าที่เลือกใช้ ความยาวของมาตราวัด ลักษณะการตอบข้อคำถามของผู้สอบ จำนวนข้อมูลที่ขาดหาย เป็นต้น

1.2 Modified Graded Response Model (M-GRM)

โมเดลนี้พัฒนาโดย Muraki (1990) เป็นโมเดลที่ปรับปรุงมาจากโมเดล GRM เหมาะสำหรับวิเคราะห์แบบสอบถามที่มีลักษณะเป็นมาตราประมาณค่า (Rating Scale) เช่น แบบวัดเจตคติ ข้อคำถามในมาตราวัดจะต้องมีจำนวนตัวเลือกการรายการคำตอบเท่ากัน แต่สามารถมีค่าพารามิเตอร์ความชันแตกต่างกันได้ โมเดล M-GRM ใช้วิธีการคำนวณความน่าจะเป็นในการเลือกการรายการคำตอบแบบสองขั้นตอนเหมือนในโมเดล GRM อย่างไรก็ตาม พารามิเตอร์เทรชโฮลด์ (β_{ij}) ในโมเดล M-GRM ถูกแบ่งออกเป็นสองส่วน ได้แก่ 1) พารามิเตอร์ตำแหน่ง (Location Parameter: b_i) ของข้อคำถามแต่ละข้อ และ 2) ชูตของพารามิเตอร์เทรชโฮลด์ (Threshold Parameter: c_j) ของมาตราวัดทั้งฉบับ ความแตกต่างระหว่างโมเดล GRM กับโมเดล M-GRM คือ โมเดล GRM จะมีการประมาณค่าชูตของพารามิเตอร์เทรชโฮลด์ (β_{ij}) สำหรับข้อคำถามแต่ละข้อ แต่โมเดล M-GRM มีชูตของพารามิเตอร์เทรชโฮลด์ (c_j) เพียงชุดเดียวที่ใช้กับมาตราวัดทั้งฉบับ แต่มีการประมาณค่าพารามิเตอร์ตำแหน่ง (b_i) หนึ่งค่าสำหรับข้อคำถามแต่ละข้อ โมเดล M-GRM ได้ชื่อว่าเป็นโมเดล GRM แบบจำกัด (Restricted Model) เพราะว่ามีจำนวนตัวเลือกการรายการคำตอบ (Category Boundaries) ของข้อคำถามแต่ละข้อเท่ากันตลอดทั้งมาตราวัด ในขณะที่โมเดล GRM นั้น ข้อคำถามแต่ละข้ออาจมีจำนวนตัวเลือกการรายการคำตอบของข้อคำถามแต่ละข้อแตกต่างกันได้ ดังนั้น โมเดล M-GRM จึงมีการประมาณค่าพารามิเตอร์น้อยกว่าโมเดล GRM จุดเด่นของโมเดล M-GRM เมื่อเปรียบเทียบกับโมเดล GRM คือ มีพารามิเตอร์ตำแหน่งแยกออกมาจากชูตของพารามิเตอร์เทรชโฮลด์ อย่างไรก็ตาม การเลือกใช้ขึ้นอยู่กับรูปแบบของข้อคำถามด้วย ถ้ามาตราวัดมีรูปแบบข้อคำถามหลากหลายแบบอยู่ในมาตราวัดฉบับเดียวกัน โมเดล GRM จะมีความเหมาะสมมากกว่าโมเดล M-GRM

1.3 Partial Credit Model (PCM)

โมเดล PCM ซึ่งพัฒนาโดย Master (1982) เป็นโมเดลที่ถูกออกแบบมาสำหรับวิเคราะห์ข้อสอบที่มีกระบวนการตอบหลายลำดับขั้น โดยมีการตรวจให้คะแนนบางส่วน กล่าวคือ มีการให้คะแนนสำหรับแต่ละขั้นตอนที่ผู้สอบทำได้ ซึ่งพบได้บ่อยในแบบวัดผลสัมฤทธิ์ทางการเรียน เช่น โจทย์ปัญหาทางคณิตศาสตร์ เป็นต้น นอกจากนี้ โมเดล PCM ยังสามารถนำมาใช้ในการวิเคราะห์มาตราวัดเจตคติ หรือมาตราวัดบุคลิกภาพที่มีการประเมินความเชื่อ หรือการตอบสนองต่อข้อความโดยมีการให้คะแนนแบบหลายค่าด้วย โมเดล PCM จะแตกต่างจากโมเดล GRM และ M-GRM เพราะเป็นโมเดลแบบทางตรง (Direct Model) คือ คำนวณค่าความ

² โมเดลการตอบสนองข้อสอบเป็นโมเดลทางคณิตศาสตร์ที่มีข้อตกลงเบื้องต้นซึ่งต้องตรวจสอบก่อนการวิเคราะห์เช่นเดียวกับการใช้สถิติอื่นๆ โดยทั่วไปจะมีข้อตกลงเบื้องต้น 2 ประการ คือ 1) ความเป็นเอกมิติ (Unidimensionality) และ 2) ความเป็นอิสระ (Local Independence) อย่างไรก็ตาม หากโมเดลการตอบสนองข้อสอบที่ใช้เป็นโมเดลแบบพหุมิติ ข้อตกลงเบื้องต้นจะมีรายละเอียดแตกต่างกันไป

น่าจะเป็นในการเลือกรายการคำตอบของแต่ละตัวเลือกได้โดยตรง โมเดลนี้เป็นโมเดลที่มีการขยายต่อมาจากโมเดลโลจิสแบบหนึ่งพารามิเตอร์ (1PL) จึงมีคุณลักษณะมาตรฐานต่าง ๆ เช่นเดียวกับ Rasch Model เช่นการแยกจากกันของค่าพารามิเตอร์ข้อสอบกับค่าพารามิเตอร์บุคคล³ ในโมเดล PCM นั้นข้อคำถามแต่ละข้อในมาตรวัดจะมีความชันเท่ากัน ดังนั้น พารามิเตอร์ความชันจึงไม่ปรากฏอยู่ในโมเดลนี้ คงเหลือเพียงพารามิเตอร์ความยากแบบลำดับขั้น (Step Difficulty: δ_{ij}) ที่ใช้อธิบายระดับความยากของการได้คะแนนจากระดับหนึ่งไปยังคะแนนอีกระดับหนึ่ง หรือจากตัวเลือกรายการคำตอบหนึ่งสู่อีกตัวเลือกหนึ่ง นอกจากนี้ ค่านี้ยังบ่งบอกถึงระดับคุณลักษณะแฝงบริเวณที่ไค้งการเลือกรายการคำตอบ 2 รายการตัดกัน บางครั้งจึงนิยมเรียกว่าพารามิเตอร์จุดตัดรายการคำตอบ (Category Intersections Parameter) โดยข้อคำถามที่มีตัวเลือกรายการคำตอบ 5 รายการ จะมีพารามิเตอร์ความยากแบบลำดับขั้น 4 ค่า จุดเด่นของโมเดลนี้ คือ สามารถใช้คะแนนดิบรวมของมาตรวัดเพื่อประมาณค่าระดับคุณลักษณะแฝงได้ ดังนั้น บุคคลที่มีคะแนนดิบรวมของมาตรวัดเท่ากันจึงมีระดับคุณลักษณะแฝงเท่ากัน อย่างไรก็ตาม โมเดลนี้นำไปใช้ในทางปฏิบัติได้ยาก เพราะในการประมาณค่าพารามิเตอร์อาจพบปัญหาว่าบริเวณจุดตัดไค้งการเลือกรายการคำตอบ ไม่ได้เรียงลำดับจากง่ายไปยาก หรือจากระดับคุณลักษณะแฝงต่ำไปสูง ทำให้เกิดปรากฏการณ์ ที่เรียกว่า การผกผัน (Reversal) ซึ่งแสดงให้เห็นว่ามีตัวเลือกรายการคำตอบอย่างน้อยหนึ่งรายการที่เป็นตัวเลือกซึ่งไม่สอดคล้องกับคุณลักษณะแฝงที่ต้องการวัด หรือข้อมูลที่เกิดขึ้นรวบรวมได้ไม่สอดคล้องกับโมเดล

1.4 Generalized Partial Credit Model (G-PCM)

โมเดล G-PCM พัฒนาโดย Muraki (1992) เป็นโมเดลที่ปรับปรุงเพิ่มเติมจากโมเดล PCM ซึ่งกำหนดให้ข้อคำถามแต่ละข้อในมาตรวัดมีพารามิเตอร์ความชัน (α_i) แตกต่างกันได้ โดยพารามิเตอร์ความชันในโมเดล G-PCM เป็นค่าที่บ่งบอกถึงระดับการตอบสนองต่อตัวเลือกรายการคำตอบที่แตกต่างกันไปในแต่ละข้อคำถาม ซึ่งขึ้นอยู่กับระดับคุณลักษณะแฝง หรือระดับความสามารถ (θ) ที่เปลี่ยนแปลงไป ในโมเดล G-PCM นี้ หากข้อคำถามใดมีค่าพารามิเตอร์ความชันน้อยกว่า 1.0 ไค้งการเลือกรายการคำตอบ (Category Response Curves) ที่ได้จะมีลักษณะแบนราบคล้ายในโมเดล PCM แต่หากค่าพารามิเตอร์ความชันมีค่ามากกว่า 1.0 ไค้งการเลือกรายการคำตอบในโมเดล G-PCM จะมีลักษณะโด่งมากกว่าไค้งการเลือกรายการคำตอบที่พบในโมเดล PCM อย่างไรก็ตาม พารามิเตอร์จุดตัดรายการคำตอบ (δ_{ij}) ในโมเดล G-PCM มีความหมายเช่นเดียวกับโมเดล PCM

1.5 Rating Scale Model (RSM)

โมเดล RSM เป็นกลุ่มของโมเดลที่มีหลากหลายรูปแบบและอาจก่อให้เกิดความสับสนได้ เพราะมีลักษณะคล้ายคลึงกับโมเดลอื่น ๆ เช่น โมเดล PCM อย่างไรก็ตาม โมเดล RSM ตามแนวคิดของ Andrich (1978) มีความแตกต่างที่สำคัญ คือ ข้อคำถามในมาตรวัดซึ่งมีรูปแบบการตอบแบบเดียวกันนั้น ข้อคำถามแต่ละข้อจะมีพารามิเตอร์ตำแหน่ง (Location Parameter: λ_i ⁴) เพียงค่าเดียว ซึ่งแสดงถึงความยากหรือง่ายของข้อ

³ พารามิเตอร์บุคคล (Person Parameter) เป็นพารามิเตอร์ที่แสดงถึงระดับความสามารถของบุคคล หรือระดับคุณลักษณะแฝง

⁴ สัญลักษณ์ (ตัวอักษรกรีก) ที่ใช้ในการแทนความหมายพารามิเตอร์ของแต่ละโมเดลมีความจำเพาะและแตกต่างกันไปในแต่ละโมเดล ดังนั้น แม้พารามิเตอร์จะมีชื่อเหมือนกันแต่อาจใช้สัญลักษณ์ที่ต่างกันก็ได้

คำถามนั้น ๆ ขณะที่พารามิเตอร์จุดตัดรายการคำตอบ (Category Intersections Parameter: δ_j) ใช้อธิบายช่วงเทรชโฮลด์ของแต่ละรายการคำตอบ (Category Threshold) ซึ่งมี J ช่วง ($J = K - 1$ โดย K คือ จำนวนตัวเลือกรายการคำตอบ) โมเดล PCM กับโมเดล RSM แตกต่างกันที่โมเดล PCM ไม่มีแนวคิดเกี่ยวกับระดับความยากง่ายของแต่ละรายการคำตอบ (แต่ละขั้นตอน หรือแต่ละจุดตัดรายการคำตอบ) ในข้อคำถามนั้น ๆ ดังนั้น ข้อคำถามแต่ละข้อจึงมีระดับความยากของแต่ละขั้นตอน (Difficulties of the Steps) แตกต่างกันได้ ในขณะที่โมเดล RSM จะมีชุดของพารามิเตอร์จุดตัดรายการคำตอบเพียงชุดเดียวที่ใช้ตลอดทั้งมาตรวัด ตัวอย่างเช่น ข้อคำถามทางคณิตศาสตร์ที่ผู้สอบต้องแก้โจทย์ปัญหาหลายขั้นตอน ขั้นตอนต่าง ๆ ภายในข้อคำถาม หรือระหว่างข้อคำถามอาจมีความยากง่ายต่างกัน得 บางขั้นตอนอาจง่าย ในขณะที่บางขั้นตอนอาจยาก อย่างไรก็ตามมาตรวัดเจตคติซึ่งส่วนใหญ่ทุกข้อมีการตอบข้อคำถามรูปแบบเดียวกัน เช่น 1 หมายถึง ไม่เห็นด้วย 2 หมายถึง เห็นด้วยปานกลาง และ 3 หมายถึง เห็นด้วยอย่างยิ่ง จะสังเกตได้ว่า ในมาตรวัดนั้น ๆ ข้อคำถามแต่ละข้อมีระดับความยากง่ายหรือช่วงของแต่ละรายการคำตอบไม่แตกต่างกัน ความยากง่ายในแต่ละขั้นตอน (Step Difficulties หรือ Intersections) ซึ่งพบในโมเดล PCM สามารถแบ่งออกเป็นสองส่วนในโมเดล RSM ได้แก่ พารามิเตอร์ตำแหน่งคุณลักษณะแฝงที่ต้องการวัดของข้อคำถาม (Location Parameter: λ_i) และพารามิเตอร์จุดตัดรายการคำตอบ (Category Intersection Parameter: δ_j) โมเดล RSM เหมาะกับข้อคำถามที่มีการกำหนดช่วงรายการคำตอบแบบคงที่ซึ่งใช้กับข้อคำถามทุกข้อเหมือนกันตลอดมาตรวัด แต่ไม่เหมาะกับมาตรวัดที่มีลักษณะการตอบข้อคำถามที่หลากหลาย

1.6 Nominal Response Model (NRM)

โมเดล NRM พัฒนาโดย Bock (1972) เป็นโมเดลสำหรับใช้วิเคราะห์ข้อสอบหรือข้อคำถามที่รายการคำตอบไม่ได้ถูกจัดเรียงลำดับ ได้แก่ ข้อสอบแบบหลายตัวเลือก (Multiple Choices) ข้อคำถามที่ใช้วัดเจตคติ ข้อคำถามสำหรับประเมินบุคลิกลักษณะ ความตั้งใจเริ่มแรกในการพัฒนาโมเดลนี้เกิดขึ้นเพื่อให้สามารถศึกษาคุณลักษณะของตัวलगซึ่งเป็นตัวเลือกหนึ่งในข้อสอบแบบหลายตัวเลือกด้วยเส้นร่องรอย (Trace Line) การคำนวณความน่าจะเป็นในการเลือกรายการคำตอบคำนวณได้จากพารามิเตอร์ที่เกี่ยวข้อง 2 พารามิเตอร์ ได้แก่ พารามิเตอร์ความชัน (Slope Parameter: α_{ix}) ซึ่งแสดงถึงความชันของเส้นร่องรอยหรือค่าอำนาจจำแนก และพารามิเตอร์จุดตัดรายการคำตอบ (Intercept Parameter: c_{ix})

ลักษณะของโมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่าสองค่าทั้ง 6 โมเดลนี้ สามารถสรุปได้ ดังตารางที่ 1

ตารางที่ 1 ตารางสรุปลักษณะของโมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่าสองค่า

โมเดล	ลักษณะรายการคำตอบ	จำนวนตัวเลือกในแต่ละ ข้อคำถามของแบบวัด		โมเดลพื้นฐาน	
		ไม่ต้องเท่ากัน	ต้องเท่ากัน	1PLM	2PLM
GRM	ตัวเลือกแบบเรียงลำดับ	✓			✓
M-GRM	ตัวเลือกแบบเรียงลำดับ		✓		✓
PCM	ตัวเลือกแบบเรียงลำดับ	✓		✓	
GPCM	ตัวเลือกแบบเรียงลำดับ	✓			✓
RSM	ตัวเลือกแบบเรียงลำดับ		✓	✓	
NRM	ตัวเลือกแบบจำแนกกลุ่ม	✓			✓

การวิเคราะห์คุณสมบัติทางการวัดตามโมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่าสองค่า เป็นการวิเคราะห์ที่มีความซับซ้อนจึงไม่สามารถใช้การคำนวณอย่างง่ายได้ โดยทั่วไปการวิเคราะห์จึงต้องอาศัยโปรแกรมสำเร็จรูปที่พัฒนาขึ้นมาโดยเฉพาะ ซึ่งโปรแกรมสำเร็จรูปที่สามารถใช้ในการวิเคราะห์มีหลากหลายโปรแกรมซึ่งจะกล่าวในส่วนต่อไป

2. โปรแกรมสำเร็จรูปที่ใช้ในการวิเคราะห์โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า

โปรแกรมสำเร็จรูปสำหรับการวิเคราะห์ข้อสอบ⁵ เป็นโปรแกรมที่พัฒนาขึ้นเพื่อใช้สำหรับวิเคราะห์ข้อมูลที่ได้จากแบบทดสอบ แบบประเมิน แบบสอบถาม หรือแบบสำรวจต่าง ๆ ซึ่งสร้างขึ้นเพื่อใช้เก็บรวบรวมข้อมูลตัวแปรแฝงที่สนใจศึกษา ข้อมูลทางการวัดบางส่วนสามารถใช้โปรแกรมสำเร็จรูปสำหรับการวิเคราะห์ข้อมูลทางสถิติทั่วไป (เช่น SPSS) ในการวิเคราะห์ได้ ขณะที่การวิเคราะห์ข้อสอบ/ข้อคำถามส่วนใหญ่ต้องใช้โปรแกรมที่มีความเฉพาะเจาะจง การวิเคราะห์ข้อมูลการวัดตัวแปรด้านเจตพิสัยด้วยโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่าเป็นอีกกรณีหนึ่งที่นิยมใช้โปรแกรมสำเร็จรูปที่มีความเฉพาะเจาะจง อย่างไรก็ดี โปรแกรมสำเร็จรูปเหล่านี้มีหลากหลายโปรแกรม แต่ละโปรแกรมมีจุดเด่นและข้อจำกัดแตกต่างกันไปขึ้นอยู่กับบริบทและความต้องการของผู้ใช้ ในที่นี้ขอนำเสนอโปรแกรมสำเร็จรูปที่เกี่ยวข้องทั้งในส่วนของการเตรียมข้อมูลและการวิเคราะห์ ได้แก่ LERTAP, Xcalibre, IRTPRO และ MULTILOG

2.1 Laboratory of Educational Research Test Analysis Program หรือ LERTAP

⁵ โปรแกรมสำเร็จรูปสำหรับการวิเคราะห์ข้อสอบ ในที่นี้หมายความว่าโปรแกรมสำเร็จรูปทุกประเภทที่เกี่ยวข้องกับการวิเคราะห์ข้อสอบ ข้อคำถาม แบบทดสอบ มาตรวัด ทั้งในมุมของตัวแปรทางด้านพุทธิพิสัย เช่น ความสามารถ และตัวแปรทางด้านเจตพิสัย เช่น เจตคติ บุคลิกภาพ

LERTAP (Nelson, 2018) เป็นโปรแกรมสำเร็จรูปที่ใช้ในการวิเคราะห์ข้อสอบตามแนวคิดของทฤษฎีการทดสอบแบบดั้งเดิมซึ่งพัฒนาขึ้นให้สามารถใช้กับ Microsoft Excel ได้ โดย LERTAP สามารถวิเคราะห์ข้อมูลการวัดได้ทั้งในระดับรายข้อและรายฉบับ มีตัวเลือกทางสถิติให้เลือกใช้หลากหลาย สามารถตรวจสอบการโยงข้อสอบได้และแสดงผลเป็นแผนภาพรูปแบบต่าง ๆ ได้ เช่น กราฟเส้น กราฟแสดงความคลาดเคลื่อนทางการวัด สกรีนพลอต (Scree Plot) บ็อกพลอต (Box Plot) ฮิสโตแกรม (Histogram) แผนภาพกระจาย (Scatter Plot) เป็นต้น นอกจากนี้ยังสามารถวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) ด้วยวิธีการ Mantel-Haenszel ได้ ในส่วนของการวิเคราะห์ข้อสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบนั้น LERTAP สามารถช่วยเตรียมข้อมูลสำหรับการวิเคราะห์ด้วยโปรแกรมอื่น ๆ ได้ (การเตรียมข้อมูลเพื่อใช้กับโปรแกรมสำเร็จรูปที่เน้นการวิเคราะห์ข้อสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบโดยเฉพาะ เช่น Xcalibre, Bilog MG เป็นต้น) ขณะเดียวกันผู้ใช้งาน LERTAP ที่มีความคุ้นเคยกับ Microsoft Excel สามารถดาวน์โหลด EIRT ซึ่งเป็นส่วนเสริม ที่ช่วยให้สามารถวิเคราะห์ข้อสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบได้โดยไม่ต้องออกจาก LERTAP ซึ่งในส่วนของการทดสอบข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า นั้น EIRT สามารถวิเคราะห์ได้ 2 โมเดล คือ Nominal Model และ Graded Response Model ผู้ใช้งานสามารถดาวน์โหลดโปรแกรม LERTAP มาทดลองใช้ได้ โดยเวอร์ชันที่ไม่เสียค่าใช้จ่ายจะมีการจำกัดจำนวนกลุ่มตัวอย่างและจำนวนข้อสอบในการวิเคราะห์ หากต้องการวิเคราะห์กลุ่มตัวอย่างและข้อสอบจำนวนมากกว่าที่กำหนดไว้จะต้องเสียค่าใช้จ่ายเพิ่มเติม

2.2 XCalibre

XCalibre (Assessment Systems Corporation, 2018) เป็นโปรแกรมสำเร็จรูปสำหรับการวิเคราะห์ข้อสอบตามแนวคิดของโมเดลการตอบสนองข้อสอบ ซึ่งสามารถวิเคราะห์โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนสองค่า (Dichotomous IRT Models) ได้แก่ One-parameter Logistic Model (1PL), Two-parameter Logistic Model (2PL), Three-parameter Logistic Model (3PL) และโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า (Polytomous IRT Models) ได้แก่ Rating Scale Model (RSM), Partial Credit Model (PCM), Generalized Rating Scale Model (GRSM), Generalized Partial Credit Model (GPCM) และ Graded Response Model (GRM) รวมทั้งสามารถวิเคราะห์ข้อสอบในกรณีที่แบบวัด/มาตรวัดมีการผสมผสานรูปแบบของข้อสอบ/ข้อคำถามระหว่างการตรวจให้คะแนนแบบสองค่ากับแบบมากกว่าสองค่าได้ด้วย (Mixed-format Tests)

2.3 IRTPRO

IRTPRO (Cai, Thissen, & Du Toit, 2018) เป็นโปรแกรมสำเร็จรูปที่พัฒนาขึ้นมาใหม่ล่าสุด ซึ่งสามารถวิเคราะห์ข้อสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบได้อย่างครอบคลุม ไม่ว่าตัวแปรนั้นจะมีโครงสร้างแบบเอกมิติ (Unidimensional) หรือพหุมิติ (Multidimensional) ตรวจให้คะแนนสองค่า (ได้แก่ 1PL, 2PL, 3PL) หรือมากกว่าสองค่า (ได้แก่ GRM, GPCM, NRM) รวมทั้งสามารถวิเคราะห์ข้อสอบที่มีการผสมผสานรูปแบบการตรวจให้คะแนนทั้งสองแบบด้วย

2.4 MULTILOG

MULTILOG (Rupp, 2003) เป็นโปรแกรมสำเร็จรูปที่พัฒนาขึ้นเพื่อใช้สำหรับการวิเคราะห์ข้อสอบตามแนวคิดทฤษฎีการตอบสนองข้อสอบ สามารถวิเคราะห์โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่าได้ โปรแกรม MULTILOG อยู่ในลิขสิทธิ์ของบริษัท Scientific Software International สามารถวิเคราะห์โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่าได้หลากหลายโมเดล ได้แก่ Graded Response Model, Multiple Response Model, Normal Linear Model for Continuous Responses ปัจจุบันมีการปรับปรุงแบบของโปรแกรมให้ง่ายต่อการใช้งานของผู้ใช้ และสามารถแสดงโค้งฟังก์ชันของโมเดลการตอบสนองข้อสอบได้ โปรแกรมนี้ นับเป็นโปรแกรมวิเคราะห์ข้อสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบในยุคแรกๆ ที่ได้รับความนิยมมาถึงปัจจุบัน

3. ตัวอย่างการนำโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่าไปใช้ในการวิเคราะห์ตัวแปรด้านเจตพิสัย

O'Connor, Crawford and Holder (2015) ได้ศึกษาเกี่ยวกับการวิเคราะห์มาตรวัด Subjective Happiness Scale (SHS) ด้วยทฤษฎีการตอบสนองข้อสอบ โดยเก็บข้อมูลจากกลุ่มตัวอย่างซึ่งเป็นนักศึกษาในระดับปริญญาตรี จำนวน 1,037 คน ที่ตอบมาตรวัด SHS ในรูปแบบออนไลน์ มาตรวัด SHS เป็นมาตรวัดที่มีข้อคำถามจำนวน 4 ข้อ แต่ละข้อมีตัวเลือกรายการคำตอบแบบมาตราประมาณค่าของลิเคิร์ท จำนวน 7 ตัวเลือก ข้อมูลการตอบนำไปวิเคราะห์ด้วยโมเดล Graded Response Model (GRM) ของ Samejima's (1969) โดยใช้โปรแกรม ltm Package in R ซึ่งใช้สำหรับการวิเคราะห์ข้อคำถามที่มีการตรวจให้คะแนนมากกว่าสองค่า (Polytomous Items response) โดยมีการประมาณค่าพารามิเตอร์เทรชโฮลด์ จำนวน 6 ค่าในแต่ละข้อคำถาม (ตัวเลือก 7 ตัวเลือก มี 6 ช่วงเทรชโฮลด์) ผลการวิเคราะห์พบว่าโดยภาพรวมมาตรวัด SHS มีคุณสมบัติทางการวัดที่ดี อย่างไรก็ตาม ผลการวิเคราะห์พบว่าข้อคำถามทั้งสี่ข้อมีค่าอำนาจจำแนกต่ำในกลุ่มผู้ตอบที่มีความสุขอยู่ในระดับสูงมาก และจำนวนตัวเลือกรายการคำตอบ 7 ตัวเลือกนั้นมีมากเกินไปและไม่จำเป็นต่อการวัด นอกจากนี้พบว่าข้อคำถาม 1 ใน 4 ข้อ ให้สารสนเทศทางการวัดน้อย อาจเนื่องมาจากข้อคำถามใช้คำกำกวม ซึ่งเมื่อตัดข้อคำถามที่มีปัญหาออกพบว่าช่วยลดความยาวของมาตรวัดได้ร้อยละ 25 และมีคุณสมบัติทางการวัดที่ดี ไม่แตกต่างจากมาตรวัดที่มีข้อคำถาม 4 ข้อ จึงเป็นทางเลือกที่ดีซึ่งช่วยลดข้อสงสัยหรือความกำกวมของข้อคำถามได้อีกด้วย

Medvedev, Siegert, Kersten, and Krägeloh (2017) ได้ศึกษาเกี่ยวกับการเพิ่มความแม่นยำในการวัดของมาตรวัด Five Facet Mindfulness Questionnaire (FFMQ) ซึ่งเป็นมาตรวัดทางด้านการมีสติ โดยใช้แนวทางการวิเคราะห์ข้อสอบตามแนวคิดของราสช์โมเดล (Rasch Model) มาตรวัด FFMQ มีข้อคำถาม 39 ข้อ ประกอบด้วย มาตรวัดย่อย จำนวน 5 มาตรวัด ได้แก่ การสังเกต (Observing) การอธิบาย (Describing) การกระทำด้วยการตระหนักรู้ (Acting with Awareness) การไม่ตัดสิน (Non-Judging) และการไม่ได้ตอบกับประสบการณ์ภายใน (Nonreacting To Inner Experience) ข้อคำถามมีลักษณะตัวเลือกรายการคำตอบเป็นมาตราประมาณค่า 5 ระดับของลิเคิร์ท โดยมีทั้งข้อคำถามเชิงบวกและเชิงลบ กลุ่มตัวอย่าง

จำนวน 296 คน เป็นนักเรียน จำนวน 200 คน และเป็นประชาชนทั่วไปจำนวน 96 คน การเก็บข้อมูลในนักเรียนดำเนินการในชั้นเรียน ขณะที่การเก็บข้อมูลในประชาชนทั่วไปใช้การส่งตู้ไปรษณีย์ในเขตพื้นที่ต่าง ๆ และให้ผู้ตอบส่งข้อมูลกลับโดยไม่เสียค่าใช้จ่าย การวิเคราะห์จำแนกเป็นสองส่วน คือ ข้อมูลทั่วไปและการวิเคราะห์ความเที่ยงใช้โปรแกรมสำเร็จรูป SPSS ขณะที่การวิเคราะห์ตามแนวคิดของราล์ชโมเดลใช้โปรแกรม RUMM2030 ของ Andrich et al (2009) ซึ่งวิเคราะห์ด้วย Unrestricted Partial Credit Model ทั้งในส่วนของการวัดและมาตรวัดย่อย ผลการวิเคราะห์พบว่าเมื่อตัดข้อคำถามที่ไม่เหมาะสมออก 2 ข้อ มาตรวัด FFMQ มีคุณสมบัติทางการวัดที่ดีและมีความตรงเชิงโครงสร้างที่ดี ผู้ที่เกี่ยวข้องสามารถใช้ตารางการแปลงค่าคะแนนแบบเรียงอันดับเป็นค่าคะแนนแบบอันตรภาคชั้น (Ordinal-to-Interval Rasch Conversion Tables) เพื่อเพิ่มความแม่นยำในการวัดให้มากขึ้นได้ ซึ่งหากต้องการความแม่นยำในการวัดเพิ่มมากขึ้นควรใช้คะแนนแบบอันตรภาคชั้น

De Melo et al. (2018) ศึกษาเกี่ยวกับมาตรวัด World Health Organization Quality of Life (WHOQOL-OLD) โดยมีวัตถุประสงค์ในการศึกษา คือ 1) เพื่อเปรียบเทียบความตรงของมาตรวัด WHOQOL-OLD แบบสามตัวเลือกเทียบกับแบบเดิมซึ่งมี 5 ตัวเลือก และ 2) เพื่อตรวจสอบคุณสมบัติทางการวัดของมาตรวัด WHOQOL-OLD แบบ 3 ตัวเลือก ด้วยทฤษฎีการตอบสนองข้อสอบ กลุ่มตัวอย่างในการศึกษา คือ ผู้สูงอายุที่มีอายุตั้งแต่ 60 ปีขึ้นไป ซึ่งอาศัยอยู่ในฝั่งตะวันออกเฉียงเหนือของบราซิล โดยแบ่งกลุ่มตัวอย่างออกเป็น 2 กลุ่ม กลุ่มหนึ่งทำมาตรวัด WHOQOL-OLD แบบ 3 ตัวเลือก จำนวน 335 คน ขณะที่อีกกลุ่มหนึ่งทำมาตรวัด WHOQOL-OLD แบบ 5 ตัวเลือก จำนวน 265 คน มาตรวัด WHOQOL-OLD เป็นมาตรวัดที่สร้างขึ้นเพื่อประเมินคุณภาพชีวิตของผู้สูงอายุ โดยแปลภาษาและตรวจสอบคุณภาพของมาตรวัดก่อนนำมาใช้กับประชากรชาวบราซิล มาตรวัดนี้มีข้อคำถามทั้งสิ้น 24 ข้อ โดยมีตัวเลือกการคำตอบเป็นแบบมาตรประมาณค่า 5 ระดับของลิเคิร์ท ตั้งแต่ 1 หมายถึง ไม่เลย 2 หมายถึง แทบจะไม่ 3 หมายถึง ปานกลาง 4 หมายถึงค่อนข้างมาก และ 5 หมายถึง อย่างมาก มาตรวัดนี้มีทั้งหมด 6 องค์ประกอบ ได้แก่ ฟังก์ชันการรับรู้สัมผัส (Sensory Function: SF) ความใกล้ชิดสนิทสนม (Intimacy: Int) ความเป็นตัวของตัวเอง (Autonomy: Aut) การมีส่วนร่วมทางสังคม (Social Participation: SP) กิจกรรมในอดีต ปัจจุบัน และอนาคต (Past, Present, and Future activities: PPF) และความตายและการควบคุมอาหาร (Death and Diet: D&D) ซึ่งในส่วนของมาตรวัด WHOQOL-OLD แบบ 3 ตัวเลือก จะมีตัวเลือกการคำตอบ 3 ตัวเลือก คือ 0 หมายถึง ไม่เลย 1 หมายถึง ปานกลาง 2 หมายถึง อย่างมาก การเก็บรวบรวมข้อมูลทำโดยนักศึกษาระดับปริญญาตรีที่ได้รับการฝึกฝนมาอย่างดี โดยเก็บข้อมูลกับผู้สูงอายุที่บ้านพัก การวิเคราะห์ข้อมูลเริ่มต้นด้วยการตรวจสอบข้อตกลงเบื้องต้นตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ โดยใช้โปรแกรม R นอกจากนี้ยังได้เปรียบเทียบโมเดลการตอบสนองข้อสอบที่มีความเหมาะสมกับการวิเคราะห์ข้อมูลระหว่าง Rating Scale Model ของ Andrich (1978) กับ Graded Response Model (GRM) โดยใช้ Chi-square เปรียบเทียบค่า Likelihood ของแต่ละโมเดล ซึ่งพบว่า Graded Response Model (GRM) เหมาะสมมากกว่าจึงนำมาใช้ในการประมาณค่าพารามิเตอร์โดยใช้โปรแกรม ltm Package ผลการวิเคราะห์ข้อมูลความสอดคล้องของโมเดลใน WHOQOL-OLD แบบสามตัวเลือกพบว่ามีค่าดัชนีความสอดคล้องอยู่ในระดับดี ขณะที่ WHOQOL-OLD แบบ

5 ตัวเลือกมีดัชนีความสอดคล้องของโมเดลอยู่ในระดับยอมรับได้ถึงต่ำ ค่าความเที่ยงของมาตรวัดแบบ 3 ตัวเลือกอยู่ระหว่าง 0.52-0.77 ขณะที่แบบ 5 ตัวเลือกมีค่าความเที่ยงอยู่ระหว่าง 0.50-0.70 ซึ่งผลการวิเคราะห์แสดงให้เห็นว่าการปรับลดตัวเลือกรายการคำตอบสามารถทำได้โดยมาตรวัดยังคงมีคุณสมบัติทางการวัดที่ไม่แตกต่างจากมาตรวัดรูปแบบเดิม

4. บทสรุป

ตัวแปรด้านเจตพิสัย เป็นตัวแปรที่แสดงถึงคุณลักษณะภายในของบุคคลที่ไม่สามารถสังเกตได้โดยตรง แต่มีอิทธิพลอยู่เบื้องหลังพฤติกรรมที่แสดงออก การวัดตัวแปรด้านเจตพิสัยจึงต้องวัดผ่านตัวแปรสังเกตได้ หรือ พฤติกรรมที่บุคคลแสดงออกผ่านสิ่งเร้าที่กำหนดให้ โดยนำทฤษฎีการวัดมาเป็นแนวทางในการแสดงความสัมพันธ์ระหว่างพฤติกรรมกับตัวแปรที่ต้องการวัดอย่างเป็นเหตุเป็นผล ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีการทดสอบแนวใหม่ที่ได้รับการนิยมนำไปใช้เพื่อพัฒนาคุณภาพของแบบวัดอย่างแพร่หลาย โดยทฤษฎีนี้เป็นทฤษฎีที่มีลักษณะเป็นกลุ่มโมเดลเชิงคณิตศาสตร์ การวิเคราะห์ข้อมูลจะต้องเลือกโมเดลให้เหมาะสมกับลักษณะข้อมูล และบริบทต่าง ๆ ตัวแปรด้านเจตพิสัยส่วนใหญ่ก็มีตัวเลือกรายการคำตอบแบบมาตรประมาณค่า โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่าจึงเป็นที่นิยมใช้ โมเดลในกลุ่มนี้ที่เป็นที่รู้จักโดยทั่วไปมี 6 โมเดล ได้แก่ Graded Response Model (GRM) Modified Graded Response Model (M-GRM) Partial Credit Model (PCM) Generalized Partial Credit Model (G-PCM) Rating Scale Model (RSM) Nominal Response Model (NRM) การวิเคราะห์ข้อมูลตามทฤษฎีการตอบสนองข้อสอบจำเป็นต้องใช้โปรแกรมสำเร็จรูปที่พัฒนาขึ้นมาโดยเฉพาะ โปรแกรมสำเร็จรูปที่เกี่ยวข้องทั้งในส่วนของการเตรียมข้อมูลและการวิเคราะห์ตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ ได้แก่ LERTAP, Xcalibre, IRTPRO, MULTILOG

บรรณานุกรม

ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 4). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

Assessment Systems Corporation. (2018). *Xcalibre: Bringing you the benefits of Item*

Response Theory (IRT). Retrieved 20 July 2018 from <http://www.assess.com/Xcalibre/>

Cai, L., Thissen, D., & Du Toit, S. (2018). *IRTPRO 2.1 for Windows*. Retrieved 20 July 2018 from http://www.ssicentral.com/irt/IRTPRO_by_SSI.pdf

Crocker, L., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. United States of America: Cengage Learning.

De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. United States of America: The Guildford Press.

- De Melo, R. L. P., Da Silva Júnior, E.G., Souto, R. Q., Leão, I. S., & Do Carmo Eulálio, M. (2018). Psychometric properties of the complete version of the World Health Organization Quality of Life Assessment (WHOQOL-OLD): reduced response scale. *Psicologia: Reflexão e Crítica*, 31(4), 1-10. DOI: 10.1186/s41155-018-0084-1
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. United States of America: Sage Publications.
- Medvedev, O. N., Siegert, R. J., Kersten, P., & Krägeloh, C. U. (2017). Improving the Precision of the Five Facet Mindfulness Questionnaire Using a Rasch Approach. *Mindfulness*, 8, 995–1008.
- Nelson, L. R. (2018). *IRT in Lertap 5*. Retrieved 20 July 2018 from <http://larrynelsonstuff.com/Documentation/IRTinLertap5.pdf>
- O'Connor, B. P., Crawford, M. R., & Holder, M. D. (2015). An Item Response Theory Analysis of the Subjective Happiness Scale. *Social Indicators Research*, 124, 249–258.
- Rupp, A. A. (2003). Item Response Modeling with BILOG-MG and MULTILOG for Windows. *International Journal of Testing*, 3(4), 365–384.