

การวิเคราะห์ข้อมูลตามทฤษฎีการตอบสนองข้อสอบ แบบตรวจให้คะแนนมากกว่า 2 ค่า โดยใช้โปรแกรมอาร์

Data Analysis in Polytomous Item Response Theory Using R

ภูรินทร์ เทพสถิตย์¹ ภัคจิรา จงสุกใส¹ ภัคจิรา บวรธรรมรัตน์¹ วรเชษฐ แซ่เจีย^{1*} และ ณภัทร ชัยมงคล²
Purin Thepsathit¹, Pakjira Jongsooksai¹, Pakjira Bowornthammarat¹, Vorachet Saejea^{1*} and
Nhabhat Chaimongkol²

(Received: July 8, 2021; Revised: August 18, 2021; Accepted: August 23, 2021)

บทคัดย่อ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) สามารถแบ่งออกเป็น 2 กลุ่มตามผลการตรวจให้คะแนนคำตอบ คือ ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (dichotomous IRT) และทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (polytomous IRT) ในการทดสอบทางการศึกษามีวิธีการตรวจให้คะแนนแบบหลายค่า เช่น มาตรฐานค่า เกณฑ์การประเมิน (รูบริก) การให้คะแนนความรู้บางส่วน บทความนี้จะนำเสนอการวิเคราะห์ข้อมูลด้วยโมเดลการให้คะแนนความรู้บางส่วน และโมเดลการให้คะแนนแบบลำดับขั้น ซึ่งเป็นโมเดลที่ครอบคลุมการตรวจสอบคุณภาพของเครื่องมือแบบตรวจให้คะแนนมากกว่า 2 ค่า ที่ใช้ทั่วไปในบริบทการศึกษา ผ่านโปรแกรมอาร์ ซึ่งเป็นโปรแกรมที่เป็นที่ยอมรับโดยทั่วไป สามารถดาวน์โหลดเพื่อนำไปใช้และพัฒนาต่อได้โดยไม่ต้องเสียค่าใช้จ่าย และรองรับระบบปฏิบัติการที่หลากหลาย ทั้งนี้ผู้เขียนได้เสนอเนื้อหาครอบคลุมถึงทฤษฎีการตอบสนองข้อสอบ โมเดลในทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า คำสั่งพื้นฐานของโปรแกรม และการวิเคราะห์ด้วยโมเดลการให้คะแนนความรู้บางส่วน และ โมเดลการให้คะแนนแบบลำดับขั้น เพื่อนำไปใช้เป็นแนวทางสำหรับการวิเคราะห์ที่เกี่ยวข้องกับการทดสอบทางการศึกษา การประเมินทางจิตวิทยา หรือทางด้านอื่น ๆ ต่อไป

คำสำคัญ: โมเดลทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า

โมเดลการให้คะแนนความรู้บางส่วน โมเดลการให้คะแนนแบบลำดับขั้น โปรแกรมอาร์

¹ นิสิตระดับปริญญาโท ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

² อาจารย์ ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

¹ Master's degree student, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

² Lecturer, Doctor, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

* Corresponding Author E-mail: v.saejea@gmail.com

Abstract

Item Response Theory (IRT) can be divided into two distinct categories based on the result of the response—Dichotomous IRT and Polytomous IRT. In educational testing contexts, polytomous scoring methods such as rating scales, rubrics, and partial credit scoring are used. This article then focuses on how to analyze data with the Partial Credit Model (PCM) and the Graded Response Model (GRM), which are models that cover quality examination of polytomous scoring tools generally used in the educational context through R. R is a well-known open-source program and is accepted by the public. It could be downloaded for use and for further development for free. In addition, it can be used in various operating systems. In the paper, we discuss the basics of the IRT, polytomous IRT models, R functions, and the analysis for PCM and GRM in educational measurement, psychological assessment, and so on.

Keywords: Polytomous IRT Model; Partial Credit Model; Graded Response Model; R programming language

บทนำ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT) เป็นทฤษฎีการทดสอบแนวใหม่ (modern test theory) พัฒนาขึ้นครั้งแรกโดย Louis Thurstone ในปี 1920s เนื่องจากข้อจำกัดของทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory; CTT) ในการประมาณค่าพารามิเตอร์ข้อสอบและแบบสอบ ได้แก่ ค่าความยากและอำนาจจำแนกแปรเปลี่ยนไปตามกลุ่มผู้สอบ กล่าวคือค่าพารามิเตอร์ของข้อสอบจากแบบสอบฉบับเดียวกันอาจมีความแตกต่างกันเมื่อนำไปใช้กับกลุ่มผู้สอบที่ต่างกัน ด้วยเหตุนี้ทฤษฎีการตอบสนองข้อสอบจึงเข้ามาช่วยแก้ไขปัญหาดังกล่าว โดยที่การประมาณค่าพารามิเตอร์ของข้อสอบจะใช้การประมาณค่าจากผลการตอบข้อสอบรายข้อ และค่าพารามิเตอร์ของข้อสอบจะไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ กล่าวคือเมื่อนำข้อสอบฉบับเดียวกันไปใช้กับกลุ่มผู้สอบที่แตกต่างกัน การประมาณค่าพารามิเตอร์ของข้อสอบจะมีความคงที่ (Kanjawasee, 2020; van der Linden, 2016)

ทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีที่เน้นอธิบายความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบถูก (probability) และตัวแปรแฝง (latent Trait) เช่น ความสามารถของผู้สอบ เจตคติของนักเรียน หรือความเครียดของผู้สอบ โดยใช้ฟังก์ชันทางคณิตศาสตร์คือฟังก์ชันโลจิสติก (logistic function) ซึ่งแสดงในรูปโค้งลักษณะข้อสอบ (Item Characteristic Curve; ICC) ในการอธิบายความสัมพันธ์ของ 1) ความสามารถของผู้สอบ (θ) กับ 2) พารามิเตอร์ข้อสอบ ซึ่งมี 3 พารามิเตอร์ ได้แก่ พารามิเตอร์ความยาก (b) พารามิเตอร์อำนาจจำแนก (a) และ พารามิเตอร์ความน่าจะเป็นในการเดาข้อสอบถูก (c) ความสัมพันธ์ดังกล่าวอาจอยู่ในรูปของโมเดล 1 พารามิเตอร์ (1PL), 2 พารามิเตอร์ (2 PL) หรือ 3 พารามิเตอร์

(3PL) (Lekhawattananpong, 2017; Pinitasakul et al., 2020; Sakolkijrungrong, n.d.; Kanjanawasee, 2020; Embretson & Reise, 2000; van der Linden, 2016)

ทฤษฎีการตอบสนองข้อสอบแบ่งออกเป็น 2 ประเภทใหญ่ ๆ คือ ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (dichotomous IRT) ซึ่งใช้กับเครื่องมือที่มีผลการตอบแบบ 2 ค่า เช่น 0, 1 (ตอบผิดได้ 0, ตอบถูกได้ 1) หรือใช่-ไม่ใช่ และทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (polytomous IRT) ใช้กับเครื่องมือที่มีผลการตอบมากกว่า 2 ค่า เช่น มาตรฐานค่า (rating scale) 5 ระดับ (เห็นด้วยอย่างยิ่ง, เห็นด้วย, ไม่แน่ใจ, ไม่เห็นด้วย และไม่เห็นด้วยอย่างยิ่ง) หรือ การตรวจให้คะแนนความรู้บางส่วน (partial credit) (Kanjanawasee, 2020) โดยทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า ศึกษา 2 พารามิเตอร์ ได้แก่ พารามิเตอร์อำนาจจำแนกและพารามิเตอร์เทรชโฮลด์ ซึ่งเปรียบเทียบกับพารามิเตอร์ความยากของทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า แต่ในการให้คะแนนมากกว่า 2 ค่า ผู้ตอบอาจมีตัวเลือกที่ต้องตอบได้มากกว่า ถูกหรือผิด เช่น การตอบข้อคำถามแบบมาตรฐานค่า 5 ระดับ ผู้ตอบสามารถเลือกตอบได้ตั้งแต่ 1 2 3 4 หรือ 5 ซึ่งช่องว่างระหว่าง 5 ระดับ จะเป็นจุดตัดที่ผู้ตอบตัดสินใจเลือกในระดับที่สูงขึ้น ซึ่งจุดตัดเรียกว่าเทรชโฮลด์ โดยในที่นี้ก็จะจะมี 4 เทรชโฮลด์

โมเดลทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า มีผู้พัฒนาโมเดลขึ้นมาหลากหลายโมเดล ขึ้นอยู่กับความเหมาะสมของการนำไปใช้ โดยโมเดลที่นิยมใช้ในการตรวจสอบคุณภาพเครื่องมือที่ให้คะแนนมากกว่า 2 ค่า มีดังนี้ (Kanjanawasee, 2020; Ostini & Nering, 2011)

1. Rating Scale Model (RSM) เป็นโมเดลแบบ 1 พารามิเตอร์ที่วิเคราะห์พารามิเตอร์เทรชโฮลด์เหมาะสมกับเครื่องมือวัดเจตคติที่เป็นลักษณะของมาตรฐานค่า (rating scale) โดยโมเดลกำหนดให้พารามิเตอร์เทรชโฮลด์ ของแต่ละจำนวนรายการคำตอบ (category) ในมาตรเท่ากันทั้งฉบับ

2. Partial Credit Model (PCM) เป็นโมเดลแบบ 1 พารามิเตอร์ที่วิเคราะห์พารามิเตอร์เทรชโฮลด์เหมาะสมกับเครื่องมือวัดผลสัมฤทธิ์หรือทักษะที่เป็นลักษณะของรูบริก โดยโมเดลกำหนดให้พารามิเตอร์เทรชโฮลด์ ของแต่ละจำนวนรายการคำตอบในมาตรแตกต่างกันได้ และมีค่าอำนาจจำแนกในแต่ละข้อเท่ากัน

3. Generalized Partial Credit Model (GPCM) เป็นโมเดลแบบ 2 พารามิเตอร์ที่พัฒนาขึ้นจากโมเดล PCM โดยมีการวิเคราะห์พารามิเตอร์อำนาจจำแนกเพิ่มเติม เหมาะสมกับเครื่องมือที่เป็นแบบวัดผลสัมฤทธิ์แบบการให้คะแนนบางส่วน

4. Graded Response Model (GRM) เป็นโมเดลแบบ 2 พารามิเตอร์ที่วิเคราะห์พารามิเตอร์เทรชโฮลด์และพารามิเตอร์อำนาจจำแนก เหมาะสมกับเครื่องมือวัดผลสัมฤทธิ์หรือเจตคติที่เป็นลักษณะของมาตรเรียงลำดับ เช่น แบบสอบที่ให้คะแนนความรู้บางส่วน หรือแบบวัดที่เป็นมาตรฐานค่า โดยโมเดลอนุญาตให้แต่ละข้อมีจำนวนรายการคำตอบที่แตกต่างกันได้

5. Modified Graded Response Model (MGRM) เป็นโมเดลแบบ 2 พารามิเตอร์ มีลักษณะการวิเคราะห์ที่คล้ายกับโมเดล GRM แต่เครื่องมือจะต้องมีจำนวนรายการคำตอบเท่ากัน เหมาะกับเครื่องมือประเภทมาตรฐานค่า

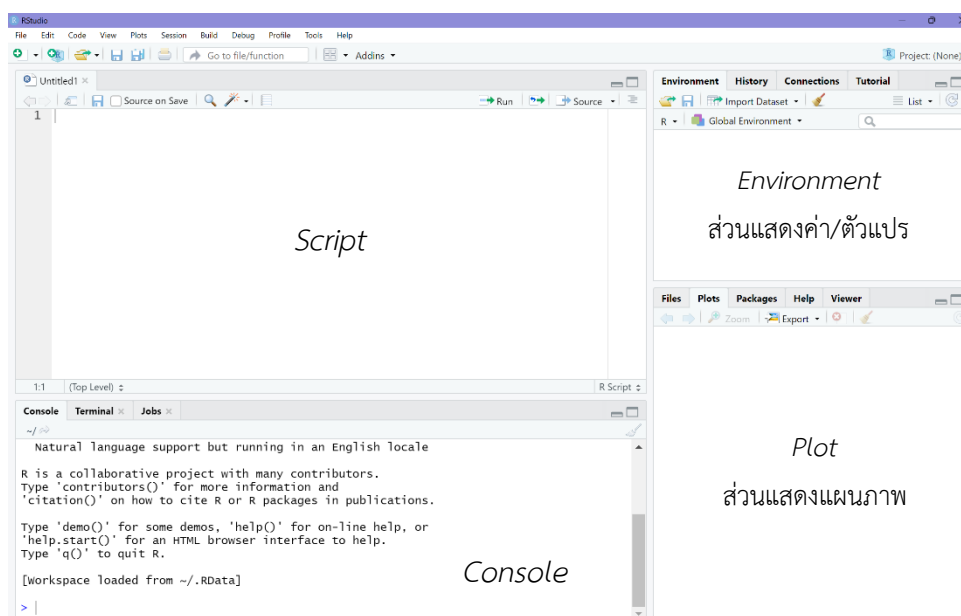
6. Nominal Response Model (NRM) เป็นโมเดลแบบ 2 พารามิเตอร์ ที่ใช้สำหรับวิเคราะห์เครื่องมือที่มีผลการตอบที่ไม่เรียงลำดับ เช่น ข้อสอบแบบหลายตัวเลือก หรือแบบวัดบุคลิกลักษณะ ลักษณะการตอบที่เหมาะสมในโมเดลนี้ จะต้องลักษณะของการตอบ เช่น จำนวนการตอบ 3 ประเภท ได้แก่ บ้านที่ทำงาน หรือห่าง

โมเดลที่เป็นที่สามารถนำมาใช้เพื่อตรวจสอบคุณภาพของเครื่องมือแบบให้คะแนนมากกว่า 2 ค่า คือ โมเดล PCM และ โมเดล GRM เนื่องจากครอบคลุมกับเครื่องมือในการวัดทางจิตวิทยาและการศึกษา ในปัจจุบัน ซึ่งมักจะอยู่ในรูปของมาตรประมาณค่าเพื่อวัดเจตคติของนักเรียน ระบุริ้วเพื่อวัดผลสัมฤทธิ์หรือทักษะใดทักษะหนึ่งโดยเฉพาะ หรือการให้คะแนนความรู้บางส่วน ซึ่งครูและนักการศึกษาสามารถพัฒนาขึ้นเองได้ ดังนั้น บทความนี้จึงนำเสนอการวิเคราะห์ด้วยโมเดล PCM และ GRM

แนะนำโปรแกรมอาร์

โปรแกรมอาร์ (R)¹ เป็นโปรแกรมสำหรับการวิเคราะห์ข้อมูล การคำนวณเชิงสถิติและการแสดงผลในรูปแบบกราฟิก โดยใช้ภาษาอาร์ถูกพัฒนาขึ้นให้ใช้งานง่ายเพราะชื่อตัวแปรและฟังก์ชันต่าง ๆ ผูกโยงกับความหมายของคำในภาษามนุษย์ ทำให้ผู้ใช้สามารถเข้าใจได้ง่าย (Sakworawit, 2017) และได้รับความน่าเชื่อถือในการวิเคราะห์ข้อมูลทางสถิติ โปรแกรมอาร์ได้รับการคิดค้นและพัฒนาโดย Ross Ihaka และ Robert Gentleman และเผยแพร่เมื่อ ค.ศ. 1995 โปรแกรมอาร์ เป็นซอฟต์แวร์ประเภทโอเพนซอร์ส (open source) ที่ทุกคนสามารถเข้าถึงและดาวน์โหลดเพื่อนำไปใช้และพัฒนาต่อได้โดยไม่ต้องเสียค่าใช้จ่าย และไม่ต้องกังวลเรื่องการละเมิดลิขสิทธิ์ รองรับระบบปฏิบัติการทั้ง Windows, MacOS และ Linux โดยโปรแกรมอาร์สามารถจำลองข้อมูล (simulation) วิเคราะห์สถิติเชิงบรรยาย (descriptive statistics) และสถิติเชิงอนุมาน (inferential statistics) และยังสามารถแสดงผลทางด้านกราฟิกที่เข้าใจง่าย จุดเด่นของโปรแกรมอาร์ คือ สามารถเขียนโปรแกรมเสริม (package) ได้มากมายครอบคลุมการวิเคราะห์ในหลากหลายสาขาวิชา และเมื่อสั่งการทำงานโปรแกรมจะแสดงผลลัพธ์ในทันที จึงได้รับความนิยมในทางวิชาการเป็นอย่างมาก ทั้งนี้การใช้งานโปรแกรมอาร์สามารถใช้งานร่วมกับโปรแกรม RStudio (ดังภาพ 1) เพื่อช่วยอำนวยความสะดวกในการใช้งานฟังก์ชันต่าง ๆ เนื่องจากมีการแยกพื้นที่การเขียนคำสั่ง (script) ออกจากส่วนแสดงผล (console) สำหรับในบทความฉบับนี้จะเป็นการสาธิตการวิเคราะห์โมเดล PCM และ GRM ด้วยโปรแกรมอาร์ร่วมกับโปรแกรม RStudio

¹ โปรแกรมอาร์สามารถดาวน์โหลดและติดตั้งได้จาก <https://cran.r-project.org/>



ภาพ 1 พื้นที่การทำงานและรายละเอียดส่วนติดต่อกับผู้ใช้ของโปรแกรม RStudio

บทความวิชาการนี้เหมาะสำหรับผู้ที่มีความรู้พื้นฐานการใช้โปรแกรมอาร์เบื้องต้น เช่น การเตรียมชุดข้อมูล และการเขียนชุดคำสั่ง (syntax) ซึ่งผู้ที่ยังไม่มีพื้นฐานการเขียนโปรแกรมอาร์สามารถศึกษาเพิ่มเติมได้จาก Chaimongkol and Kanjanawasee (2020) อย่างไรก็ตามในบทความนี้จะขอแนะนำคำสั่งพื้นฐานเบื้องต้นที่จำเป็นในการวิเคราะห์และการเตรียมข้อมูล ดังนี้

```
#R01
getwd()
setwd("C:/Users/lenkk/OneDrive/documents/R Project/606")

#R02
install.packages("mirt")

#R03
library(mirt)
library(stats4)
library(lattice)

#R04
dat <- read.csv(file = "C:/Users/lenkk/OneDrive/documents/
R Project/606/new-resp.csv", header = TRUE)

#R05
dat <- read.table(file = "C:/Users/lenkk/OneDrive/documents/
R Project/606/new-resp.txt", header = TRUE)
```

ภาพ 2 ตัวอย่างฟังก์ชันกำหนดพื้นที่ทำงาน การติดตั้งแพ็คเกจ และการนำเข้าข้อมูล

1. การกำหนดพื้นที่ทำงาน (working directory) เมื่อเปิดโปรแกรมอาร์ขึ้นมาแล้ว ควรกำหนดพื้นที่ทำงานให้เป็นพื้นที่จัดเก็บข้อมูลในการทำงานแต่ละครั้ง สามารถตรวจสอบได้โดยใช้ฟังก์ชัน `getwd()` เพื่อเรียกชื่อพื้นที่ทำงานที่กำหนดไว้ให้ปรากฏในส่วนแสดงผล (console) และหากต้องการเปลี่ยนพื้นที่ก็สามารถทำได้โดยใช้ฟังก์ชัน `setwd()` โดยกำหนดตำแหน่งในเครื่องหมายคำพูด (double quotation marks) ระหว่างพื้นที่วงเล็บ รายละเอียดเป็นดังส่วน #R01 ในภาพ 2

2. การติดตั้งแพ็คเกจ (package installation) ในการประมวลผลและแสดงผลในโปรแกรมอาร์สามารถทำได้ฟังก์ชันเพียงเล็กน้อยเท่านั้น จึงจำเป็นต้องติดตั้งแพ็คเกจเพิ่มเติมเพื่อเพิ่มประสิทธิภาพและขอบเขตของการทำงาน โดยใช้คำสั่ง `install.packages()` และใส่ชื่อแพ็คเกจ ในเครื่องหมายคำพูด และเมื่อติดตั้งเรียบร้อยแล้ว ก่อนจะใช้คำสั่งในแพ็คเกจนั้นจะต้องเรียกใช้งานด้วยฟังก์ชัน `library()` ข้อมูลตัวอย่างการติดตั้งและเรียกใช้แพ็คเกจ “mirt” แสดงในส่วน #R02 และ #R03 ของภาพ 2 ตามลำดับ โดยเมื่อติดตั้งแล้วจะมีแพ็คเกจ “stats4” และ “lattice” ติดตั้งมาให้พร้อมกันด้วย เมื่อใช้แพ็คเกจ “mirt” จึงต้องเรียกใช้อีกสองแพ็คเกจด้วยเพื่อให้ทำงานได้สมบูรณ์

3. การเตรียมและนำเข้าข้อมูล (data preparation and input) ในการคำนวณชุดข้อมูลที่รวบรวมมาได้ จำเป็นต้องจัดกระทำข้อมูลให้อยู่ในรูปแบบที่โปรแกรมอาร์รองรับได้ นั่นคืออยู่ในไฟล์นามสกุล .csv หรือ .txt ที่สามารถสร้างผ่านโปรแกรม Microsoft Excel ซึ่งมีฟังก์ชันนำเข้าข้อมูลทั้งสองรูปแบบต่างกัน ได้แก่ `read.csv()` และ `read.table()` ตามลำดับ โดยใส่ชื่อไฟล์หรือตำแหน่งของไฟล์ในเครื่องหมายคำพูด และควรจัดเก็บให้อยู่ในรูปตัวแปรหนึ่งในระบบมีรายละเอียดในส่วน #R04 และ #R05 ของภาพ 2 โดยตัวอย่างชุดคำสั่งเป็นดังภาพ 3

	A	B	C	D	E	F
1	ID	V1	V2	V3	V4	V5
2	1	3	3	2	1	1
3	2	3	1	1	1	1
4	3	3	3	2	1	0
5	4	3	3	3	1	0
6	5	3	1	0	0	0
7	6	3	3	2	0	0
8	7	3	1	1	0	0
9	8	3	3	2	1	0
10	9	3	2	0	1	0
11	10	3	1	0	0	0

ภาพ 3 ตัวอย่างชุดคำสั่งที่ใช้ในการวิเคราะห์ (ไฟล์นามสกุล .csv)

การวิเคราะห์ด้วยโมเดลการให้คะแนนบางส่วน (Partial Credit Model: PCM) และโมเดลการให้คะแนนแบบลำดับขั้น (Graded Response Model: GRM)

โมเดลการให้คะแนนบางส่วน (PCM) มีเป้าหมายเพื่อศึกษาพารามิเตอร์เทรซโฮลด์ของข้อคำถาม (item) และจำนวนรายการคำตอบ (scale/categories) ภายในข้อคำถาม ส่วนโมเดลการให้คะแนนแบบลำดับขั้น (GRM) จะให้ความสำคัญกับการเรียงลำดับของระดับความสามารถ/คุณลักษณะในการตอบในแต่ละตัวเลือก (scale/category) ภายในข้อคำถามเดียวกันว่าเป็นไปตามอันดับรายการคำตอบที่วางไว้หรือไม่ ในการวิเคราะห์ด้วยโมเดลทั้งสองด้วยโปรแกรมอาร์ มีแพ็คเกจสำหรับการวิเคราะห์เป็นจำนวนมาก เช่น mirt, ltm, eRm เป็นต้น สำหรับบทความนี้นำเสนอแพ็คเกจ mirt ซึ่งมีความครอบคลุมในการวิเคราะห์ และมีการนำมาใช้ในงานวิจัยในหลากหลายมุมมอง (Freeman et al., 2021; Satıcı et al., 2020; Weiss & Roberts, 2018) โดยมีรายละเอียดคำสั่งและคำอธิบายของแต่ละขั้นตอน ดังนี้

1. การกำหนดค่าสำหรับการสร้างโมเดลการให้คะแนนบางส่วน (PCM) และโมเดลการให้คะแนนแบบลำดับขั้น (GRM)

คำสั่งในการสร้างโมเดลการให้คะแนนบางส่วนจากข้อมูลผลการตอบที่ได้จากการใช้เกณฑ์การประเมิน โดยในตัวอย่างนี้ชุดข้อมูลถูกจำลองขึ้นจำนวน 100 ชุด มีลักษณะเป็นมาตราประมาณค่า 5 ข้อ แต่ละข้อมี 4 ระดับการเลือกตอบ มีรายละเอียดการวิเคราะห์ดังภาพ 4

```
#R06
library(mirt)
library(stats4)
library(lattice)

data <- read.csv("C:/R/data.csv", header = TRUE) #R07
head(data) #R08

#R09
model <- mirt(data, 1, itemtype = 'Rasch')
summary(model)
```

ภาพ 4 ชุดคำสั่งการเรียกแพ็คเกจ การนำเข้าข้อมูล และการสร้างโมเดล

#R06 เรียกแพ็คเกจ mirt เพื่อใช้งาน โดยการเรียกใช้แพ็คเกจ mirt ซึ่งจะต้องติดตั้งและเรียกแพ็คเกจเสริม stats4 และ lattice ขึ้นมาใช้งานร่วมด้วยจึงจะทำให้การทำงานสมบูรณ์ โดยใช้คำสั่ง library()

#R07 ฟังก์ชันนำเข้าคะแนนของผู้สอบซึ่งบันทึกเป็นไฟล์ .csv ที่อยู่ใน Drive C -> โฟลเดอร์ R -> ไฟล์ data.csv ซึ่งถูกบันทึกให้อยู่ในตัวแปรชื่อ “data” โดยใช้คำสั่ง data <- read.csv (“C:/R/data.csv”, header = TRUE) โดยกำหนดให้ห้อยค์ประกอบ header เป็น TRUE เพื่อแจ้งกับโปรแกรมว่าข้อมูลมีหัวตาราง หากไม่ต้องการให้แสดงข้อมูลหัวตารางให้กำหนดเป็น FALSE

#R08 เรียกดูตัวอย่างข้อมูลภายในตัวแปร data ซึ่งจะแสดงผลให้เห็น 6 แถวแรก โดยใช้คำสั่ง head() แสดงผลดังภาพ 5A

#R09 นำข้อมูลที่ได้นำเข้ามาวิเคราะห์ด้วยโมเดลโดยใช้ฟังก์ชัน mirt ซึ่งต้องประกอบไปด้วย (1) ตัวแปรที่เป็นข้อมูล (2) จำนวนมิติ (Dimension) ที่ต้องการวัด ซึ่งในที่นี้กำหนดไว้ที่ 1 มิติหรือเอกมิติ (Unidimensionality) (3) ประเภทของโมเดลซึ่งในที่นี้หากต้องการวิเคราะห์ด้วยโมเดล PCM ให้ใช้คำสั่ง itemtype = ‘Rasch’ แต่หากต้องการวิเคราะห์ด้วยโมเดล GRM ให้ใช้คำสั่ง itemtype = ‘graded’ จากนั้นใช้คำสั่ง summary(model) เพื่อเรียกดูค่าที่ได้จากการวิเคราะห์ ดังภาพ 5B

5A

```
> head(data) #R08
```

	V1	V2	V3	V4	V5
1	3	3	2	1	1
2	3	1	1	1	1
3	3	3	2	1	0
4	3	3	3	1	0
5	3	1	0	0	0
6	3	3	2	0	0

5B

```
> summary(model)
```

	F1	h2
V1	0.402	0.162
V2	0.402	0.162
V3	0.402	0.162
V4	0.402	0.162
V5	0.402	0.162

SS loadings: 0.808
 Proportion Var: 0.162

Factor correlations:

	F1
F1	1

ภาพ 5 ผลการแสดงผลการนำเข้าข้อมูลด้วยคำสั่ง head(data) และคุณสมบัติของโมเดล PCM

2. การตรวจสอบความสอดคล้องของโมเดล PCM และ GRM

ในการตรวจสอบคุณลักษณะและความสอดคล้องของโมเดล PCM และ GRM มีประเด็นที่ควรให้ความสำคัญทั้งหมด 4 ประเด็น ได้แก่ การตรวจสอบความสอดคล้องกลมกลืน (fit statistic) ของข้อรายการประเมินที่ใช้โมเดล PCM การวิเคราะห์พารามิเตอร์ความยากและอำนาจจำแนก การวิเคราะห์คุณลักษณะแฝง (θ) มีชุดของคำสั่งแสดงดังภาพ 6

```
#R10
itemfit(model)

#R11
coef(model, IRTpars = TRUE)

#R12
theta <- matrix(seq(-3,3,.1))
tinfo <- testinfo(model, theta)
tinfo

#R13
theta_se <- fscores(model, full.scores.SE = TRUE, method = 'EAP')
empirical_rxx(theta_se)
```

ภาพ 6 ชุดคำสั่งในการวิเคราะห์โมเดล PCM

#R10 วิเคราะห์ความสอดคล้องกลมกลืนด้วยสถิติทดสอบ Chi-square จากตัวแปร model โดยใช้คำสั่ง itemfit()

#R11 เรียกดูพารามิเตอร์เทรชโฮลด์และอำนาจจำแนก โดยจะแสดงเป็นค่าพารามิเตอร์เทรชโฮลด์ที่อยู่บนค่า theta และกำหนดให้แสดงผลของค่าพารามิเตอร์ให้อยู่ในรูปแบบดั้งเดิม โดยใช้คำสั่ง IRTpars = TRUE

#R12 กำหนดตัวแปร theta โดยใช้คำสั่ง matrix ให้ค่า theta อยู่ในช่วง -3 ถึง 3 และกำหนดให้เพิ่มครั้งละ 0.1 จากนั้นใช้คำสั่ง testinfo วิเคราะห์สารสนเทศของแบบสอบโดยกำหนดให้อยู่ในตัวแปร tinfo โดยให้แสดงค่าตามตัวแปร theta ที่กำหนดไว้

#R13 ใช้คำสั่ง fscores ประมาณค่า factor score ของความสามารถและตัวแปรแฝง และบันทึกให้อยู่ในตัวแปร theta_se โดยใช้คำสั่ง 1) full.scores.SE = TRUE เพื่อกำหนดให้ความคลาดเคลื่อนมาตรฐานสัมพันธ์กับผู้สอบเป็นรายบุคคล และ 2) method = 'EAP' เพื่อประมาณค่าโดยใช้วิธีการของ expected a posteriori (Chianchana, 2009) จากนั้นเรียกดูค่าความเที่ยงในตัวแปร theta_se โดยใช้คำสั่ง empirical_rxx

2.1 ผลการวิเคราะห์ข้อมูล

1) การวิเคราะห์ความสอดคล้องกลมกลืน (fit statistic) ของข้อสอบ

ในการตรวจสอบจะใช้สถิติ Chi-square test ในการพิจารณา โดยมีเกณฑ์ที่ใช้ตรวจสอบ 3 รายการหลัก ประกอบด้วย (1) ค่า Chi-squareหารด้วย Degree of Freedom (χ^2/df) ต้องไม่เกิน 2 (2) ค่า RMSEA มีค่าน้อยกว่า .05 (3) ค่า p-value มีค่ามากกว่า .05 ซึ่งหากข้อรายการประเมินไม่สอดคล้องกลมกลืนกับโมเดล แสดงว่าข้อรายการประเมินนั้นไม่มีความตรงเชิงโครงสร้าง (Suksawang, 2013)

ผลการวิเคราะห์ความสอดคล้องกลมกลืนของข้อรายการประเมินที่ใช้โมเดล PCM พบว่า ทุกข้อสอดคล้องกับข้อมูลเชิงประจักษ์โดยพิจารณาจากค่าสถิติ 2 ใน 3 สถิติ ได้แก่ χ^2/df ($df.S_X2$) มีค่าไม่เกิน 2 RMSEA ($RMSEA.S_X2$) มีค่าน้อยกว่า .05 และค่า p-value ($p.S_X2$) มีค่ามากกว่า .05 ซึ่งเป็นการสนับสนุนว่าโมเดล PCM สอดคล้องกับข้อมูลเชิงประจักษ์ ผลการวิเคราะห์ทั้งหมดเป็นไปดังภาพ 7

> itemfit(model)					
	item	S_X2	df.S_X2	RMSEA.S_X2	p.S_X2
1	V1	4.785	7	0.000	0.686
2	V2	5.512	8	0.000	0.702
3	V3	12.572	11	0.038	0.322
4	V4	10.496	11	0.000	0.486
5	V5	9.678	7	0.062	0.208

ภาพ 7 ผลการวิเคราะห์ความสอดคล้องกลมกลืนของข้อสอบรายข้อของโมเดล PCM

2) การวิเคราะห์พารามิเตอร์เทรซโฮลด์และอำนาจจำแนก

2.1) แนวการวิเคราะห์โมเดล PCM

โมเดล PCM ศึกษาพารามิเตอร์เทรซโฮลด์เพียงพารามิเตอร์เดียว โดยกำหนดให้พารามิเตอร์อำนาจจำแนกมีค่าเท่ากับ 1 ทั้งหมด โดยผลการวิเคราะห์ด้วยโมเดล PCM ตามชุดข้อมูลตัวอย่างพบว่า ค่าพารามิเตอร์อำนาจจำแนกเท่ากับ 1 ทุกข้อ ซึ่งตรงกับทฤษฎีของโมเดล PCM ที่กำหนดให้ค่าพารามิเตอร์อำนาจจำแนกมีค่าเท่ากับ 1 ส่วนค่าพารามิเตอร์เทรซโฮลด์ แสดงผลเป็นเทรซโฮลด์การตอบ โดยข้อที่ 1 รายการคำตอบที่ 0 คะแนนไม่ได้ถูกเลือก จึงเหลือรายการคำตอบที่ 1 2 และ 3 เท่านั้น ดังนั้นจึงแสดงเทรซโฮลด์รายการตอบเพียง 2 เทรซโฮลด์ และข้อที่ 4 พบว่าเทรซโฮลด์รายการคำตอบไม่เรียงลำดับ ซึ่งละเมิดข้อตกลงเรื่อง monotonicity

โดยพิจารณาจากเทรซโฮลด์รายการคำตอบที่ 2 (1.367) ยากกว่าจุดตัดที่ 3 (1.263) ดังนั้นควรมีการปรับปรุงรายการคำตอบที่ 2 ให้สามารถวัดผู้ที่ความสามารถระดับนั้นได้ตรงมากขึ้น ส่วนข้อที่ 2 3 และ 5 มีค่าพารามิเตอร์เทรซโฮลด์ที่ใช้ได้ ดังภาพ 8

2.2) แนวการวิเคราะห์โมเดล GRM

โมเดล GRM ศึกษาพารามิเตอร์เทรซโฮลด์และพารามิเตอร์อำนาจจำแนก โดยจะมีค่าพารามิเตอร์แตกต่างกันรายข้อ โดยผลการวิเคราะห์ด้วยโมเดล PCM ตามชุดข้อมูลตัวอย่างพบว่าผลการวิเคราะห์ด้วยโมเดล GRM พบว่าข้อที่ 1 พารามิเตอร์อำนาจจำแนก 0.879 และมีพารามิเตอร์ความยาก -4.354, -2.664 และ -0.840 ตามลำดับ ข้อที่ 2 พารามิเตอร์อำนาจจำแนก 1.078 และมีพารามิเตอร์ความยาก -3.197, -1.442 และ 0.347 ตามลำดับ ข้อที่ 3 พารามิเตอร์อำนาจจำแนก 2.190 และมีพารามิเตอร์ความยาก -1.403, -0.015 และ 1.154 ตามลำดับ ข้อที่ 4 พารามิเตอร์อำนาจจำแนก 2.093 และมีพารามิเตอร์ความยาก -0.213, 0.974 และ 2.119 ตามลำดับ และข้อที่ 5 พารามิเตอร์อำนาจจำแนก 2.209 และมีพารามิเตอร์ความยาก 0.313, 1.521 และ 2.215 ตามลำดับ ผลการวิเคราะห์ทั้งหมดเป็นดังภาพ 8

> coef(model, IRTpars = TRUE)				> coef(model, IRTpars = TRUE)			
\$V1				\$V1			
	a	b1	b2		a	b1	b2
par	1	-1.806	-0.564		par	0.879	-4.354
\$V2				\$V2			
	a	b1	b2		a	b1	b2
par	1	-2.274	-0.431		par	1.078	-3.197
\$V3				\$V3			
	a	b1	b2		a	b1	b2
par	1	-1.094	0.357		par	2.19	-1.403
\$V4				\$V4			
	a	b1	b2		a	b1	b2
par	1	-0.353	1.367		par	2.093	-0.213
\$V5				\$V5			
	a	b1	b2		a	b1	b2
par	1	1.469	1.652		par	2.209	0.313
\$GroupPars				\$GroupPars			
	MEAN_1	COV_11			MEAN_1	COV_11	
par	0	0.63		par	0	1	

ภาพ 8 ผลการวิเคราะห์พารามิเตอร์ของโมเดล PCM (ซ้าย) และ GRM (ขวา)

จากผลการวิเคราะห์สามารถสรุปได้ว่า ข้อมูลเชิงประจักษ์มีความสอดคล้องกับโมเดล PCM และเมื่อพิจารณาถึงเทรซโฮลด์รายข้อคำถามพบว่า อาจต้องปรับปรุงรายการคำตอบในข้อคำถามที่ 1 และ 4 เนื่องจากมีบางรายการคำตอบไม่ถูกเลือกใช้ ซึ่งรายการคำตอบอาจไม่เป็นตัวแทนที่ดีของระดับความสามารถผู้ตอบ หรือรายการคำตอบที่ไม่เรียงลำดับ หมายถึง รายการคำตอบที่อยู่ระดับก่อนหน้าอาจต้องใช้ระดับความสามารถ (θ) สูงกว่าในการตอบ

เมื่อพิจารณาโมเดล GRM พบว่า เทพโซลด์มีการเรียงลำดับชั้นอย่างเหมาะสม กล่าวคือรายการคำตอบมีการเรียงลำดับตามระดับความสามารถที่ต้องใช้ในการตอบข้อคำถามจากต่ำไปสูง และแต่ละข้อมีการกระจายของผลการตอบที่หลากหลายในแต่ละระดับความสามารถ เมื่อพิจารณาค่าเทพโซลด์ซึ่งมีค่าอยู่ระหว่าง -4.354 ถึง 2.215 และค่าอำนาจจำแนกซึ่งมีค่าอยู่ระหว่าง 0.879 ถึง 2.209 พบว่า ค่าเทพโซลด์และค่าอำนาจจำแนกเป็นไปตามเกณฑ์ กล่าวคือข้อคำถามมีความเหมาะสมและสามารถจำแนกความสามารถของผู้ตอบได้ในระดับดี (Kanjanaawasee, 2020)

3) การวิเคราะห์สารสนเทศของข้อสอบ

จากผลการวิเคราะห์พบว่า มีค่าสูงสุดอยู่ตำแหน่งที่ 42 มีค่าสารสนเทศเท่ากับ 2.971 หมายถึงแบบสอบสามารถให้สารสนเทศของผู้สอบที่มีความสามารถค่อนข้างไปทางเก่ง ($\theta = 1.1$) ได้ดีที่สุด ผลการวิเคราะห์ทั้งหมดแสดงดังภาพ 9

```
> theta <- matrix(seq(-3,3,.1))
> tinfo <- testinfo(modelPCM,theta)
> tinfo
[1] 0.7323654 0.7877477 0.8458913 0.9067637 0.9703130 1.0364674 1.1051339
[8] 1.1761964 1.2495139 1.3249173 1.4022050 1.4811382 1.5614347 1.6427617
[15] 1.7247289 1.8068823 1.8886999 1.9695918 2.0489071 2.1259488 2.2000000
[22] 2.2703604 2.3363928 2.3975768 2.4535618 2.5042140 2.5496486 2.5902399
[29] 2.6266048 2.6595592 2.6900486 2.7190595 2.7475164 2.7761739 2.8055087
[36] 2.8356168 2.8661230 2.8961069 2.9240563 2.9478604 2.9648590 2.9719667
[43] 2.9658832 2.9433929 2.9017287 2.8389547 2.7542975 2.6483518 2.5231034
[50] 2.3817539 2.2283798 2.0674952 1.9036068 1.7408367 1.5826601 1.4317708
[57] 1.2900619 1.1586890 1.0381845 0.9285906 0.8295915
```

ภาพ 9 ผลการวิเคราะห์สารสนเทศของข้อสอบของโมเดล PCM

4) การวิเคราะห์ความเที่ยง

ผลการวิเคราะห์ค่าความเที่ยงแบบทฤษฎีการตอบสนองข้อสอบด้วยวิธีการ expected a posteriori (EAP) พบว่ามีค่าความเที่ยงเท่ากับ 0.622 ซึ่งอยู่ในเกณฑ์พอใช้ (Kanjanaawasee, 2005) โดยค่าความเที่ยงแบบทฤษฎีการตอบสนองข้อสอบสามารถเทียบเคียงได้กับความเที่ยงในทฤษฎีการทดสอบแบบดั้งเดิม จึงใช้วิธีการแปลผลแบบเดียวกันได้ (Adams, 2005) ผลการวิเคราะห์ฟังก์ชันเป็นดังภาพ 10

```
> theta_se <- fscores(modelPCM, full.scores.SE = TRUE, method = 'EAP')
> empirical_rxx(theta_se)
F1
0.6218714
```

ภาพ 10 ผลการวิเคราะห์ความเที่ยงของโมเดล PCM

2.2 การพล็อตกราฟ

การพล็อตกราฟสำหรับโมเดล IRT นิยมวาดกราฟสารสนเทศและกราฟโค้งคุณลักษณะ ซึ่งสามารถวาดกราฟได้ทั้งฉบับของแบบสอบและรายข้อสอบ รวมถึงสามารถวาดกราฟแสดงค่าความเที่ยงแบบ IRT ของแบบสอบได้อีกด้วย มีรายละเอียดของคำสั่งดังภาพ 11 และการแสดงกราฟในภาพ 12

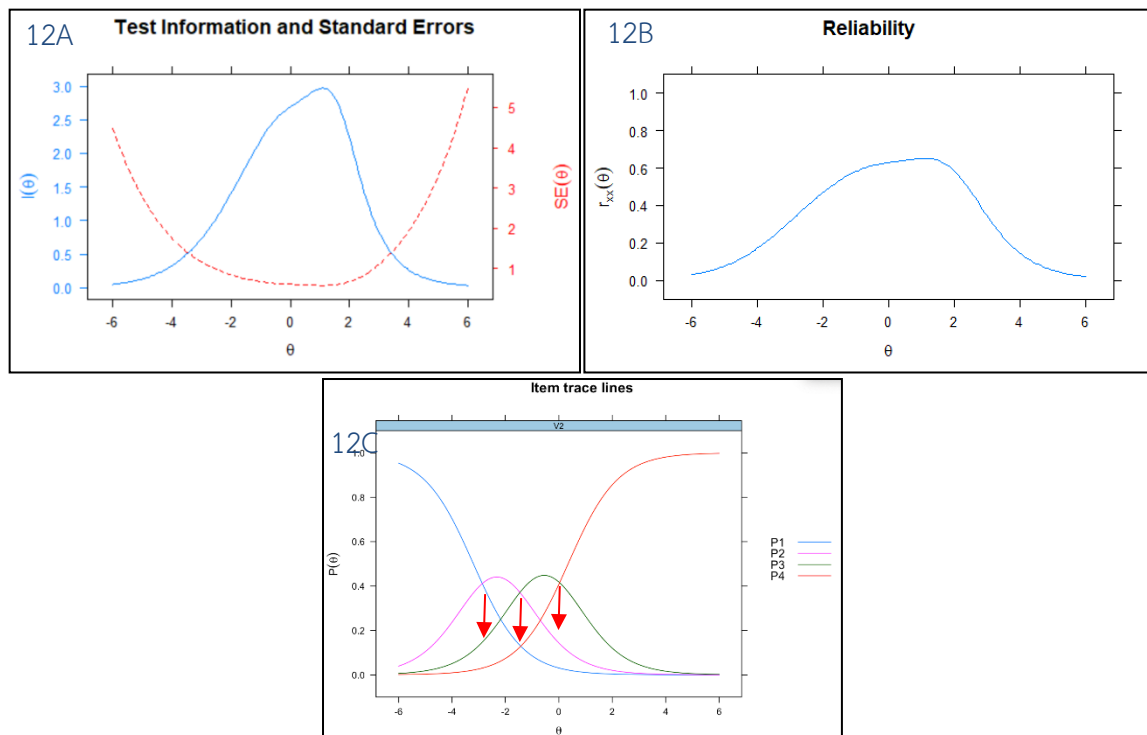
```
plot(model, type = 'infoSE') #R14.1
plot(model, type = 'rxx') #R14.2
plot(model, type = 'trace', which.items = 1) #R14.3
```

ภาพ 11 ชุดคำสั่งในการสร้างกราฟของโมเดล PCM

#R14.1 แสดงกราฟสารสนเทศและความคลาดเคลื่อนมาตรฐานของแบบสอบ โดยใช้คำสั่ง plot(model, type = 'infoSE') โดยเส้นทึบสีฟ้าแทนสารสนเทศแบบสอบ (test information) และ เส้นประสีแดงแทนค่าความคลาดเคลื่อนมาตรฐาน (standard error)

#R14.2 แสดงกราฟความเที่ยงของแบบสอบ โดยใช้คำสั่ง plot(model, type = 'rxx')

#R14.3 วาดกราฟพารามิเตอร์เทรซโฮลด์ภายในข้อที่ 2 ของแบบสอบ โดยใช้คำสั่ง plot(model, type = 'trace', which.items = 2) โดย P1 P2 P3 และ P4 แทนโอกาสในการเลือกตอบระดับที่ 1 2 3 และ 4 ตามลำดับ



ภาพ 12 ผลการแสดงผลกราฟของโมเดล PCM ตามคำสั่ง #R14.1 – #R14.3

จากกราฟผลการวิเคราะห์ พบว่า แบบสอบมีความเหมาะสมกับผู้ที่มีความสามารถปานกลางค่อนข้างไปทางเก่ง โดยพิจารณาจากกราฟสารสนเทศและความคลาดเคลื่อนมาตรฐานของแบบสอบ (ภาพ 12A) ซึ่งกราฟสารสนเทศ (เส้นทึบสีฟ้า) มีค่าสูงสุดในตำแหน่ง θ ซึ่งมีค่าเท่ากับ 1.1 และกราฟความคลาดเคลื่อนมาตรฐาน (เส้นประสีแดง) ที่เกิดขึ้นจากผู้ตอบที่มีความสามารถปานกลางค่อนข้างไปทางเก่งจะเกิดความคลาดเคลื่อนของการวัดที่ต่ำ นอกจากนี้ กราฟความเที่ยง (ภาพ 12B) ก็มีความสอดคล้องกับกราฟสารสนเทศของแบบสอบ ซึ่งมีความเที่ยงสูงสุดในตำแหน่งใกล้เคียงกัน

สำหรับกราฟรายการคำตอบหรือ trace lines (ภาพ 12C) แสดงถึงรายการคำตอบและจุดตัดหรือเทอร์ชโฮลด์ (ลูกศรสีแดง) ของแต่ละข้อรายการ ซึ่งจากกราฟ พบว่า รายการคำตอบและเทอร์ชโฮลด์มีการเรียงลำดับที่เหมาะสม แสดงให้เห็นว่าแต่ละรายการคำตอบต้องใช้ระดับความสามารถในการตอบเรียงลำดับจากต่ำไปสูงในการตอบข้อคำถาม

สรุป

ในบริบททางการศึกษามีทั้งทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า และทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า โดยบทความนี้นำเสนอการวิเคราะห์ด้วยโมเดล PCM และ GRM ตามแนวคิดทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (polytomous IRT) ที่มีความเหมาะสมสำหรับการนำไปใช้ในการวิเคราะห์ข้อมูลที่ได้จากเครื่องมือการประเมินทางการศึกษามักมีรูปแบบเป็นการตรวจให้คะแนนคำตอบถูกบางส่วน ในแบบวัดผลสัมฤทธิ์ หรือมาตรฐานค่า ในแบบวัดเจตคติ บุคลิกภาพ และเชาวน์ปัญญา เป็นต้น โดยแบ่งการนำเสนอออกเป็น 3 ตอน คือ ทฤษฎีการตอบสนองข้อสอบ โปรแกรมอาร์ และการวิเคราะห์ด้วยโมเดลการให้คะแนนบางส่วนและโมเดลการให้คะแนนแบบลำดับขั้น ซึ่งทั้งหมดนี้ทำการวิเคราะห์ผ่านโปรแกรมอาร์

โปรแกรมประเภทโอเพนซอร์สอย่างโปรแกรมอาร์มีผู้พัฒนาแพ็คเกจสำหรับการวิเคราะห์ขึ้นมาอย่างต่อเนื่อง ไม่ว่าจะเป็นการวิเคราะห์ด้วยโมเดล PCM GRM หรือโมเดลอื่น ๆ ผู้ที่สนใจยังสามารถศึกษาและวิเคราะห์ผ่านแพ็คเกจอื่นนอกจาก “mirt” ที่ใช้นำเสนอเป็นหลักในบทความนี้ได้ เช่น “ltm” (สามารถวิเคราะห์ได้ทั้งสองโมเดล) “eRm” (สามารถวิเคราะห์ได้เฉพาะ PCM) นอกจากนี้หากข้อมูลที่ต้องการวิเคราะห์มีลักษณะที่แตกต่างออกไปอาจต้องทำการวิเคราะห์ด้วยโมเดลอื่นที่มีความเหมาะสมกับลักษณะข้อมูลและเป้าหมายในการวิเคราะห์ของท่านได้ เช่น หากต้องการวิเคราะห์เครื่องมือมาตรฐานค่าที่มีช่วงกว้างของคะแนนเท่ากัน ก็สามารถใช้ Rating Scale Model (RSM) ในการวิเคราะห์ได้

References

- Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation, 31*, 162-172. <https://doi.org/10.1016/j.stueduc.2005.05.008>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates Publishers.
- Freeman, D., Loe, B. S., Kingdon, D., Startup, H., Molodynski, A., Rosebrock, L., Brown, P., Sheaves, B., Waite, F., & Bird, J. C. (2021). The revised Green et al., Paranoid Thoughts Scale (R-GPTS): Psychometric properties, severity ranges, and clinical cut - offs. *Psychological Medicine, 51*(2), 244-253. <https://doi.org/10.1017/s0033291719003155>
- Nering, M. L., & Ostini, R. (2011). New Perspectives and Applications. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 3-20). Routledge.

- Satici, B., Gocet-Tekin, E., Deniz, M. E., Satici, S. A. (2020). Adaptation of the fear of COVID-19 scale: Its association with psychological distress and life satisfaction in Turkey. *International Journal of Mental Health and Addiction*.
<https://doi.org/10.1007/s11469-020-00294-0>
- van der Linden, W. J. (2016). *Handbook of item response theory, volume one: Models*. Chapman & Hall/CRC.
- Weiss, S., & Roberts, R. D. (2018). Using anchoring vignettes to adjust self-reported personality: A comparison between countries. *Frontiers in Psychology*, 9.
<https://doi.org/10.3389/fpsyg.2018.00325>

Translated Thai Reference

- Chianchana, C. (2009). Multidimensional analysis. *Journal of Education Khon Kaen University*, 32(4), 13-22. (in Thai)
- Chaimongkol, N., & Kanjanawasee, S. (2020). The application of R programming for educational testing. *Journal of Research Methodology*, 33(1), 21-45.
<https://portal.edu.chula.ac.th/pub/jrm/index.php/jrm/article/view/648>. (in Thai)
- Lekhawatthanapong, N. (2017). *A comparison of partial knowledge scoring method quality: A development of modified number right-elimination scoring method* [Master's Thesis]. Chulalongkorn University Intellectual Repository (CUIR).
<http://cuir.car.chula.ac.th/handle/123456789/59866> (in Thai)
- Pinitasakul, P., Senarat, S., & Senarat, B. (2020). Item response theory and testlet test. *Ubon Ratchathani Journal of Research and Evaluation*, 9(2), 74-84. <https://so06.tci-thaijo.org/index.php/ubonreseva/article/view/243461> (in Thai)
- Suksawang, P. (2013). The Basics of Structural Equation Modeling. *Princess Naradhiwas University Journal*, 6(2), 136-145. (in Thai)
- Kanjanawasee, S. (2005). Classical test theories (5th Ed.). Chulalongkorn University Press. (in Thai)
- Kanjanawasee, S. (2020). *Modern test theories* (5th Ed.). Chulalongkorn University Press. (in Thai)
- Sakolkijrungrong, S. (n.d.). *Polytomous item response theory models for affective variables*.
<https://www.stou.ac.th/Offices/ore/info/cae/viewkb.aspx?id=291>. (in Thai)
- Sakworawit, A. (2017). *Descriptive statistics, exploratory data analysis, and statistical graphic with basic R Programming*. Chulalongkorn University Press. (in Thai)