

บทความวิจัย (Research Article)

การพัฒนาโมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ
โดยใช้โมเดลทฤษฎีการตอบสนองข้อสอบทางเลือกที่คำนึงถึงการเดา
DEVELOPMENT OF A MULTILEVEL MIXTURE IRT MODEL
USING ALTERNATIVE IRT MODEL WITH GUESSING

Received: September 12, 2020

Revised: October 8, 2020

Accepted: October 14, 2020

ธีรุตม์ สุขสกุลวัฒน์^{1*} ณัฐภรณ์ หลาวทอง² และสิวะโชติ ศรีสุทธียากร³
Teerut Suksakulwat^{1*} Nuttaporn Lawthong² and Siwachoat Srisuttiyakorn³

^{1,2,3}คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย^{1,2,3}Faculty of Education, Chulalongkorn University, Bangkok 10330, Thailand

*Corresponding Author, E-mail: teerut1234@gmail.com

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์ 3 ประการ คือ 1) เพื่อพัฒนาโมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับที่ประยุกต์จากโมเดล 2PLE (MMix2PLE) และวิธีการประมาณค่าที่เหมาะสม 2) เพื่อเปรียบเทียบความถูกต้องของการตรวจสอบจำนวนกลุ่มแฝงระหว่าง AIC กับ BIC สำหรับโมเดล MMix3PL และ MMix2PLE และ 3) เพื่อเปรียบเทียบความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง ความสามารถเฉลี่ย และพารามิเตอร์ของข้อสอบ ระหว่างโมเดล MMix3PL กับ MMix2PLE ข้อมูลที่นำมาศึกษาเป็นข้อมูลจำลองจากวิธีการมอนติคาร์โลภายใต้สถานการณ์จำลอง 32 เงื่อนไข (โมเดลจำลอง 2 เงื่อนไข x ขนาดตัวอย่าง 2 เงื่อนไข x โมเดลแข่งขัน 8 เงื่อนไข) แต่ละเงื่อนไขถูกจำลองซ้ำ 10 รอบ รวมจึงมีข้อมูลทั้งสิ้น 320 ชุด ผลการวิจัยสรุปได้ดังนี้

1. ขั้นตอนวิธีการประมาณค่าพารามิเตอร์แบบเบสสำหรับโมเดล MMix2PLE ที่พัฒนาในงานวิจัยนี้พบว่าค่อนข้างมีความเหมาะสม สามารถหาค่าประมาณพารามิเตอร์ได้เพียงหนึ่งเดียว (Unique) และประมาณค่าพารามิเตอร์ย้อนกลับได้ค่อนข้างดีภายใต้เงื่อนไขการจำลอง
2. การเปรียบเทียบความถูกต้องของการตรวจสอบจำนวนกลุ่มแฝง พบว่า AIC มีแนวโน้มมีความถูกต้องมากกว่า BIC สำหรับทั้งโมเดล MMix3PL และ MMix2PLE
3. การเปรียบเทียบความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง ความสามารถเฉลี่ย และพารามิเตอร์ของข้อสอบ พบว่าโมเดล MMix2PLE มีแนวโน้มมีความแม่นยำสูงกว่าโมเดล MMix3PL โดยเฉพาะเมื่อขนาดตัวอย่างใหญ่ ($N = 2,500$)

คำสำคัญ: โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ โมเดลทฤษฎีการตอบสนองข้อสอบทางเลือกที่คำนึงถึงการเดา การคัดเลือกโมเดล

Abstract

This study had three purposes: 1) to develop a multilevel mixture IRT model applied from the 2PLE model (MMix2PLE) and an appropriate estimation method; 2) to compare the accuracy of detection of latent classes between AIC and BIC for the MMix3PL and MMix2PLE models; and 3) to compare the precision of estimation: the latent classes, mean ability, and item parameters between the MMix3PL and MMix2PLE models. The data were simulated from the Monte Carlo method under 32 conditions (2 simulation models x 2 sample sizes x 8 candidate models), each condition was replicated 10 times, yielding a total of 320 data sets. The research results were as follows:

1. A Bayesian estimation algorithm for the MMix2PLE model developed in this research was found to be quite suitable. It was able to estimate unique parameter and recover parameter quite well under simulated conditions.
2. Comparison of the accuracy of detection of latent classes, AIC tends to be more accurate than BIC for both the MMix3PL and MMix2PLE models.
3. Comparison of the precision of estimation: the latent classes, mean ability, and item parameters, the MMix2PLE model tended to be more precise than the MMix3PL model, especially when the sample size was large ($N = 2,500$).

Keywords: MMixIRT Model, Alternative IRT Model with Guessing, Model Selection

บทนำ

โมเดลทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นตัวแบบที่นิยมใช้อย่างแพร่หลายในการวัดผลทางการศึกษา เนื่องจากมีข้อตกลงเบื้องต้นที่ได้เปรียบกว่าโมเดลทฤษฎีการทดสอบแบบดั้งเดิม เช่น ความเป็นอิสระ (Local Independence) ความไม่แปรเปลี่ยน (Invariance) และการเป็นฟังก์ชันต่อเนื่อง (Continuous Function) (Kanjana-wasee, 2012, pp. 75-78) อย่างไรก็ตามการนำโมเดล IRT เช่น โมเดล Rasch โมเดลโลจิส 2 พารามิเตอร์ (2PL: Two-Parameter Logistic) และโมเดลโลจิส 3 พารามิเตอร์ (3PL) มาวิเคราะห์ข้อสอบปรนัยซึ่งมีการตรวจให้คะแนนสองค่าคือ 0 (ผิด) กับ 1 (ถูก) อาจพบว่า มีข้อจำกัดในการวัดผล ดังนี้ ประการที่หนึ่ง ข้อจำกัดเกี่ยวกับความไม่แปรเปลี่ยนของโมเดล IRT ซึ่งมีสมมติฐานว่าผู้สอบมาจากประชากรที่มีภาวะเอกพันธ์กันเพียงกลุ่มเดียว (Single Homogeneous Population) สมมติฐานดังกล่าวอาจเป็นการฝ่าฝืนธรรมชาติของข้อมูล เนื่องจากโดยทั่วไปผู้สอบในประชากรเดียวกันอาจมีคุณลักษณะที่แตกต่างกันได้ เช่น นักเรียนจากโรงเรียนเดียวกันอาจมีแบบแผนการตอบ (Response Pattern) ข้อสอบที่แตกต่างมาก จนเชื่อได้ว่าผู้สอบน่าจะมาจากประชากรกลุ่มย่อย (Subpopulation) คนละกลุ่มมากกว่า (Tay et al., 2011, pp. 184-185) ประการที่สอง ข้อจำกัดเกี่ยวกับการละเลยต่อโครงสร้างพหุระดับของข้อมูล (Multilevel Data Structure) เนื่องจากธรรมชาติของข้อมูลทางการศึกษา มักพบว่าข้อมูลในระดับล่างจะซ้อน (Nested) อยู่ภายใต้ข้อมูลในระดับบน เช่น ข้อมูลนักเรียนซ้อนอยู่ภายใต้ห้องเรียน ห้องเรียนซ้อนอยู่ภายใต้โรงเรียน เป็นต้น อย่างไรก็ตามการวิเคราะห์ด้วยโมเดล IRT โดยทั่วไปยังไม่มีคำนึงถึงโครงสร้างดังกล่าว ผลการวัดในระดับหน่วย (Micro-Level) เช่น ระดับนักเรียน จึงอาจถูกรบกวนจากตัวแปรในระดับกลุ่ม (Macro-Level) เช่น ระดับโรงเรียน (Maier, 2001, pp. 308-309) ประการที่สาม ข้อจำกัดเกี่ยวกับการวัดการเดา เนื่องจากข้อสอบมาตรฐานทั่วไปมักเป็นข้อสอบปรนัยแบบหลายตัวเลือกที่มีการตรวจให้คะแนนเป็นลูกกับผิด ซึ่งเปิดช่องให้มีการนำความรู้บางส่วน (Partial Knowledge) เข้ามาช่วยในการเดาคำตอบได้ เช่น การตัดตัวเลือกที่ทราบว่าจะไม่ถูกออกไปก่อน

แล้วจึงเดาตอบแบบสุ่มจากตัวเลือกที่เหลือ เป็นต้น การนำข้อมูลผลการตอบข้อสอบดังกล่าวมาวิเคราะห์ด้วยโมเดล Rasch หรือ 2PL ซึ่งไม่มีการคำนึงถึงการเดา จึงอาจทำให้สารสนเทศเกี่ยวกับความรู้บางส่วนของผู้สอบที่ใช้ในการเดาถูกละเลยไป ส่วนโมเดล 3PL แม้จะมีการคำนึงถึงการเดา แต่ก็อาจมีข้อจำกัดเกี่ยวกับการประมาณค่าพารามิเตอร์การเดาย้อนกลับ ซึ่งมักพบว่าไม่ค่อยมีความสัมพันธ์กับค่าจริงภายใต้เงื่อนไขการศึกษาเชิงจำลอง รวมถึงปรัชญาความเชื่อว่าโอกาสในการเดาถูกเป็นคุณลักษณะของข้อสอบ มิได้ผันแปรตามความรู้ของผู้สอบ การแปลผลการเดาของโมเดล 3PL จึงยังไม่สอดคล้องกับสภาพธรรมชาติเท่าที่ควร (Lee & Bolt, 2018, pp. 91-92)

จากข้อจำกัดข้างต้น แนวทางในการหลีกเลี่ยงข้อจำกัดของโมเดล IRT เกี่ยวกับการฝ่าฝืนความไม่แปรเปลี่ยนและการละเลยโครงสร้างพหุระดับของข้อมูล คือการเปลี่ยนมาวิเคราะห์ด้วยโมเดล IRT แบบผสมพหุระดับ (MMixIRT) (Cho & Cohen, 2010, pp. 336-338) ส่วนการหลีกเลี่ยงข้อจำกัดเกี่ยวกับการวัดการเดา คือการเปลี่ยนมาวิเคราะห์ด้วยโมเดล IRT ทางเลือกที่คำนึงถึงการเดา (Zhu et al., 2019, pp. 450-452) งานวิจัยนี้จึงต้องการบูรณาการแนวทางดังกล่าวมาพัฒนาเป็นโมเดล MMixIRT ที่มีการประยุกต์จากโมเดล IRT ทางเลือกที่คำนึงถึงการเดา โดยจะทดลองนำโมเดล 2PLE (2PL Extension; Zhu et al., 2019, pp. 452-453) มาใช้ในการศึกษา เนื่องจากโมเดลมีความน่าสนใจหลายประการ อาทิ การนิยามให้โอกาสในการเดาถูกมีความผันแปรตามความสามารถของผู้สอบและคุณลักษณะของข้อสอบ การแปลผลการเดาจึงมีแนวโน้มสอดคล้องกับสภาพธรรมชาติมากขึ้น รวมถึงการนำค่าดังกล่าวมาคำนวณเป็นโอกาสในการเดาถูกยังเป็นการหลีกเลี่ยงปัญหาที่อาจจะเกิดขึ้นจากการประมาณค่าพารามิเตอร์การเดาโดยตรงด้วย

ทั้งนี้เพื่อเป็นการตรวจสอบถึงประสิทธิภาพของโมเดล MMixIRT ที่พัฒนาจากโมเดล 2PLE ในงานวิจัยนี้จะนำโมเดลดังกล่าวมาเปรียบเทียบกับโมเดล MMix3PL ภายใต้สถานการณ์จำลอง เพื่อศึกษาว่าเกณฑ์สารสนเทศ (information criterion) ใดมีความเหมาะสมสำหรับการคัดเลือกโมเดล หรือมีความถูกต้องของการตรวจสอบจำนวนกลุ่มแฝงจริงของข้อมูลมากที่สุด และเพื่อศึกษาว่าการประมาณค่าพารามิเตอร์ในโมเดลมีความแม่นยำเพียงใด ผลจากการวิจัยจะช่วยให้สารสนเทศสำหรับการตัดสินใจเลือกใช้โมเดลในบริบทการทดสอบขนาดใหญ่ที่มีความไววิพันซ์และมีการเดาเกิดขึ้น และจะเป็นต้นแบบในการบูรณาการระหว่างโมเดล IRT กับโมเดลอื่นๆ ต่อไป

วัตถุประสงค์การวิจัย

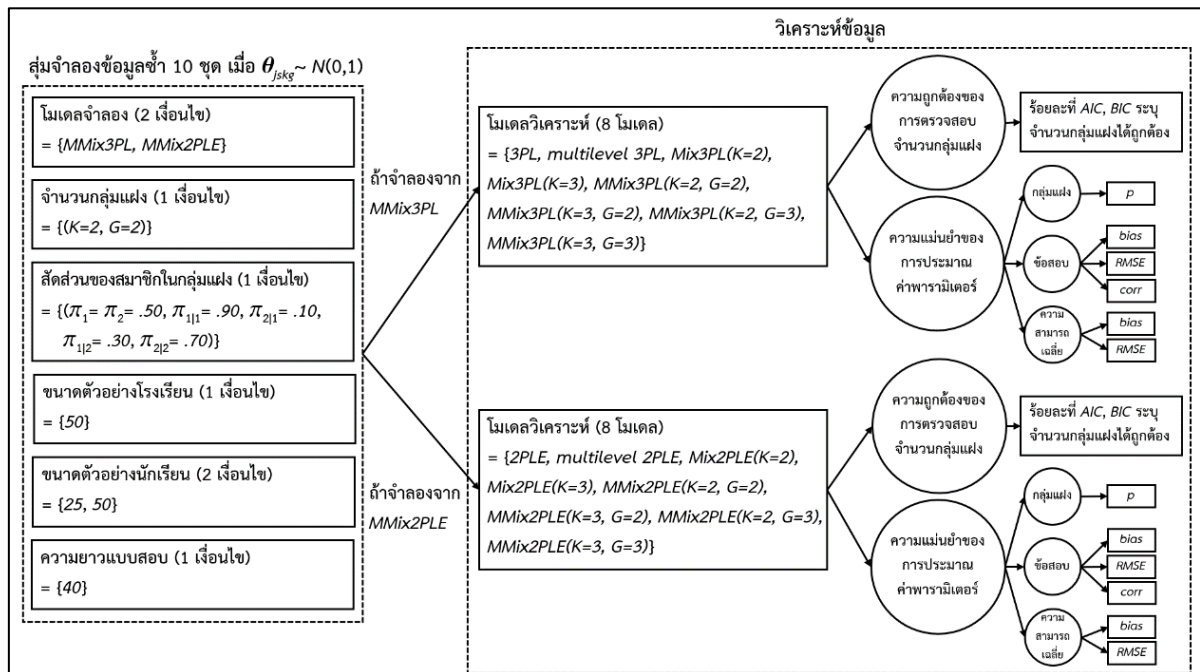
1. เพื่อพัฒนาโมเดล MMixIRT ที่ประยุกต์จากโมเดล 2PLE (MMix2PLE) และวิธีการประมาณค่าที่เหมาะสม
2. เพื่อเปรียบเทียบความถูกต้องของการตรวจสอบจำนวนกลุ่มแฝงระหว่าง AIC กับ BIC สำหรับโมเดล MMix3PL และ MMix2PLE
3. เพื่อเปรียบเทียบความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง ความสามารถเฉลี่ย และพารามิเตอร์ของข้อสอบ ระหว่างโมเดล MMix3PL กับ MMix2PLE

สมมติฐานการวิจัย

1. เกณฑ์สารสนเทศ BIC มีความถูกต้องของการตรวจสอบจำนวนกลุ่มแฝงมากกว่า AIC ทั้งสำหรับโมเดล MMix3PL และ MMix2PLE
2. โมเดล MMix2PLE มีความแม่นยำของการประมาณค่าพารามิเตอร์สูงกว่าโมเดล MMix3PL และการเพิ่มขนาดตัวอย่างน่าจะช่วยให้ความแม่นยำของทั้งสองโมเดลสูงขึ้น

กรอบแนวคิดในการวิจัย

กรอบแนวคิดในการวิจัยประกอบด้วย 2 ส่วน ได้แก่ 1) การออกแบบสถานการณ์จำลอง เป็นการกำหนดเงื่อนไขสำหรับจำลองข้อมูลพหุระดับที่มีความวิวิธพันธุ์และมีการเดาเข้ามาเกี่ยวข้อง โดยดำเนินการจำลองข้อมูลด้วยโปรแกรม R และ 2) การวิเคราะห์ข้อมูล เป็นการนำข้อมูลจำลองมาวิเคราะห์ด้วยโมเดลแข่งขันที่มีปรัชญาความเชื่อการเดาและจำนวนกลุ่มแฝงแตกต่างกัน เพื่อศึกษาผลตามวัตถุประสงค์การวิจัย โดยดำเนินการวิเคราะห์ตามแนวทางแบบเบย์ส์ (Bayesian approach) ด้วยโปรแกรม MultiBUGS ดังแสดงในภาพ 1



ภาพ 1 กรอบแนวคิดในการวิจัย

วิธีดำเนินการวิจัย

งานวิจัยนี้ใช้ระเบียบวิธีวิจัยเชิงทดลองด้วยการจำลองสถานการณ์แบบมอนติคาร์โล แผนการวิจัย ประกอบด้วย 3 ส่วน ได้แก่ 1) การออกแบบสถานการณ์จำลอง 2) การวิเคราะห์และการประมาณค่าพารามิเตอร์ และ 3) การพิจารณาความถูกต้องและแม่นยำ รายละเอียด ดังนี้

1. การออกแบบสถานการณ์จำลอง

1.1 โมเดล กำหนดโมเดลประชากร (population model) เพื่อจำลองข้อมูลพหุระดับที่มีความวิวิธพันธุ์และคำนึงถึงการเดาคือ MMix3PL และ MMix2PLE ดังสมการ (1) - (2)

โมเดล MMix3PL

$$P(y_{ijs} = 1 | \theta_{jskg}) = \sum_{g=1}^G \sum_{k=1}^K \pi_g \cdot \pi_{k|g} \left(c_{ikg} + (1 - c_{ikg}) \frac{1}{1 + \exp[-a_{ikg} (\theta_{jskg} - b_{ikg})]} \right) \quad (1)$$

โมเดล MMix2PLE

$$P(y_{ijs} = 1 | \theta_{jskg}) = \sum_{g=1}^G \sum_{k=1}^K \pi_g \cdot \pi_{k|g} \left(g_{ijskg} + (1 - g_{ijskg}) \frac{1}{1 + \exp[-a_{ikg} (\theta_{jskg} - b_{ikg})]} \right) \quad (2)$$

เมื่อ π_g คือสัดส่วนของสมาชิกในกลุ่มแฝงระดับโรงเรียน g ; $\pi_{k|g}$ คือสัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียน k ในกลุ่มแฝงระดับโรงเรียน g ; θ_{jskg} คือพารามิเตอร์ความสามารถของนักเรียน j ในโรงเรียน s ซึ่งอยู่ในกลุ่มแฝงระดับนักเรียน k และระดับโรงเรียน g เมื่อ $\theta_{jskg} \sim N(\mu_{kg}, \sigma_{kg}^2)$; μ_{kg} และ σ_{kg}^2 คือค่าเฉลี่ยและความแปรปรวนของความสามารถในกลุ่มแฝงระดับนักเรียนและระดับโรงเรียน; a_{ikg} , b_{ikg} และ c_{ikg} คือพารามิเตอร์อำนาจจำแนก ความยาก และการเดาของข้อสอบในกลุ่มแฝงระดับนักเรียนและระดับโรงเรียน; $g_{ijskg} = (1/m_i)^{1/(1+\exp[a_{ikg}(\theta_{jskg}-b_{ikg})])}$ คือโอกาสในการเดาถูก (Zhu et al., 2019, pp. 453) ซึ่งมีค่าผันแปรตามความสามารถของผู้สอบ อำนาจจำแนก ความยาก และจำนวนตัวเลือกของข้อสอบ (m)

1.2 จำนวนกลุ่มแฝง จำลองข้อมูลเป็นข้อมูลที่มี 2 ระดับ ได้แก่ ระดับกลุ่ม (ระดับโรงเรียน) และระดับหน่วย (ระดับนักเรียน) โดยกำหนดให้กลุ่มแฝงระดับโรงเรียน (G) 2 กลุ่ม และระดับนักเรียน (K) 2 กลุ่ม ซึ่งเป็นสถานการณ์ขั้นพื้นฐานสำหรับการศึกษาดูด้วยโมเดล MMixIRT

1.3 สัดส่วนของสมาชิกในกลุ่มแฝง กำหนดค่าสัดส่วนในระดับโรงเรียนทั้งสองกลุ่มเท่ากันคือ .50 โดยในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 ($g = 1$) สมมติให้กลุ่มแฝงระดับนักเรียนกลุ่มแรก ($k = 1$) เป็นกลุ่มเด่นที่มีค่าสัดส่วนเป็น .90 ส่วนในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 2 ($g = 2$) สมมติให้กลุ่มแฝงระดับนักเรียนกลุ่มที่สอง ($k = 2$) เป็นกลุ่มเด่นที่มีค่าสัดส่วนเป็น .70 ดังตาราง 1

ตาราง 1 ค่าจำลองสัดส่วนของสมาชิกในกลุ่มแฝง

สัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียน	สัดส่วนของสมาชิกในกลุ่มแฝงระดับโรงเรียน	
	$P(g = 1) = \pi_1 = .50$	$P(g = 2) = \pi_2 = .50$
$P(k = 1)$	$\pi_{1 1} = .90$	$\pi_{1 2} = .30$
$P(k = 2)$	$\pi_{2 1} = .10$	$\pi_{2 2} = .70$

1.4 ขนาดตัวอย่าง มี 2 ขนาดคือ 1,250 และ 2,500 คน คำนวณมาจากผลคูณระหว่างขนาดตัวอย่างของโรงเรียน 1 ระดับคือ 50 โรงเรียน กับขนาดตัวอย่างของนักเรียน 2 ระดับคือ 25 และ 50 คน ทั้งนี้เหตุผลในการเลือกขนาดตัวอย่างของโรงเรียนเป็น 50 เนื่องจากจำนวนดังกล่าวเป็นที่นิยมสำหรับการวิจัยเชิงทดลองทั่วไป ส่วนการเลือกขนาดตัวอย่างของนักเรียนเป็น 25 และ 50 เนื่องจากโดยทั่วไปพบว่าโรงเรียนแต่ละแห่งมีนักเรียนเข้าร่วมการทดสอบระดับชาติอย่างน้อย 25 คนโดยประมาณ (Cho, 2007, p. 1) และ 50 คือจำนวนที่ใช้ศึกษาเมื่อมีนักเรียนมากขึ้น

1.5 พารามิเตอร์ความสามารถ จำลองมาจากตัวแปรสุ่ม $\theta_{jskg} \sim N(0, 1)$

1.6 ความยาวแบบสอบ กำหนดเป็น 40 ข้อ เพื่อให้ใกล้เคียงกับแบบสอบมาตรฐานที่ใช้ในการทดสอบระดับชาติ เช่น ข้อสอบ O-NET โดยสมมติให้มีการวัดสาระการเรียนรู้ 4 ด้าน แต่ละด้านวัดด้วยข้อสอบ 10 ข้อ

1.7 พารามิเตอร์ของข้อสอบ กำหนดระดับสมรรถภาพของผู้สอบในกลุ่มแฝงสำหรับจำลองค่าพารามิเตอร์ของข้อสอบดังตาราง 2 และค่าจำลองพารามิเตอร์ของข้อสอบที่สมนัยกับระดับสมรรถภาพดังกล่าวดังตาราง 3 ซึ่งการกำหนดค่าดังกล่าวมีลักษณะใกล้เคียงกับงานวิจัยก่อนหน้านี้คือ Li et al. (2009, p. 359) และ Cho (2007, p. 84)

ตาราง 2 การจำลองรูปแบบสมรรถภาพในกลุ่มแฝง

สาระ	กลุ่มแฝงโรงเรียนที่ 1 ($g = 1$)				กลุ่มแฝงระดับโรงเรียนที่ 2 ($g = 2$)			
	กลุ่มแฝงนักเรียนที่ 1 ($k = 1$)		กลุ่มแฝงนักเรียนที่ 2 ($k = 2$)		กลุ่มแฝงนักเรียนที่ 1 ($k = 1$)		กลุ่มแฝงนักเรียนที่ 2 ($k = 2$)	
1	สูงมาก		ต่ำ		ปานกลาง		ต่ำ	
2	สูง		ต่ำมาก		ต่ำ		ต่ำมาก	
3	ต่ำ		สูง		ปานกลาง		สูง	
4	ต่ำมาก		สูงมาก		ต่ำ		ต่ำมาก/สูงมาก	

ตาราง 3 ค่าจำลองพารามิเตอร์ของข้อสอบในกลุ่มแฝง

สาระ	ข้อ	$k = 1, g = 1$			$k = 2, g = 1$			$k = 1, g = 2$			$k = 2, g = 2$		
		b_{11}	a_{11}	c_{11}	b_{21}	a_{21}	c_{21}	b_{12}	a_{12}	c_{12}	b_{22}	a_{22}	c_{22}
1	1	-2.50	2	0.10	0.60	1	0.20	-0.50	1	0.20	0.60	1	0.20
	2	-2.40	2	0.10	0.70	1	0.20	-0.40	1	0.20	0.70	1	0.20
	3	-2.30	2	0.10	0.80	1	0.20	-0.30	1	0.20	0.80	1	0.20
	4	-2.20	2	0.10	0.90	1	0.20	-0.20	1	0.20	0.90	1	0.20
	5	-2.10	2	0.10	1.00	1	0.25	-0.10	1	0.20	1.00	1	0.25
	6	-2.00	2	0.10	1.10	1	0.25	0.00	1	0.20	1.10	1	0.25
	7	-1.90	2	0.10	1.20	1	0.25	0.10	1	0.20	1.20	1	0.25
	8	-1.80	2	0.10	1.30	1	0.25	0.20	1	0.20	1.30	1	0.25
	9	-1.70	2	0.10	1.40	1	0.25	0.30	1	0.20	1.40	1	0.25
	10	-1.60	2	0.10	1.50	1	0.25	0.40	1	0.20	1.50	1	0.25
2	11	-1.50	2	0.10	1.60	1	0.25	0.50	1	0.20	1.60	1	0.25
	12	-1.40	2	0.10	1.70	1	0.25	0.60	1	0.20	1.70	1	0.25
	13	-1.30	2	0.10	1.80	1	0.25	0.70	1	0.20	1.80	1	0.25
	14	-1.20	2	0.10	1.90	1	0.25	0.80	1	0.20	1.90	1	0.25
	15	-1.10	2	0.10	2.00	1	0.25	0.90	1	0.20	2.00	1	0.25
	16	-1.00	2	0.10	2.10	1	0.25	1.00	1	0.25	2.10	1	0.25
	17	-0.90	1	0.20	2.20	1	0.25	1.10	1	0.25	2.20	1	0.25
	18	-0.80	1	0.20	2.30	1	0.25	1.20	1	0.25	2.30	1	0.25
	19	-0.70	1	0.20	2.40	1	0.25	1.30	1	0.25	2.40	1	0.25
	20	-0.60	1	0.20	2.50	1	0.25	1.40	1	0.25	2.50	1	0.25
3	21	0.60	1	0.20	-1.50	2	0.10	-0.40	1	0.20	-1.50	2	0.10
	22	0.70	1	0.20	-1.40	2	0.10	-0.30	1	0.20	-1.40	2	0.10
	23	0.80	1	0.20	-1.30	2	0.10	-0.20	1	0.20	-1.30	2	0.10
	24	0.90	1	0.20	-1.20	2	0.10	-0.10	1	0.20	-1.20	2	0.10
	25	1.00	1	0.25	-1.10	2	0.10	0.00	1	0.20	-1.10	2	0.10
	26	1.10	1	0.25	-1.00	2	0.10	0.10	1	0.20	-1.00	2	0.10
	27	1.20	1	0.25	-0.90	1	0.20	0.20	1	0.20	-0.90	1	0.20
	28	1.30	1	0.25	-0.80	1	0.20	0.30	1	0.20	-0.80	1	0.20
	29	1.40	1	0.25	-0.70	1	0.20	0.40	1	0.20	-0.70	1	0.20
	30	1.50	1	0.25	-0.60	1	0.20	0.50	1	0.20	-0.60	1	0.20
4	31	1.60	1	0.25	-2.50	2	0.10	0.60	1	0.20	1.50	1	0.25
	32	1.70	1	0.25	-2.40	2	0.10	0.70	1	0.20	1.60	1	0.25
	33	1.80	1	0.25	-2.30	2	0.10	0.80	1	0.20	1.70	1	0.25
	34	1.90	1	0.25	-2.20	2	0.10	0.90	1	0.20	1.80	1	0.25
	35	2.00	1	0.25	-2.10	2	0.10	1.00	1	0.25	1.90	1	0.25
	36	2.10	1	0.25	-2.00	2	0.10	1.10	1	0.25	-2.00	2	0.10
	37	2.20	1	0.25	-1.90	2	0.10	1.20	1	0.25	-1.90	2	0.10
	38	2.30	1	0.25	-1.80	2	0.10	1.30	1	0.25	-1.80	2	0.10
	39	2.40	1	0.25	-1.70	2	0.10	1.40	1	0.25	-1.70	2	0.10
	40	2.50	1	0.25	-1.60	2	0.10	1.50	1	0.25	-1.60	2	0.10

1.8 โมเดลแข่งขัน กำหนดโมเดลเพื่อศึกษาการคัดเลือกโมเดลเป็น 8 โมเดล คือ IRT, multilevel IRT, mixture IRT (MixIRT) ที่มีกลุ่มแฝงระดับนักเรียน (K) เท่ากับ 2 และ 3 กลุ่ม และ Multilevel Mixture IRT (MMixIRT) ที่มีกลุ่มแฝงระดับนักเรียนเท่ากับ 2 และ 3 กลุ่ม ซึ่งซ่อนอยู่ในกลุ่มแฝงระดับโรงเรียน (G) เท่ากับ 2 และ 3 กลุ่ม

จากสถานการณ์จำลองมีเงื่อนไขการศึกษาคือ โมเดลจำลอง 2 เงื่อนไข $\times$ จำนวนกลุ่มแฝง 1 เงื่อนไข $\times$ สัดส่วนสมาชิกในกลุ่มแฝง 1 เงื่อนไข $\times$ ขนาดตัวอย่าง 2 เงื่อนไข $\times$ ความยาวแบบสอบ 1 เงื่อนไข $\times$ โมเดลแข่งขัน 8 เงื่อนไข รวมทั้งสิ้น 32 เงื่อนไข แต่ละเงื่อนไขจะถูกจำลองซ้ำ 10 รอบ ทำให้มีจำนวนชุดข้อมูลที่ต้องวิเคราะห์รวม 320 ชุด

2. การวิเคราะห์และการประมาณค่าพารามิเตอร์

2.1 การกำหนดค่าพารามิเตอร์เพื่อประมาณค่าแบบเบสด้วย MultiBUGS จากโมเดล MMix3PL และ MMix2PLE ในสมการ (1) และ (2) โครงสร้างการแจกแจงความสามารถของผู้สอบคือ $\theta_{jskg} \sim N(\mu_{kg}, \sigma_{kg}^2)$ อย่างไรก็ตามในงานวิจัยนี้จะแปลงพารามิเตอร์ดังกล่าวให้มีความแปรปรวนเป็น 1 โดยกำหนดให้ $\eta_{jskg} = \theta_{jskg} / \sigma_{kg} \sim N(\mu_{kg}, 1)$ เมื่อ $\theta_{jskg} = \sigma_{kg} \cdot \eta_{jskg}$ และ $m_i = 5$ สำหรับโมเดล MMix2PLE เพื่อให้สอดคล้องกับจำนวนตัวเลือกของข้อสอบมาตรฐาน และกำหนดการแจกแจงก่อนหน้าและการแจกแจงภายหลังในลักษณะเดียวกับงานของ Cho (2010, pp. 339-340) และ Choi (2014, p. 40) ดังนี้

การแจกแจงก่อนหน้า (prior distribution) ของพารามิเตอร์ในโมเดล

ผู้สอบ: $\eta_{jskg} \sim N(\mu_{kg}, 1)$; $\mu_{kg} \sim N(0, 1)$; $\mu_{11} = 0$; $\sigma_{kg} \sim N(0, 1) T(0,)$

ข้อสอบ: $a_{ikg} \sim N(0, 1) T(0,)$; $b_{ikg} \sim N(0, 1)$; $c_{ikg} \sim Beta(5, 17)$

กลุ่มแฝง: $g \sim \text{Mult}(1, \pi_g); (\pi_1, \dots, \pi_G) \sim \text{Dir}(1, \dots, 1);$

$k \sim \text{Mult}(1, \pi_{k|g}); (\pi_{1|1}, \dots, \pi_{K|G}) \sim \text{Dir}(1, \dots, 1)$

ความน่าจะเป็นของการแจกแจงภายหลัง (posterior distribution) แสดงดัง (3)

$$P(S|U) \propto L(k, g, \theta_{jskg}) P(\eta_{jskg} | \mu_{kg}) P(\mu_{kg}) P(\sigma_{kg}) P(a_{ikg}) P(b_{ikg}) P(c_{ikg}) P(g | \pi_g) P(\pi_g) P(k | \pi_{k|g}) P(\pi_{k|g}) \quad (3)$$

เมื่อ $L(k, g, \theta_{jskg})$ คือ ฟังก์ชันภาวะควรจะเป็นของโมเดล MMixRT (Cho & Cohen, 2010, p. 345)

และ $S = \{\eta_{jskg}, \mu_{kg}, \sigma_{kg}, a_{ikg}, b_{ikg}, c_{ikg}, g, k, \pi_g, \pi_{k|g}\}$ คือเซตของพารามิเตอร์ที่สนใจ

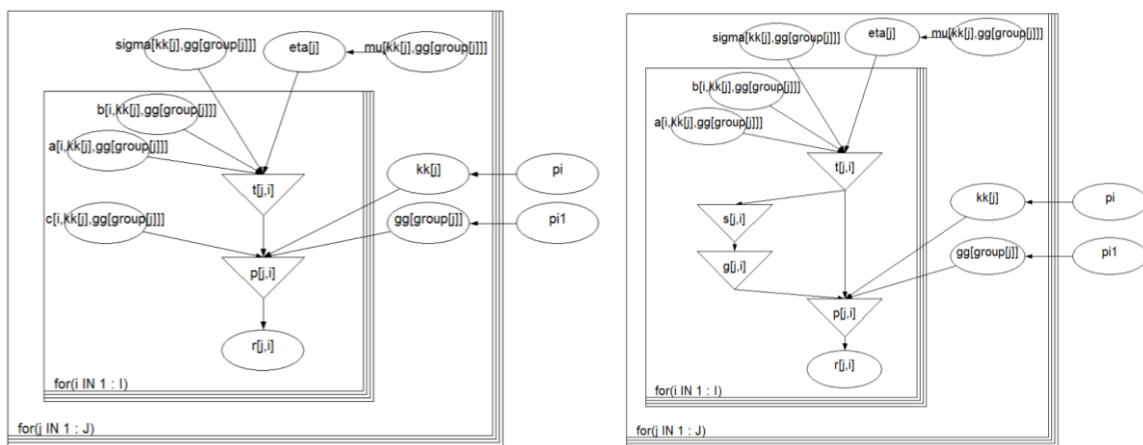
การประมาณค่าแบบเบสใน MultiBUGS

จากโมเดล MMix3PL ในสมการ (1) ถ้ากำหนด $t_{ij} = 1 / (1 + \exp[-a_{ikg}(\theta_{jskg} - b_{ikg})])$ อาจเขียนความสัมพันธ์ในรูปแบบเชิงเส้นได้เป็น $\text{logit}(t_{ij}) = a_{ikg}(\theta_{jskg} - b_{ikg})$ ทั้งนี้ s, k และ g จะไม่นำมาพิจารณาเป็นดัชนีสำหรับ t_{ij} เนื่องจากโดยทั่วไปนักเรียนแต่ละคนจะสังกัดอยู่ในโรงเรียนเพียงแห่งเดียว รวมถึงจะอยู่ในกลุ่มแฝงระดับนักเรียนหรือระดับโรงเรียนได้เพียงกลุ่มเดียวเช่นกัน การใช้ดัชนี j จึงเพียงพอที่จะอธิบายแทน s, k และ g ได้ ดังนั้นโมเดล MMix3PL ในรูปแบบเชิงเส้นจึงเขียนได้เป็น $p_{ij} = \sum_{g=1}^G \sum_{k=1}^K \pi_g \pi_{k|g} (c_{ikg} + (1 - c_{ikg}) t_{ij})$ ดังแสดงในภาพ 2 (ซ้าย)

ในทำนองเดียวกันสำหรับโมเดล MMix2PLE ในสมการ (2) ถ้ากำหนด $\text{logit}(t_{ij}) = a_{ikg}(\theta_{jskg} - b_{ikg})$;

$\text{logit}(s_{ij}) = -\text{logit}(t_{ij})$ และ $g_{ij} = .20^{s_{ij}}$ โมเดล MMix2PLE ในรูปแบบเชิงเส้นจะเขียนได้เป็น

$$p_{ij} = \sum_{g=1}^G \sum_{k=1}^K \pi_g \pi_{k|g} (g_{ij} + (1 - g_{ij}) t_{ij}) \quad \text{ดังแสดงในภาพ 2 (ขวา)}$$



ภาพ 2 กราฟการประมาณค่าแบบเบสสำหรับโมเดล MMix3PL (ซ้าย) และMMix2PLE (ขวา)

หมายเหตุ: ○ คือ stochastic node หมายความว่าตัวแปรถูกสุ่มมาจากการแจกแจงก่อนหน้าที่ทราบค่า; ▽ คือ logical node หมายความว่าตัวแปรมีความสัมพันธ์กับตัวแปรก่อนหน้าแบบระบุรูปแบบได้; □ คือ การดำเนินการทวนซ้ำ

2.2 การตรวจสอบการสลับสัญลักษณ์ (Label Switching) ในการประมาณค่าแบบเบสมีความเป็นไปได้ว่าพารามิเตอร์กลุ่มแฝงที่ต่างกันอาจมีการแจกแจงภายหลังเหมือนกัน พารามิเตอร์ที่ประมาณได้จึงอาจมีได้เพียงหนึ่งเดียว และอาจเกิดปัญหาการสลับสัญลักษณ์ ซึ่งแบ่งเป็น 2 ประเภทหลัก (Cho, 2010, pp. 343-344) ดังนี้

ประเภทที่ 1 เป็นการสลับสัญลักษณ์จากการทวนซ้ำในลูกโซ่ ส่งผลทำให้การประมาณค่าไม่สามารถระบุได้ (Non-identifiable) อาจมีสาเหตุมาจากความไม่เพียงพอของสารสนเทศที่นำมาใช้ในขั้นตอนวิธีเพื่อจำแนกกลุ่มแฝง การสลับสัญลักษณ์ประเภทนี้สามารถตรวจสอบได้จากกราฟความหนาแน่น (Density Plot) ของการแจกแจงความน่าจะเป็น

ภายหลังของค่าพารามิเตอร์ หากมีการสลับสัญลักษณ์เกิดขึ้น กราฟจะมีลักษณะเป็นพหุฐานนิยม (Multimodal) ซึ่งแสดงถึงการที่ค่าประมาณพารามิเตอร์นั้นไม่ได้มีค่าเพียงหนึ่งเดียว หรือไม่สามารถระบุค่าที่แน่นอนได้

ประเภทที่ 2 เป็นการสลับสัญลักษณ์จากการทวนซ้ำระหว่างลูกโซ่ ทำให้ค่าประมาณพารามิเตอร์กลุ่มแฝงในแต่ละลูกโซ่มีการสลับสัญลักษณ์กัน อาจมีสาเหตุมาจากการกำหนดค่าเริ่มต้นของพารามิเตอร์ในลูกโซ่ที่มีความแตกต่างกัน อย่างไรก็ตาม การสลับสัญลักษณ์ประเภทนี้สามารถจัดการได้ด้วยการนำค่าที่ประมาณได้มาเทียบเคียงกับค่าจริงที่จำลองขึ้น

2.3 การวินิจฉัยการลู่เข้า การวินิจฉัยการลู่เข้าของลูกโซ่ในงานวิจัยนี้ ใช้การพิจารณาด้วยวิธี Gelman and Rubin ซึ่งดำเนินการวิเคราะห์ด้วยโปรแกรม MultiBUGS และอัตตสหสัมพันธ์ (Autocorrelation) โดยในการประมาณค่าด้วยลูกโซ่มาร์คอฟ 2 ลูกโซ่ จะกำหนดให้มีการตัดค่าสังเกต (Burn-in) 7,000 ค่าแรกทิ้ง และนำค่าสังเกตภายหลังการตัด (Post-burn-in) 8,000 ค่ามาประมาณหาค่าเฉลี่ย

3. การพิจารณาความถูกต้องและแม่นยำ

3.1 ความถูกต้องของการตรวจสอบจำนวนกลุ่มแฝง พิจารณาจากค่าร้อยละที่ AIC และ BIC ระบุจำนวนกลุ่มแฝงของข้อมูลได้ถูกต้อง (AIC, BIC มีค่าน้อยที่สุด) จากข้อมูล 10 ชุด ทั้งนี้เกณฑ์สารสนเทศ AIC และ BIC สามารถคำนวณได้จากสูตรดังสมการ (4) และ (5) ตามลำดับ

$$AIC = -2 \log L + 2p \quad (4)$$

$$BIC = -2 \log L + p \cdot \ln(N) \quad (5)$$

เมื่อ p คือ จำนวนพารามิเตอร์; N คือ ขนาดตัวอย่างของข้อมูล; $\log L$ คือ ค่า Log-likelihood ซึ่งคำนวณได้จากค่าเฉลี่ยภายหลังของความเบี่ยงเบน (Deviance) จากขั้นตอนวิธีของลูกโซ่มาร์คอฟ

3.2 ความแม่นยำของการประมาณค่าพารามิเตอร์

พารามิเตอร์กลุ่มแฝง คำนวณจากค่าร้อยละของความใกล้เคียง (p) ระหว่างค่าประมาณกับค่าจริงของสัดส่วนสมาชิกในกลุ่มแฝงจากข้อมูลแต่ละชุด มีสูตรคำนวณดังสมการ (6) หาก p เข้าใกล้ 100 จะหมายความว่า การประมาณค่ามีความแม่นยำสูง

$$p = 100 - \left(\frac{\sum_{r=1}^R \sum_{g=1}^G \sum_{k=1}^K |\hat{\pi}_{k|g,r} - \pi_{k|g,r}|}{\sum_{r=1}^R \sum_{g=1}^G \sum_{k=1}^K \pi_{k|g,r}} \times 100 \right) \quad (6)$$

เมื่อ $\pi_{k|g,r}$ และ $\hat{\pi}_{k|g,r}$ คือ ค่าจริงและค่าประมาณของสัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียน k ในกลุ่มแฝงระดับโรงเรียน g จากข้อมูลจำลองชุดที่ r

พารามิเตอร์ของข้อสอบและความสามารถเฉลี่ย คำนวณจาก bias และ RMSE ซึ่งบอกถึงความคลาดเคลื่อนและรากที่สองของความคลาดเคลื่อนกำลังสองจากข้อมูลแต่ละชุด และค่า corr ซึ่งบอกถึงความสัมพันธ์ระหว่างค่าประมาณกับค่าจริง มีสูตรคำนวณดังสมการ (7) - (9) หาก bias/RMSE เข้าใกล้ 0 และ corr เข้าใกล้ 1 จะหมายความว่า การประมาณค่ามีความแม่นยำสูง

$$\text{bias} = \frac{\sum_{r=1}^R \sum_{g=1}^G \sum_{k=1}^K \sum_{i=1}^I (\hat{\theta}_{ikgr} - \theta_{ikg})}{RGKI} \quad (7)$$

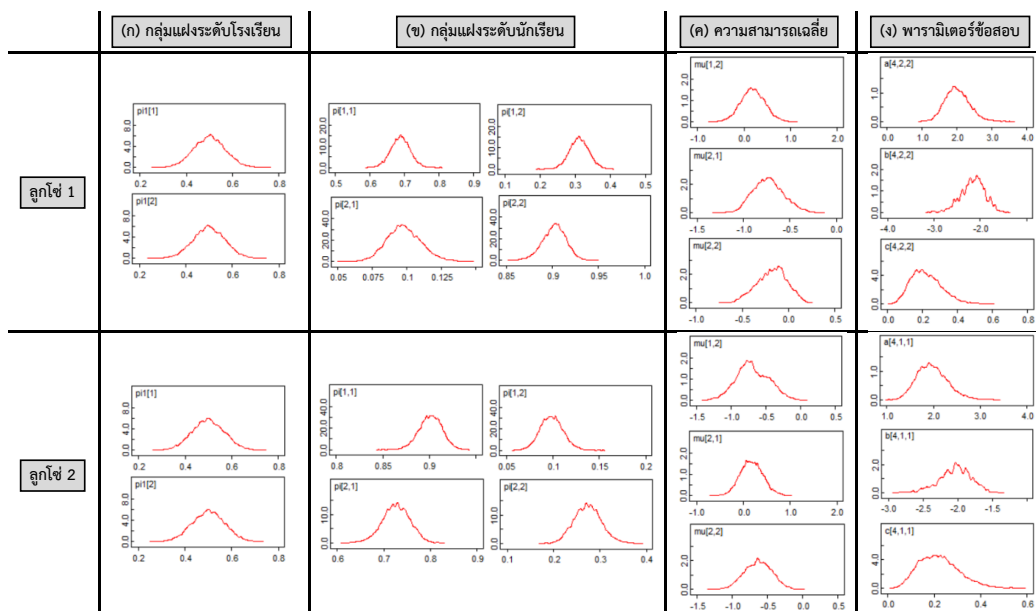
$$\text{RMSE} = \sqrt{\frac{\sum_{r=1}^R \sum_{g=1}^G \sum_{k=1}^K \sum_{i=1}^I (\hat{\theta}_{ikgr} - \theta_{ikg})^2}{RGKI}} \quad (8)$$

$$\text{corr} = \frac{1}{R} \sum_{r=1}^R \frac{\text{Cov}(\hat{\Theta}_{ikgr}, \Theta_{ikg})}{\sigma_{\hat{\Theta}_{ikgr}} \sigma_{\Theta_{ikg}}} \quad (9)$$

เมื่อ Θ_{ikg} และ $\hat{\Theta}_{ikgr}$ คือค่าจริงและค่าประมาณของพารามิเตอร์ของข้อสอบ i ในกลุ่มแฝงระดับนักเรียน k และระดับโรงเรียน g จากข้อมูลจำลองชุดที่ r (ส่วนความสามารถเฉลี่ยก็สามารถคำนวณได้จากสูตรเดียวกัน เพียงแต่จะไม่คำนึงถึงดัชนีของข้อสอบ i) ทั้งนี้ก่อนคำนวณ bias และ RMSE ค่าพารามิเตอร์ที่ประมาณได้จะถูกปรับเทียบให้อยู่บนสเกลเดียวกับค่าพารามิเตอร์จริงก่อน โดยงานวิจัยนี้เลือกใช้การปรับเทียบด้วยวิธีค่าเฉลี่ยและซิกมา (Choi, 2014, pp. 52-53)

ผลการวิจัย

1. การตรวจสอบการสลับสัญลักษณ์ พบว่า กราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์มีแนวโน้มเป็นเอกฐานนิยม (Unimodal) ดังภาพ 3 และมีลักษณะเดียวกันนี้กับทุกเงื่อนไขสำหรับโมเดล MMix3PL และ MMix2PLE ซึ่งแสดงถึงการที่ค่าประมาณพารามิเตอร์สามารถระบุค่าที่แน่นอนได้ จึงอาจอนุมานได้ว่าไม่มีการสลับสัญลักษณ์ประเภทที่ 1 หรือสารสนเทศที่นำมาใช้เพื่อประมาณค่าพารามิเตอร์สำหรับโมเดลในงานวิจัยนี้มีความเพียงพอแล้ว อย่างไรก็ตาม บางเงื่อนไขการศึกษา พบว่า ค่าประมาณพารามิเตอร์กลุ่มแฝงอาจมีการสลับสัญลักษณ์ประเภทที่ 2 สังเกตได้จากค่าประมาณระหว่างลูกโซ่ 1 และ 2 ที่สลับกัน เช่น ลูกโซ่ 1: $\pi_{22} \approx .90$; ลูกโซ่ 2: $\pi_{11} \approx .90$ (ดูภาพ 3ข) และผลดังกล่าวยังทำให้ค่าประมาณความสามารถเฉลี่ยและพารามิเตอร์ของข้อสอบในกลุ่มแฝงมีการสลับสัญลักษณ์เช่นกัน (ดูภาพ 3ค, 3ง) อย่างไรก็ตาม ในงานวิจัยนี้จะแก้ปัญหาดังกล่าวด้วยการนำค่าที่ประมาณได้ไปเทียบเคียงกับค่าจริงที่จำลองขึ้น เพื่อศึกษาถึงความแม่นยำของการประมาณค่าพารามิเตอร์ต่างๆ ต่อไป



ภาพ 3 ตัวอย่างกราฟความหนาแน่นของการแจกแจงภายหลังของค่าพารามิเตอร์

2. ความถูกต้องของการตรวจสอบจำนวนกลุ่มแฝง พบว่า AIC สามารถระบุจำนวนกลุ่มแฝงได้สอดคล้องกับกลุ่มแฝงจริง ($K = 2$ และ $G = 2$) ทั้ง 10 ชุดข้อมูล (ร้อยละ 100) สำหรับทั้งกรณี $N = 1,250$ และ $2,500$ และมีลักษณะ

เดียวกันนี้สำหรับโมเดล MMix3PL และ MMix2PLE ในขณะที่ BIC ไม่สามารถระบุจำนวนกลุ่มแฝงได้ถูกต้องแม้แต่ชุดเดียว ยกเว้นกรณี $N = 2,500$ สำหรับโมเดล MMix2PLE ที่สามารถระบุจำนวนกลุ่มแฝงได้ถูกต้องทุกชุดข้อมูล ดังตาราง 4

ตาราง 4 ค่าเฉลี่ย AIC และ BIC กับความถูกต้องของการตรวจสอบจำนวนกลุ่มแฝง

(ก) โมเดลที่มีปรัชญาการเดาตามแนวคิดโมเดล 3PL					(ข) โมเดลที่มีปรัชญาการเดาตามแนวคิดโมเดล 2PLE				
โมเดล	$N = 1,250$		$N = 2,500$		โมเดล	$N = 1,250$		$N = 2,500$	
	AIC	BIC	AIC	BIC		AIC	BIC	AIC	BIC
3PL	57722 (0.0)	58342 (0.0)	115092 (0.0)	115796 (0.0)	2PLE	53501 (0.0)	53917 (0.0)	106067 (0.0)	106537 (0.0)
multilevel 3PL	58509 (0.0)	59140 (0.0)	116875 (0.0)	117592 (0.0)	multilevel 2PLE	54123 (0.0)	54547 (0.0)	108001 (0.0)	108486 (0.0)
Mix3PL ($K=2$)	52438 (0.0)	53401 (100.0)	104295 (0.0)	105715 (0.0)	Mix2PLE ($K=2$)	46467 (0.0)	47308 (0.0)	92477 (0.0)	93431 (0.0)
Mix3PL ($K=3$)	51964 (0.0)	53846 (0.0)	102986 (0.0)	105122 (100.0)	Mix2PLE ($K=3$)	45279 (0.0)	46548 (100.0)	89966 (0.0)	91406 (0.0)
MMix3PL ($K=2, G=2$)	51797 (100.0)	54312 (0.0)	102431 (100.0)	105283 (0.0)	MMix2PLE ($K=2, G=2$)	45006 (100.0)	46700 (0.0)	89212 (100.0)	91133 (100.0)
MMix3PL ($K=3, G=2$)	52090 (0.0)	55866 (0.0)	102663 (0.0)	107051 (0.0)	MMix2PLE ($K=3, G=2$)	45203 (0.0)	47751 (0.0)	89473 (0.0)	92361 (0.0)
MMix3PL ($K=2, G=3$)	52295 (0.0)	56071 (0.0)	102929 (0.0)	107215 (0.0)	MMix2PLE ($K=2, G=3$)	45354 (0.0)	47900 (0.0)	89579 (0.0)	92467 (0.0)
MMix3PL ($K=3, G=3$)	52828 (0.0)	58498 (0.0)	103401 (0.0)	109834 (0.0)	MMix2PLE ($K=3, G=3$)	45694 (0.0)	49515 (0.0)	89956 (0.0)	94295 (0.0)

หมายเหตุ: ตัวหนาและขีดเส้นใต้ คือ กรณีที่ AIC, BIC ระบุจำนวนกลุ่มแฝงของข้อมูลได้ถูกต้อง; ตัวเลขในวงเล็บ คือ ร้อยละที่โมเดลถูกเลือกจากข้อมูล 10 ชุด

3. ความแม่นยำของการประมาณค่าพารามิเตอร์ ในส่วนนี้จะนำผลการประมาณค่าของโมเดล MMix3PL และ MMix2PLE ที่มีกลุ่มแฝงระดับนักเรียน 2 กลุ่ม และระดับโรงเรียน 2 กลุ่ม มาศึกษา เนื่องจากเป็นโมเดลที่มีจำนวนกลุ่มแฝงสอดคล้องกับข้อมูลจริง รายละเอียดดังนี้

กลุ่มแฝง พบว่า โมเดล MMix3PL และ MMix2PLE มีความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝงระดับโรงเรียน (π_g) สูงร้อยละ 100.0 เท่ากัน ขณะที่กลุ่มแฝงระดับนักเรียน (π_{kg}) มีค่าสูงร้อยละ 96.02 ถึง 97.71 และ 98.92 ถึง 99.25 ตามลำดับ (ดูตาราง 5ก)

ความสามารถเฉลี่ย พบว่า โมเดล MMix2PLE (bias = -.117 ถึง -.084, RMSE = .082 ถึง .097) มีแนวโน้มให้ความแม่นยำในการประมาณค่าพารามิเตอร์ความสามารถเฉลี่ย (μ) สูงกว่าโมเดล MMix3PL (bias = -.193 ถึง -.156, RMSE = .135) (ดูตาราง 5ข)

ข้อสอบ พบว่า โมเดล MMix2PLE (bias = .000, RMSE = .037 ถึง .055, corr = .987 ถึง .991) มีแนวโน้มให้ความแม่นยำในการประมาณค่าความยาก (b) สูงกว่าโมเดล MMix3PL (bias = .000, RMSE = .053 ถึง .082, corr = .980 ถึง .987) ส่วนค่าอำนาจจำแนก (a) พบว่าโมเดล MMix2PLE (bias = -.039 ถึง -.010, RMSE = .078 ถึง .114, corr = .667 ถึง .782) มีแนวโน้มให้ความแม่นยำใกล้เคียงกับโมเดล MMix3PL (bias = -.154 ถึง -.101, RMSE = .077 ถึง .115, corr = .772 ถึง .838) เมื่อขนาดตัวอย่างเป็น 2,500 อย่างไรก็ตามค่าการเดา (c) ในโมเดล MMix3PL พบว่ายังค่อนข้างมีปัญหาในการประมาณค่าย้อนกลับ เนื่องจากค่าประมาณที่ได้ไม่มีความสัมพันธ์กับค่าจริง (corr น้อยกว่า 0) (ดูตาราง 5ข)

ตาราง 5 ความแม่นยำของการประมาณค่าพารามิเตอร์

(ก) พารามิเตอร์กลุ่มแฝงในระดับต่าง ๆ						(ข) ความสามารถเฉลี่ยและพารามิเตอร์ของข้อสอบ						
	N	โมเดล MMix3PL		โมเดล MMix2PLE		N	โมเดล MMix3PL			โมเดล MMix2PLE		
		p (โรงเรียน)	p (นักเรียน)	p (โรงเรียน)	p (นักเรียน)		bias	RMSE	corr	bias	RMSE	corr
π	1,250	100	96.02	100	98.92	μ	-0.156	0.135	-	-0.117	0.097	-
	2,500	100	97.71	100	99.25		-0.193	0.135	-	-0.084	0.082	-
α	1,250					β	-0.154	0.115	.772	-0.039	0.114	.667
	2,500						-0.101	0.077	.838	-0.010	0.078	.782
β	1,250					γ	0.000	0.082	.980	0.000	0.055	.987
	2,500						0.000	0.053	.987	0.000	0.037	.991
γ	1,250					δ	0.027	0.007	-.290	-	-	-
	2,500						0.025	0.006	-.186	-	-	-

สรุปและอภิปรายผล

1. ขั้นตอนวิธีการประมาณค่าพารามิเตอร์แบบเบสสำหรับโมเดล MMix2PLE ที่พัฒนาในงานวิจัยนี้พบว่าค่อนข้างมีความเหมาะสม สังเกตได้จากกราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์ที่มีลักษณะเป็นเอกฐานนิยม ซึ่งแสดงถึงค่าพารามิเตอร์ที่ประมาณได้ว่ามีเพียงหนึ่งเดียวหรือระบุค่าที่แน่นอนได้ ประกอบกับการพิจารณาการประมาณค่าพารามิเตอร์ย้อนกลับก็พบว่าสามารถทำได้ค่อนข้างดีภายใต้เงื่อนไขการศึกษาเชิงจำลองเช่นกัน ซึ่งผลดังกล่าวสอดคล้องกับงานวิจัยของ Cho and Cohen (2010, p. 338) ที่ศึกษาโมเดล MMixRasch และใช้ขั้นตอนวิธีการประมาณค่าแบบเบสเช่นกัน ดังพิจารณาได้จากความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝงสำหรับโมเดล MMix2PLE ที่มีค่าร้อยละ 96.0 ถึง 100.0 ซึ่งใกล้เคียงกับโมเดล MMixRasch ที่มีค่าร้อยละ 98.0 ถึง 100.0 และความแม่นยำของการประมาณค่าพารามิเตอร์ความยากสำหรับโมเดล MMix2PLE ที่มี bias = .000, RMSE = .037 ถึง .055 และ corr = .987 ถึง .991 ซึ่งใกล้เคียงกับโมเดล MMixRasch ที่มี bias = -.001 ถึง .033, RMSE = .094 ถึง .115 และ corr = .997 (Cho & Cohen, 2010, p. 352)

2. การเปรียบเทียบความถูกต้องของการตรวจสอบจำนวนกลุ่มแฝง พบว่า AIC มีแนวโน้มมีความถูกต้องมากกว่า BIC ซึ่งผลดังกล่าวขัดแย้งกับสมมติฐานที่ตั้งไว้ ทั้งนี้ อาจเนื่องจากสมมติฐานดังกล่าวเป็นการอิงผลมาจากการวิจัยก่อนหน้านี้ที่ศึกษาโมเดล MMixRasch (Cho & Cohen, 2010, p. 351) ซึ่งไม่มีการคำนึงถึงการเดา อย่างไรก็ตามเมื่อพิจารณางานวิจัยของ Choi (2014, p. 57) ที่ศึกษาในบริบทโมเดล MixIRT จะพบว่า BIC มีแนวโน้มมีความถูกต้องมากกว่า หากเป็นโมเดล MixRasch หรือ Mix2PL ในขณะที่ AIC จะมีแนวโน้มมีความถูกต้องมากกว่า หากเป็นโมเดล Mix3PL ซึ่งผลดังกล่าวน่าจะอนุมานถึงโมเดล MMix3PL และ MMix2PLE ในงานวิจัยนี้ได้เช่นกัน

3. การเปรียบเทียบความแม่นยำของการประมาณค่าพารามิเตอร์ พบว่าโมเดล MMix2PLE มีแนวโน้มมีความแม่นยำสูงกว่า MMix3PL และการเพิ่มขนาดตัวอย่างยังมีส่วนช่วยให้ความแม่นยำสูงขึ้นด้วย ซึ่งผลดังกล่าวค่อนข้างสอดคล้องกับสมมติฐานที่ตั้งไว้ ทั้งนี้ อาจเนื่องมาจากการนิยามให้โอกาสในการเดาถูก (g) ในโมเดล MMix2PLE ว่าไม่ใช่พารามิเตอร์ที่ต้องประมาณค่า แตกต่างจาก c ในโมเดล MMix3PL ซึ่งเป็นพารามิเตอร์ของข้อสอบที่ต้องประมาณค่า ดังนั้นในแง่ของการประมาณค่า โมเดล MMix3PL จึงมีความซับซ้อนมากกว่าโมเดล MMix2PLE และอาจส่งผลทำให้ความแม่นยำของการประมาณค่าพารามิเตอร์มีแนวโน้มต่ำกว่า (Sen, 2014, pp. 56-57) ส่วนการประมาณค่า c ในโมเดล MMix3PL ที่มีค่าสหสัมพันธ์เป็นลบ พบว่าผลดังกล่าวมีความสอดคล้องกับงานวิจัยของ Choi (2014, p. 76) ซึ่งศึกษาโมเดล Mix3PL และได้

ค่าสหสัมพันธ์เป็นลบเช่นกัน ทั้งนี้อาจเป็นเพราะในทางทฤษฎี C คือ ความน่าจะเป็นที่ผู้สอบความสามารถต่ำมากจะตอบข้อสอบได้ถูกต้อง แต่โดยธรรมชาติของข้อมูล ผู้สอบที่มีคุณลักษณะดังกล่าวมักมีอยู่น้อยมาก การประมาณค่าโดยอิงจากข้อมูลที่มีอยู่น้อยจึงมักมีแนวโน้มไม่สัมพันธ์กับค่าจริง (Lee & Bolt, 2018, p. 91)

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำไปใช้

1.1 นักวัดและประเมินผลได้แนวทางในการตัดสินใจเลือกใช้โมเดล MMixIRT กับข้อมูลทางการศึกษา ซึ่งเป็นข้อมูลพหุระดับที่มีความวิวิธพันธุ์ในประชากรและมีการเดาเข้ามาเกี่ยวข้อง

1.2 หน่วยงานทางการศึกษา เช่น สทศ. สพฐ. สามารถนำโมเดลดังกล่าวไปใช้ เพื่อหากลุ่มแฝงระดับโรงเรียนหรือนักเรียน ซึ่งอาจได้สารสนเทศที่สะท้อนถึงข้อบกพร่องของหลักสูตรของโรงเรียนแต่ละกลุ่ม อันจะนำไปสู่การวินิจฉัยเพื่อวางแผนให้นักเรียนมีผลสัมฤทธิ์ทางการเรียนที่สูงขึ้น

1.3 โมเดลในงานวิจัยนี้ถูกพัฒนาด้วยโปรแกรม MultiBUGS เวอร์ชัน 1.0.0 อย่างไรก็ตามโปรแกรมในเวอร์ชันดังกล่าวอาจมีข้อจำกัดเกี่ยวกับขนาดหน่วยความจำในการรองรับค่าพารามิเตอร์ที่ต้องการประมาณค่าหรือจุดต่อ (Node) ในขั้นตอนวิธีการประมาณค่าแบบเบย์ ผู้ใช้งานจึงควรตรวจสอบความเหมาะสมระหว่างเวอร์ชันของโปรแกรมกับจำนวนพารามิเตอร์ในโมเดล รวมถึงขนาดตัวอย่างของข้อมูลที่เหมาะสมก่อนวิเคราะห์ประมาณค่า

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 การวิจัยนี้เป็นต้นแบบของการบูรณาการระหว่างโมเดล IRT ทางเลือกที่คำนึงถึงการเดากับโมเดล MMixIRT ซึ่งมีการวัดเป็นแบบเอกมิติ (Unidimensionality) เท่านั้น ในการทำวิจัยครั้งต่อไปจึงควรมีการพัฒนาโมเดลการวัดเป็นแบบพหุมิติ ซึ่งอาจช่วยให้ผลการวัดมีความสอดคล้องกับข้อสอบในปัจจุบันที่มีการวัดความสามารถหลายองค์ประกอบมากขึ้น

2.2 การวิจัยครั้งต่อไปอาจนำแนวคิดไปบูรณาการกับโมเดล IRT ที่มีการตรวจให้คะแนนหลายค่า (Polytomous Scoring) ซึ่งน่าจะนำไปประยุกต์ใช้กับแบบวัดทางจิตวิทยาได้

References

- Cho, S. J., & Cohen, A. S. (2010). A multilevel mixture model with applications to DIF. *Journal of Educational and Behavioral Statistics*, 35(3), 336-370.
- Cho, S. J. (2007). *A multilevel mixture IRT model for DIF analysis (Unpublished doctoral dissertation)*. Athens, GA: University of Georgia.
- Choi, Y. (2014). *Metric identification in mixture IRT models (Unpublished doctoral dissertation)*. Athens, GA: University of Georgia.
- Kanjanawasee, S. (2012). *Modern test theories* (4th ed.). Bangkok: Chulalongkorn University Press. [in Thai]
- Lee, S., & Bolt, D. (2018). An alternative to the 3PL: Using asymmetric item characteristic curves to address guessing effects. *Journal of Educational Measurement*, 55(1), 90-111.
- Li, F., Cohen, A., Kim, S., & Cho, S. J. (2009). Model selection methods for mixture dichotomous IRT models. *Applied Psychological Measurement*, 33(5), 353-373.
- Maier, K. S. (2001). A Rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics*, 26(3), 307-330.

- Sen, S. (2014). *Robustness of mixture IRT models to violations of latent normality (Unpublished doctoral dissertation)*. Athens, GA: University of Georgia.
- Tay, L., Diener, E., Drasgow, F., & Vermunt, J. (2011). Multilevel mixed-measurement IRT analysis: An explication and application to self-reported emotions across the world. *Organizational Research Methods*, 14(1), 177-207.
- Zhu, Z., Wang, C., & Tao, J. (2019). A two-parameter logistic extension model: an efficient variant of the three-parameter logistic model. *Applied Psychological Measurement*, 43(6), 449-463.