

การวิเคราะห์โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพระดัดด้วย
โมเดลทฤษฎีการตอบสนองข้อสอบทางเลือกที่คำนึงถึงการเดา



นายธีรุตม์ สุขสกุลวัฒน์

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต

สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2563

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

An Analysis of Multilevel Mixture Item Response Theory Model using
Alternative IRT Model with Guessing



A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Educational Measurement and Evaluation

Department of Educational Research and Psychology

FACULTY OF EDUCATION

Chulalongkorn University

Academic Year 2020

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การวิเคราะห์โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสม พหุระดับด้วยโมเดลทฤษฎีการตอบสนองข้อสอบทางเลือกที่ คำนึงถึงการเดา
โดย	นายธีรุตม์ สุขสกุลวัฒน์
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	อาจารย์ ดร.สิวะโชติ ศรีสุทธิยากร

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัย เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต

..... คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.อวยพร เรืองตระกูล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(อาจารย์ ดร.สิวะโชติ ศรีสุทธิยากร)

..... กรรมการ
(ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

..... กรรมการ
(อาจารย์ ดร.ชยุตม์ ภิรมย์สมบัติ)

..... กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.สังวรณ์ ังคกระโทก)

ธีรุตม์ สุขสกุลวัฒน์: การวิเคราะห์โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับด้วยโมเดลทฤษฎีการตอบสนองข้อสอบทางเลือกที่คำนึงถึงการเดา. (An Analysis of Multilevel Mixture Item Response Theory Model using Alternative IRT model with Guessing) อ.ที่ปรึกษาหลัก: รศ.ดร.ณัฐภรณ์ หลาวทอง, อ.ที่ปรึกษาร่วม: อ.ดร.สวitez โขติ ศรีสุธธยากร

การวิจัยนี้มีวัตถุประสงค์ 4 ประการ คือ 1) เพื่อศึกษารูปแบบของโมเดล MMix2PLE และวิธีการประมาณค่าพารามิเตอร์ 2) เพื่อเปรียบเทียบความถูกต้องของการคัดเลือกโมเดลระหว่างเกณฑ์สารสนเทศ AIC, BIC, CAIC และ DBIC สำหรับโมเดล MMix3PL และ MMix2PLE ภายใต้สถานการณ์จำลอง 3) เพื่อศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง ระหว่างโมเดล MMix3PL กับ MMix2PLE ภายใต้สถานการณ์จำลอง และ 4) เพื่อวิเคราะห์หาจำนวนกลุ่มแฝง พารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง จากข้อมูลเชิงประจักษ์ด้วยโมเดล MMix3PL และ MMix2PLE ข้อมูลที่ศึกษาประกอบด้วยข้อมูลจำลองจากวิธีมอนติคาร์โล 32 เจียนไซ (2 โมเดล x 2 ขนาดตัวอย่าง x 8 โมเดลแข่งขัน) แต่ละเจียนไซถูกจำลองซ้ำ 10 รอบ รวมมีข้อมูลทั้งสิ้น 320 ชุด และข้อมูลเชิงประจักษ์จากข้อมูลคะแนนสอบ O-NET วิชาคณิตศาสตร์ ม.6 (2562) ของนักเรียน 2,500 คน จากโรงเรียน 50 แห่ง ใน 4 สังกัด คือ สพฐ. สข. อว. และ กสท. ผลการวิจัยสรุปได้ดังนี้

1. รูปแบบของโมเดล MMix2PLE ที่อธิบายโอกาสในการเดาถูกด้วยคุณลักษณะของผู้สอบและข้อสอบ สามารถช่วยหลีกเลี่ยงปัญหาที่อาจเกิดขึ้นจากการประมาณค่าพารามิเตอร์ c ของโมเดล MMix3PL ได้ ส่วนวิธีการประมาณค่าแบบเบสที่นำมาใช้พบว่ามีความแม่นยำให้ค่าประมาณพารามิเตอร์ได้ค่อนข้างแม่นยำ สัมพันธ์กับค่าจริง ระบุค่าที่แน่นอนได้ และมีความแม่นยำในการประมาณค่าเมื่อขนาดตัวอย่างในกลุ่มแฝงมีขนาดเล็ก

2. การเปรียบเทียบความถูกต้องของการคัดเลือกโมเดล พบว่า AIC มีความแม่นยำมากที่สุด รองลงมาคือ DBIC และต่ำที่สุดคือ BIC และ CAIC ภายใต้เงื่อนไขข้อมูล 2 ระดับ ขนาดตัวอย่างเป็น 1,250 และ 2,500 คน และมีจำนวนกลุ่มแฝงระดับโรงเรียน 2 กลุ่ม และกลุ่มแฝงระดับนักเรียน 2 กลุ่ม

3. ความแม่นยำของการประมาณค่าพารามิเตอร์ พบว่าโมเดล MMix2PLE และ MMix3PL มีความแม่นยำค่อนข้างสูง ยกเว้นพารามิเตอร์ c ของโมเดล MMix3PL ที่ให้ค่าประมาณไม่สัมพันธ์กับค่าจริง ภายใต้สถานการณ์ที่โมเดลวิเคราะห์เป็นโมเดลเดียวกับโมเดลประชากร

4. การวิเคราะห์ข้อมูลเชิงประจักษ์ด้วยโมเดล MMix3PL และ MMix2PLE ให้ผลสรุปเกี่ยวกับจำนวนกลุ่มแฝงและความสามารถสอดคล้องกัน โดยพบว่าข้อมูลมีกลุ่มแฝงระดับโรงเรียน 2 กลุ่ม และกลุ่มแฝงระดับนักเรียน 2 กลุ่ม รวมมีกลุ่มแฝงทั้งสิ้น 4 กลุ่ม ได้แก่ 1) กลุ่มความสามารถสูง) 2) กลุ่มความสามารถต่ำ 3) กลุ่มความสามารถปานกลาง และ 4) กลุ่มความสามารถต่ำมาก ทั้งนี้สมาชิกในกลุ่มแฝงสองกลุ่มแรกส่วนใหญ่เป็นนักเรียนที่มีผู้ปกครองรายได้มากกว่า 300,000 บาท ศึกษาในโรงเรียนสังกัด อว. และอยู่ใน กทม. ขณะที่สมาชิกในกลุ่มหลังส่วนใหญ่เป็นนักเรียนที่มีผู้ปกครองรายได้ไม่เกิน 300,000 บาท ศึกษาในโรงเรียนสังกัด สพฐ. สข. และ กสท. และมีได้อยู่ใน กทม. ส่วนการประมาณค่าพารามิเตอร์ของข้อสอบ พบว่าโมเดลทั้งสองให้ค่าประมาณไม่ค่อยสอดคล้องกัน ซึ่งเป็นผลมาจากการจัดสมาชิกในกลุ่มแฝงที่มีความแตกต่างกัน

สาขาวิชา การวัดและประเมินผลการศึกษา

ลายมือชื่อนิสิต

ปีการศึกษา 2563

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ลายมือชื่อ อ.ที่ปรึกษาร่วม

6084208827 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION

KEYWORD: multilevel mixture IRT model, alternative IRT model with guessing, model selection

Teerut Suksakulwat: An Analysis of Multilevel Mixture Item Response Theory Model using
Alternative IRT Model with Guessing. Advisor: Assoc. Prof. Nuttaporn Lawthong, Ph.D.
Co-advisor: Siwachoat Srisuttiyakorn, Ph.D.

This study had four purposes: 1) to study the form of the MMix2PLE model and its parameter estimation method; 2) to compare the accuracy of model selection between AIC, BIC, CAIC, and DBIC for the MMix3PL and MMix2PLE models under simulated conditions; 3) to study the precision of parameter estimation: classes, items, and mean ability between the MMix3PL and MMix2PLE models under simulated conditions; and 4) to determine the number of classes, latent class parameters, item parameters, and mean ability in classes from empirical data using the MMix3PL and MMix2PLE models. The data consisted of 32 simulated data from the Monte Carlo method (2 models x 2 sample sizes x 8 candidate models), each data was replicated 10 times, yielding a total of 320 data sets, and the empirical data from O-NET in Mathematics Grade 12 (2019), of 2,500 students, 50 schools, 4 affiliations: OBEC, OPEC, MHESI, and DLA. The research results were as follows:

1. The form of the MMix2PLE model, which described the chance of a correct guess with the characteristics of respondents and items can avoid problems may arise from estimation of c parameters of the MMix3PL model. As for the Bayesian estimation tended to provide fairly accurate estimates, correlate to real values, be specifiable, and be robust in parameter estimation when the sample size in the class was small.

2. Comparison of the accuracy of model selection, AIC tended to be most accurate, followed by DBIC, and the lowest being BIC and CAIC under 2 school-level and 2 student-level latent classes conditions.

3. The precision of parameter estimation, the MMix2PLE and MMix3PL model tended to be highly precise, except c parameters for the MMix3PL model which did not correlate with the real values when the analyzed model was the same as the population model.

4. Empirical data analysis using the MMix3PL and MMix2PLE models yielded consistent results regarding latent classes and ability. The data consisted of 2 school-level and 2 student-level latent classes, total of 4 ability groups: 1) high, 2) low, 3) moderate, and 4) very low. Most of the first two groups were students whose parents earn more than 300,000 baht, studied in MHESI schools, Bangkok, while most of the latter two groups were students whose parents earn no more than 300,000 baht, studied in OBEC, OPEC, and DLA schools, not Bangkok. The item parameter estimation found that both models gave inconsistent estimates as a result of different arrangement of members in the latent classes.

Field of Study: Educational Measurement and Student's Signature

Evaluation

Academic Year: 2020 Advisor's Signature

Co-advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์นี้เป็นส่วนสำคัญในการศึกษาระดับดุษฎีบัณฑิต ผู้วิจัยต้องอาศัยความรู้ ประสบการณ์ และความอดทนเพื่อให้สำเร็จเป็นรูปเล่ม อย่างไรก็ตามความสำเร็จครั้งนี้จะเกิดขึ้นไม่ได้ หากปราศจากความช่วยเหลือจากคณะบุคคลดังต่อไปนี้ ผู้วิจัยจึงขอกราบขอบพระคุณมา ณ ที่นี้

กราบขอบพระคุณ รองศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง อาจารย์ที่ปรึกษาหลัก และ อาจารย์ ดร.สิวะโชติ ศรีสุทธิยากร อาจารย์ที่ปรึกษาร่วม ที่คอยดูแลช่วยเหลือ ให้คำปรึกษา ชี้แนะ และ ตรวจสอบแก้ไข จนบรรลุวัตถุประสงค์ตามที่ตั้งไว้ ตลอดระยะเวลาที่ทำวิทยานิพนธ์ฉบับนี้

กราบขอบพระคุณ รองศาสตราจารย์ ดร.อวยพร เรืองตระกูล ประธานกรรมการสอบ วิทยานิพนธ์ ศาสตราจารย์ ดร.ศิริชัย กาญจนवासี อาจารย์ ดร.ชยุตม์ ภิรมย์สมบัติ และ รองศาสตราจารย์ ดร.สังวรณ์ ังตกระโทก กรรมการสอบวิทยานิพนธ์ ที่คอยให้คำชี้แนะที่เป็นประโยชน์ยิ่งต่อ ผู้วิจัยในการปรับแก้งานให้มีความสมบูรณ์ขึ้น

กราบขอบพระคุณอาจารย์ทุกท่านในภาควิชาวิจัยและจิตวิทยาการศึกษาที่คอยประสิทธิ์ ประสาทวิชาความรู้ เพื่อให้สามารถนำไปใช้ในการดำเนินชีวิตต่อไปในอนาคต

กราบขอบพระคุณ ศาสตราจารย์กิตติคุณ ดร.ปราณี กุลละวณิช ที่ปรึกษาสถาบันภาษาไทย สิรินคร ผู้ช่วยศาสตราจารย์ดุष्ฎีพร ชำนิโรคนานต์ อดีตผู้อำนวยการสถาบันภาษาไทยสิรินคร และรอง ศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์ อาจารย์ที่ปรึกษาระดับปริญญาโท ที่ให้ความกรุณาเขียน จดหมายรับรองเพื่อใช้สมัครเข้าศึกษาต่อในระดับปริญญาเอก

กราบขอบพระคุณ สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ที่ให้ความ อนุเคราะห์ข้อมูลคะแนนสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 วิชาคณิตศาสตร์ เพื่อสถิติการใช้โมเดลกับ ข้อมูลเชิงประจักษ์ อันทำให้เห็นถึงประโยชน์ในเชิงปฏิบัติได้ชัดเจนขึ้น

กราบขอบพระคุณ คุณแม่ กัลยา สุขสกุลวัฒน์ คุณป้า พรรณี วัลภา และญาติพี่น้องทุกคน ใน ความรักความอบอุ่นที่มีให้ตลอดเวลาในยามที่อดอยและการสนับสนุนอย่างดีมาโดยตลอด

และที่ขาดไม่ได้คือ ขอขอบคุณเพื่อน ๆ พี่ ๆ น้อง ๆ ทุกคนจากสถาบันภาษาไทยสิรินครและ ภาควิชาวิจัยและจิตวิทยาการศึกษาที่คอยให้กำลังใจและคำแนะนำตลอดมา ทำให้ผู้วิจัยสามารถทำ วิทยานิพนธ์จนสำเร็จลุล่วงออกมาได้

ธีรุตม์ สุขสกุลวัฒน์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ค
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ฌ
สารบัญรูปภาพ.....	ฎ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญ	1
1.2 คำถามวิจัย.....	4
1.3 วัตถุประสงค์การวิจัย.....	5
1.4 สมมติฐานการวิจัย	5
1.5 ขอบเขตการวิจัย	6
1.6 นิยามศัพท์ที่ใช้ในการวิจัย.....	7
1.7 ประโยชน์ที่ได้รับ.....	8
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง	10
ตอนที่ 1 แนวคิดการเดาคำตอบในข้อสอบปรนัย	11
1.1 รูปแบบของข้อสอบปรนัย	11
1.2 การเดาคำตอบในข้อสอบปรนัย	12
1.3 ปัจจัยที่ส่งผลต่อการเดาคำตอบ	13
1.4 แนวทางการป้องกันการเดาคำตอบ.....	15
ตอนที่ 2 แนวคิดของโมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ	18

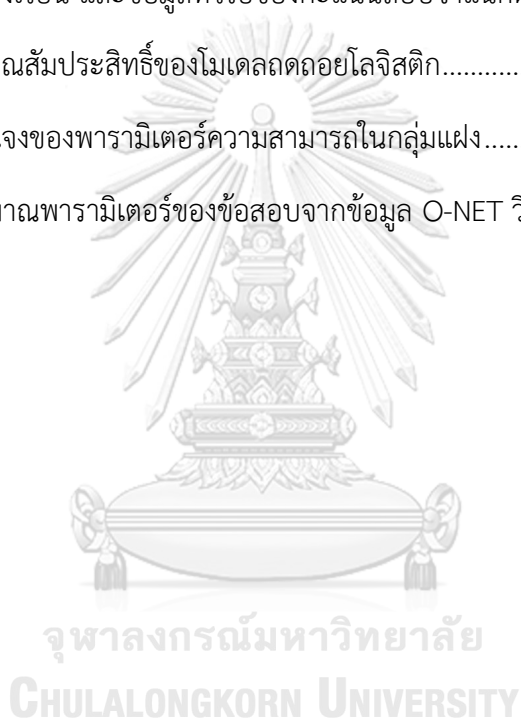
2.1 โมเดลทฤษฎีการตอบสนองข้อสอบ (IRT) และโมเดล IRT ทางเลือกที่คำนึงถึงการเดา..	18
2.2 โมเดลกลุ่มแฝง (LC).....	27
2.3 โมเดลพหุระดับ (multilevel).....	28
2.4 โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสม (MixIRT).....	29
2.5 โมเดลทฤษฎีการตอบสนองข้อสอบพหุระดับ (multilevel IRT)	31
2.6 โมเดลกลุ่มแฝงพหุระดับ (multilevel LC)	33
2.7 โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ (MMixIRT)	33
ตอนที่ 3 วิธีการประมาณค่าพารามิเตอร์.....	42
3.1 วิธีภาวะควรจะเป็นสูงสุด (maximum likelihood)	42
3.2 วิธีการประมาณค่าแบบเบย์ (Bayesian estimation).....	42
ตอนที่ 4 การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo simulation).....	48
ตอนที่ 5 การทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน (O-NET).....	50
ตอนที่ 6 กรอบแนวคิดในการวิจัย.....	54
บทที่ 3 วิธีดำเนินการวิจัย.....	56
ส่วนที่ 1 การศึกษาภายใต้สถานการณ์จำลอง	56
ส่วนที่ 2 การศึกษาจากข้อมูลเชิงประจักษ์	77
บทที่ 4 ผลการวิจัย.....	79
ตอนที่ 1 ผลการศึกษาภายใต้สถานการณ์จำลอง	79
ตอนที่ 2 ผลการศึกษาจากข้อมูลเชิงประจักษ์	81
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	90
5.1 สรุปผลการวิจัย.....	90
5.2 อภิปรายผลการวิจัย	92
5.3 ข้อเสนอแนะ	95
บรรณานุกรม.....	98

ภาคผนวก.....	106
ภาคผนวก ก คำสั่งที่ใช้ในการจำลองข้อมูลแบบมอนติคาร์โลด้วยโปรแกรม R	107
ภาคผนวก ข คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์ด้วยโปรแกรม MultiBUGS	113
ข.1 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล 3PL	113
ข.2 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล 2PLE	114
ข.3 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล multilevel 3PL.....	115
ข.4 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล multilevel 2PLE	116
ข.5 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล Mix3PL	117
ข.6 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล Mix2PLE	119
ข.7 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล MMix3PL	121
ข.8 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล MMix2PLE.....	123
ประวัติผู้เขียน	126

สารบัญตาราง

	หน้า
ตาราง 2.1 การเปรียบเทียบการป้องกันการเดาทั้ง 4 แนวทาง	18
ตาราง 2.2 รูปแบบของโมเดล IRT	19
ตาราง 2.3 การเปรียบเทียบระหว่างโมเดล AG, FG3PL, 2PLG, 3P-RH และ 2PLE	26
ตาราง 2.4 โมเดลการวิเคราะห์ตัวแปรแฝง.....	35
ตาราง 2.5 โครงสร้างสัดส่วนของสมาชิกในกลุ่มแฝงสำหรับโมเดล MMixIRT.....	36
ตาราง 2.6 การเปรียบเทียบโมเดล MMixRasch รูปแบบทั่วไปและรูปแบบเฉพาะ.....	37
ตาราง 2.7 การสังเคราะห์งานวิจัยที่ศึกษาเกี่ยวกับโมเดล MMixIRT	40
ตาราง 2.8 การเปรียบเทียบพารามิเตอร์ของโมเดลและโปรแกรมวิเคราะห์.....	41
ตาราง 2.9 จำนวนข้อสอบส่วนปรนัย O-NET วิชาคณิตศาสตร์ ระดับ ม.6.....	53
ตาราง 2.10 ตัวแปรที่ใช้ในการจำแนกกลุ่มโรงเรียนและนักเรียน	53
ตาราง 3.1 สถานการณ์การศึกษาเมื่อข้อมูลมีจำนวนกลุ่มแฝงแตกต่างกัน	57
ตาราง 3.2 ค่าจำลองสัดส่วนของสมาชิกในกลุ่มแฝง.....	57
ตาราง 3.3 การจำลองรูปแบบสมรรถภาพของผู้สอบในกลุ่มแฝง	59
ตาราง 3.4 ค่าจำลองพารามิเตอร์ของข้อสอบในกลุ่มแฝง.....	59
ตาราง 3.5 ค่าประมาณพารามิเตอร์กลุ่มแฝงสำหรับโมเดล MMix3PL กรณี $N = 2,500$	72
ตาราง 3.6 ค่าประมาณพารามิเตอร์กลุ่มแฝงสำหรับโมเดล MMix2PLE กรณี $N = 2,500$	73
ตาราง 3.7 ค่าประมาณพารามิเตอร์ของข้อสอบสำหรับโมเดล MMix3PL กรณี $N = 2,500$	73
ตาราง 3.8 ค่าประมาณพารามิเตอร์ของข้อสอบสำหรับโมเดล MMix2PLE กรณี $N = 2,500$	74
ตาราง 3.9 การเปรียบเทียบค่าโอกาสในการเดาถูกระหว่างโมเดล MMix3PL กับ MMix2PLE.....	74
ตาราง 3.10 จำนวนตัวอย่างของโรงเรียนและนักเรียน	77
ตาราง 4.1 ค่าเฉลี่ยเกณฑ์สารสนเทศกับความถูกต้องของการคัดเลือกโมเดล.....	80

ตาราง 4.2 ความแม่นยำของการประมาณค่าพารามิเตอร์.....	81
ตาราง 4.3 ข้อมูลคะแนนสอบจำแนกตามสังกัดและขนาดของโรงเรียน.....	82
ตาราง 4.4 การคัดเลือกโมเดลกรณีการศึกษาข้อมูลเชิงประจักษ์.....	84
ตาราง 4.5 ค่าสัดส่วนของสมาชิกในกลุ่มแฝงระดับโรงเรียน (π_g) และระดับนักเรียน ($\pi_{k g}$).....	85
ตาราง 4.6 จำนวนและค่าสัดส่วนของโรงเรียนจำแนกตามตัวแปรและกลุ่มแฝงระดับโรงเรียน	86
ตาราง 4.7 จำนวนและค่าสัดส่วนของนักเรียนจำแนกตามตัวแปรและกลุ่มแฝงระดับนักเรียนซึ่งซ้อนอยู่ในกลุ่มแฝงระดับโรงเรียน และข้อมูลทั่วไปของคะแนนสอบจำแนกตามกลุ่มแฝง	87
ตาราง 4.8 ค่าประมาณสัมประสิทธิ์ของโมเดลถดถอยโลจิสติก.....	87
ตาราง 4.9 การแจกแจงของพารามิเตอร์ความสามารถในกลุ่มแฝง.....	88
ตาราง 4.10 ค่าประมาณพารามิเตอร์ของข้อสอบจากข้อมูล O-NET วิชาคณิตศาสตร์	89



สารบัญรูปภาพ

	หน้า
ภาพ 2.1 การอธิบายการเดาตามแนวคิดของ Hutchinson (1991).....	12
ภาพ 2.2 ปัจจัยที่ส่งผลต่อการเดาคำตอบ	15
ภาพ 2.3 ตัวอย่าง ICCs ของโมเดล Rasch, 2PL และ 3PL	20
ภาพ 2.4 มโนทัศน์ของกระบวนการแก้ปัญหาและกระบวนการเดาในโมเดล AG.....	22
ภาพ 2.5 ตัวอย่าง ICCs ของโมเดล AG.....	22
ภาพ 2.6 มโนทัศน์ของกระบวนการแก้ปัญหาและกระบวนการเดาในโมเดล FG3PL.....	23
ภาพ 2.7 ตัวอย่าง ICCs ของโมเดล 2PLG เมื่อ w (ซ้าย) และ b (ขวา) มีการเปลี่ยนแปลง.....	24
ภาพ 2.8 ตัวอย่าง ICCs ของโมเดล 3P-RH (เส้นทึบ) กับ 2PNO (เส้นประ).....	25
ภาพ 2.9 ตัวอย่าง ICCs ของโมเดล 2PLE (เส้นทึบ) และ 3PL (เส้นประ).....	26
ภาพ 2.10 แผนภาพองค์ประกอบของโมเดล MMixIRT.....	34
ภาพ 2.11 โครงสร้างของโมเดล MMixRasch (บน) และ MMix2PL (ล่าง).....	39
ภาพ 2.12 ลูกโซ่มีแนวโน้มเข้าสู่สถานะคงตัว	44
ภาพ 2.13 ลูกโซ่ควรถูกตัดทิ้งก่อนประมาณค่า.....	44
ภาพ 2.14 ลูกโซ่มีจำนวนน้อยเกินไป	44
ภาพ 2.15 ลูกโซ่มีปัญหาการเข้าสู่สถานะคงตัว	44
ภาพ 2.16 ตัวอย่างการแจกแจงยูนิฟอร์ม.....	48
ภาพ 2.17 ตัวอย่างการแจกแจงปกติตัวแปรเดียว	48
ภาพ 2.18 ตัวอย่างการแจกแจงปกติสองตัวแปร.....	48
ภาพ 2.19 ตัวอย่างการแจกแจงแกมมา	48
ภาพ 2.20 ตัวอย่างการแจกแจงบีตา.....	48
ภาพ 2.21 ตัวอย่างการแจกแจงตรีเคิล.....	48

ภาพ 2.22	ขั้นตอนการจำลองข้อมูล.....	49
ภาพ 2.23	กรอบแนวคิดในการวิจัย	55
ภาพ 3.1	กราฟิกการประมาณแบบเบส์ค่าสำหรับโมเดล 3PL	66
ภาพ 3.2	กราฟิกการประมาณค่าแบบเบส์สำหรับโมเดล 2PLE	66
ภาพ 3.3	กราฟิกการประมาณค่าแบบเบส์สำหรับโมเดล multilevel 3PL	67
ภาพ 3.4	กราฟิกการประมาณค่าแบบเบส์สำหรับโมเดล multilevel 2PLE	67
ภาพ 3.5	กราฟิกการประมาณค่าแบบเบส์สำหรับโมเดล Mix3PL	67
ภาพ 3.6	กราฟิกการประมาณค่าแบบเบส์สำหรับโมเดล Mix2PLE	68
ภาพ 3.7	กราฟิกการประมาณค่าแบบเบส์สำหรับโมเดล MMix3PL	68
ภาพ 3.8	กราฟิกการประมาณค่าแบบเบส์สำหรับโมเดล MMix2PLE.....	68
ภาพ 3.9	กราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์.....	70
ภาพ 3.10	กราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์กลุ่มแฝง เมื่อโมเดลวิเคราะห์ที่มีจำนวนกลุ่มแฝงมากกว่าข้อมูล	71
ภาพ 3.11	การวินิจฉัยการลู่เข้าของลูกโซ่: (ก) โมเดล IRT, (ข) โมเดล multilevel IRT, (ค) โมเดล MixIRT และ (ง) โมเดล MMixIRT	72
ภาพ 4.1	กลุ่มสาระข้อสอบ (ซ้าย) และผลการตรวจสอบมิติของความสามารถที่มุ่งวัด (ขวา)	82
ภาพ 4.2	ผลการวิเคราะห์ด้วยโมเดลพหุระดับ (ซ้าย) และโมเดล multilevel IRT (ขวา).....	83
ภาพ 4.3	กราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์กลุ่มแฝง กรณีการศึกษาข้อมูลเชิงประจักษ์.....	84
ภาพ 4.4	การวินิจฉัยการลู่เข้าของลูกโซ่กรณีการศึกษาข้อมูลเชิงประจักษ์	85
ภาพ 4.5	โปรไฟล์ความยากของข้อสอบกลุ่มสาระพีชคณิตในกลุ่มแฝงระดับนักเรียนซึ่งซ่อนอยู่ใน: (ก) กลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 ($g = 1$) และ (ข) กลุ่มแฝงระดับโรงเรียนกลุ่มที่ 2 ($g = 2$).....	89

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญ

แบบสอบปรนัยแบบหลายตัวเลือก (multiple-choice test) เป็นเครื่องมือที่นิยมใช้อย่างแพร่หลายในการทดสอบเพื่อวัดความถนัดและผลสัมฤทธิ์ทางการศึกษาของผู้เรียน เนื่องจากแบบสอบมีจุดเด่นหลายประการ เช่น วัดความรู้ได้หลายระดับ บริหารจัดการเวลาได้สะดวก ตรวจให้คะแนนได้อย่างเป็นปรนัย ใช้เครื่องจักรตรวจได้อย่างรวดเร็วและแม่นยำ อย่างไรก็ตามแบบสอบดังกล่าวยังมีข้อจำกัดสำคัญ 2 ประการ คือ 1) มักเกิดการเดาคำตอบ (guessing) เมื่อผู้สอบมีความรู้ไม่เพียงพอในการตอบข้อคำถาม ความแปรปรวนของคะแนนความคลาดเคลื่อน (error variance) ที่วัดได้จากแบบสอบจึงมีแนวโน้มมีค่าสูง ส่งผลทำให้ความเที่ยงและความตรงต่ำ (Budescu & Bo, 2015) และ 2) ไม่สามารถวัดผู้ตอบที่มีความรู้บางส่วน (partial knowledge) เนื่องจากการตรวจให้คะแนนโดยทั่วไปมีเพียงสองค่าคือ 0 (ผิด) กับ 1 (ถูก) จึงไม่เพียงพอต่อการจำแนกผู้ตอบกลุ่มดังกล่าวออกจากกลุ่มที่มีความรู้สมบูรณ์ (complete knowledge) และกลุ่มที่ไม่มีความรู้เลย (no knowledge) (Bo, Lewis, & Budescu, 2015)

จากข้อจำกัดที่กล่าวมาข้างต้น การเดาคำตอบถือเป็นสิ่งที่ส่งผลต่อการวัดและประเมินด้วยแบบสอบปรนัยที่มีการตรวจให้คะแนนแบบสองค่ามากที่สุด เนื่องจากการเปิดช่องให้ผู้ตอบที่มีความรู้ไม่เพียงพอมีโอกาสได้คะแนนจากการตอบ คะแนนที่วัดได้จึงอาจไม่สะท้อนถึงความรู้จริงของผู้ตอบ และอาจเกิดข้อผิดพลาดในการนำผลคะแนนไปใช้เพื่อประเมินหรือตัดสิน โดยเฉพาะเรื่องที่มีความสำคัญและไม่มีโอกาสติดตามตรวจสอบได้อีก เช่น การสอบเพื่อสำเร็จการศึกษา การสอบคัดเลือกเพื่อเข้าศึกษาต่อหรือทำงาน (ศิริชัย กาญจนวาสี, 2556)

ในอดีตแม้จะมีการป้องกันการเดาด้วยวิธีต่าง ๆ เช่น การหักคะแนนเมื่อเดาผิด การเปลี่ยนรูปแบบการตอบจากการเลือกตัวถูกเป็นตัวเลข การปรับให้มีตัวถูกหลายตัว (Bo et al., 2015) อย่างไรก็ตามวิธีการดังกล่าวพบว่ายังมีข้อจำกัดในเชิงปฏิบัติหลายประการ เช่น การสอบต้องใช้เวลานานกว่าปกติ การสร้างเครื่องมือมีค่าใช้จ่ายสูง การคำนวณคะแนนมีความซับซ้อน รวมถึงผลการวัดยังอาจเปลี่ยนไปตามกลุ่มผู้สอบ เนื่องจากในสถานการณ์จริงอาจมีผู้สอบหลายประเภท เช่น กลุ่มไม่ชอบเสี่ยง กลุ่มไม่ชอบสูญเสีย กลุ่มอนุรักษ์นิยมการทำข้อสอบแบบดั้งเดิม เป็นต้น ผู้สอบแต่ละกลุ่มจึงอาจถูกลดโทษอย่างไม่เป็นสัดส่วน การประมาณค่าคะแนนจึงมักมีความลำเอียง (Budescu & Bo, 2015) นอกจากนี้ยังมีปัจจัยอื่น ๆ ที่อาจส่งผลต่อการเดาเช่นกัน อาทิ ความยากของข้อสอบ อำนาจจำแนกของข้อสอบ คุณภาพตัวलग ความรู้ของผู้สอบ แรงจูงใจในการสอบ ดังนั้นจึงเป็นเรื่องยากในเชิงปฏิบัติที่จะป้องกันการเดาด้วยวิธีการดังกล่าว

โมเดลทฤษฎีการตอบสนองข้อสอบ (IRT: item response theory) จึงเริ่มเข้ามามีบทบาทในการวัดผลทางการศึกษาและอธิบายการเดาแทนทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory) เนื่องจากโมเดล IRT มีข้อได้เปรียบเหนือกว่าหลายประการ อาทิ ความเป็นอิสระ (local independence) ความไม่แปรเปลี่ยน (invariance) และการเป็นฟังก์ชันต่อเนื่อง (continuous function) (ศิริชัย กาญจนวาสี, 2555) โดยจุดเด่นของโมเดล IRT คือการนำผลการตอบข้อสอบของบุคคลซึ่งเป็นค่าสังเกตได้มาอธิบายเป็นความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบถูกกับคุณลักษณะแฝงของผู้สอบและข้อสอบในรูปแบบฟังก์ชันโลจิสติก ตัวอย่างของโมเดล IRT ที่ใช้กับข้อสอบที่มีการตรวจให้คะแนนแบบสองค่า เช่น โมเดล Rasch โมเดลโลจิส 2 พารามิเตอร์ (2PL: two-parameter logistic) และโมเดลโลจิส 3 พารามิเตอร์ (3PL)

จากคุณสมบัติความไม่แปรเปลี่ยน โมเดล IRT มีสมมติฐานหนึ่งว่าผู้สอบต้องมาจากประชากรที่มีภาวะเอกพันธ์เพียงกลุ่มเดียว (single homogeneous population) เพื่อแสดงนัยว่าโค้งลักษณะข้อสอบ (ICCs: item characteristic curves) เพียงชุดเดียวสามารถอธิบายความสัมพันธ์ระหว่างผลการตอบข้อสอบกับคุณลักษณะแฝงที่เกี่ยวข้องได้ พารามิเตอร์ของข้อสอบจึงไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ และพารามิเตอร์ความสามารถไม่แปรเปลี่ยนไปตามชุดข้อสอบ (Lord & Novick, 1968; ศิริชัย กาญจนวาสี, 2555) อย่างไรก็ตามสมมติฐานนี้อาจเป็นการฝ่าฝืนธรรมชาติของข้อมูล เนื่องจากโดยทั่วไปผู้สอบในประชากรอาจมีคุณลักษณะที่ต่างกันก็ได้ เช่น นักเรียนจากโรงเรียนเดียวกันอาจมีแบบแผนการตอบ (response pattern) ที่แตกต่างกัน และความแตกต่างนี้อาจมีมากจนเชื่อได้ว่าผู้สอบน่าจะมาจากประชากรกลุ่มย่อย (subpopulation) หลายกลุ่มมากกว่า (Tay et al., 2011)

จากปัญหาดังกล่าวจึงมีการพัฒนาโมเดลรูปแบบใหม่ด้วยการนำโมเดล IRT บูรณาการเข้ากับโมเดลกลุ่มแฝง (LC: latent class) เรียกว่า โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสม (MixIRT: mixture IRT) โดยมีการผ่อนคลายคุณสมบัติความไม่แปรเปลี่ยนลง (Rost, 1990) กล่าวคือโมเดล MixIRT จะยอมให้พารามิเตอร์ของผู้สอบและข้อสอบแปรผันตามกลุ่มแฝงของประชากรได้ การวิเคราะห์ด้วยโมเดล MixIRT จึงให้สารสนเทศทั้งเชิงปริมาณเกี่ยวกับคุณลักษณะแฝงของผู้สอบและข้อสอบ และเชิงคุณภาพเกี่ยวกับกลุ่มแฝงของประชากรซึ่งสมาชิกภายในกลุ่มมีแบบแผนการตอบที่คล้ายคลึงกัน (Mislevy & Verhelst, 1990) ด้วยคุณสมบัติดังกล่าว โมเดล MixIRT จึงมักถูกนำไปใช้เพื่อแก้ปัญหาทางการทดสอบ เช่น การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Cohen & Bolt, 2005) การศึกษากลวิธีการตอบข้อสอบ (Bolt, Cohen, & Wollack, 2001) การตรวจสอบความเร็วในการทำแบบสอบ (Bolt, Cohen, & Wollack, 2002)

นอกจากนี้ธรรมชาติของข้อมูลทางการศึกษา มักพบว่าข้อมูลในระดับล่างจะซ้อน (nested) อยู่ภายใต้ข้อมูลในระดับบน เช่น ข้อมูลนักเรียนซ้อนอยู่ภายใต้ห้องเรียน ห้องเรียนซ้อนอยู่ภายใต้โรงเรียน เป็นต้น เรียกข้อมูลลักษณะนี้ว่าข้อมูลพหุระดับ (multilevel data) ทั้งนี้หากโมเดล MixIRT

มีการละเลยต่อโครงสร้างพหุระดับของข้อมูล ก็อาจส่งผลให้การคัดเลือกโมเดล (model selection) ของเกณฑ์สารสนเทศบางตัวขาดความถูกต้อง และการประมาณค่าพารามิเตอร์อาจคลาดเคลื่อนไปจากค่าจริง (Lee, Cho, & Sterba, 2018; Sen, Cohen, & Kim, 2019) ด้วยเหตุนี้จึงมีการนำโมเดล MixIRT บูรณาการเข้ากับโมเดลพหุระดับ เรียกว่า โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ (MMixIRT: multilevel mixture IRT) เพื่อผ่อนคลายปัญหาดังกล่าว (Cho, 2007; Vermunt, 2007) โมเดล MMixIRT จึงถูกนำมาใช้อย่างแพร่หลายในการวิจัยเกี่ยวกับข้อมูลทางการศึกษา เช่น การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเพื่อจำแนกกลุ่มโรงเรียนตามกลุ่มนักเรียนที่มีแบบแผนการตอบคล้ายกัน (Cho & Cohen, 2010; Finch & Finch, 2013) การจำแนกกลุ่มประเทศตามกลุ่มบุคคลที่มีภาวะอารมณ์คล้ายกัน (Tay et al., 2011) การจำแนกกลุ่มหลักสูตรตามกลุ่มนักศึกษาที่มีความพึงพอใจคล้ายกัน (Bacci & Ginaldi, 2015)

ยิ่งไปกว่านั้นข้อมูลการทดสอบทางการศึกษาส่วนใหญ่ ยังคงนิยมวัดด้วยข้อสอบปรนัยแบบหลายตัวเลือกที่มีการตรวจให้คะแนนแบบสองค่าดังกล่าวไปแล้วข้างต้น ซึ่งข้อสอบรูปแบบนี้เปิดช่องให้มีการนำความรู้บางส่วนเข้ามาช่วยในการเดาได้ เช่น การตัดตัวเลือกที่ทราบว่าจะไม่ถูกออกไปก่อนแล้วจึงเดาตอบแบบสุ่มจากตัวเลือกที่เหลือ เป็นต้น การนำข้อมูลจากข้อสอบรูปแบบนี้มาวิเคราะห์ด้วยโมเดล IRT ทั่วไป เช่น โมเดล Rasch หรือ 2PL ซึ่งไม่มีการคำนึงถึงการเดา จึงอาจทำให้สารสนเทศเกี่ยวกับความรู้บางส่วนของผู้สอบที่ใช้ในการเดาถูกละเลยไป (Andrich & Marais, 2014) ส่วนโมเดล 3PL แม้จะมีการคำนึงถึงการเดา แต่ก็อาจมีข้อจำกัดเกี่ยวกับการประมาณค่าพารามิเตอร์การเดาย้อนกลับ ซึ่งจากการศึกษาเชิงจำลองในหลายงานวิจัยที่ผ่านมาพบว่าค่าประมาณที่ได้มักมีแนวโน้มไม่สัมพันธ์กับค่าจริง รวมถึงปรัชญาความเชื่อว่าโอกาสในการเดาถูกเป็นคุณลักษณะของข้อสอบที่มีได้ผันแปรตามความรู้ของผู้สอบ การแปลผลการเดาของโมเดล 3PL จึงอาจไม่สอดคล้องกับสภาพธรรมชาติเท่าที่ควร (Lee & Bolt, 2018) จากข้อจำกัดดังกล่าว ผู้วิจัยจึงอนุมานว่าโมเดล MMixIRT ที่ประยุกต์จากโมเดล IRT ทั่วไปก็น่าจะมีปัญหาในลักษณะดังกล่าวด้วยเช่นกัน

เพื่อเป็นการหลีกเลี่ยงข้อจำกัดเกี่ยวกับการวัดค่าการเดาของโมเดล MMixIRT งานวิจัยนี้จึงต้องการนำโมเดล MMixIRT มาประยุกต์ใช้กับโมเดลทฤษฎีการตอบสนองข้อสอบทางเลือกที่คำนึงถึงการเดา (alternative IRT model with guessing) (Han, 2012; San Martín, Del Pino, & De Boeck, 2006) โดยจะทดลองนำโมเดล 2PLE (2PL extension) (Zhu, Wang, & Tao, 2019) มาใช้ในการศึกษา เนื่องจากโมเดลมีความน่าสนใจหลายประการ อาทิ การนิยามให้โอกาสในการเดาถูกมีความผันแปรตามความสามารถของผู้สอบและคุณลักษณะของข้อสอบ ได้แก่ ความยาก อำนาจจำแนก และจำนวนตัวเลือกของข้อสอบ การแปลผลการเดาจึงมีแนวโน้มสอดคล้องกับธรรมชาติของการตอบข้อสอบมากขึ้น และโมเดลยังสามารถจำแนกการเดาแบบอิงความสามารถ (ability-based guessing) กับการเดาแบบสุ่ม (random guessing) ออกจากกันได้ รวมถึงการนำความสามารถของ

ผู้สอบและคุณลักษณะของข้อสอบมาคำนวณเป็นโอกาสในการเดาถูกยังเป็นการหลีกเลี่ยงปัญหาที่อาจเกิดขึ้นจากการประมาณค่าพารามิเตอร์การเดาโดยตรง ค่าพารามิเตอร์ความยากและอำนาจจำแนกที่ประมาณได้จากโมเดล 2PLE จึงไม่ได้รับผลกระทบจากปัญหาการประมาณค่าพารามิเตอร์การเดา และน่าจะส่งผลทำให้ความแม่นยำในการประมาณค่าพารามิเตอร์ในภาพรวมสูงกว่าโมเดล 3PL

นอกจากนี้ในการวิเคราะห์ด้วยโมเดลที่มีกลุ่มแฝง จำนวนกลุ่มแฝงในประชากรจะถือว่าเป็นสิ่งที่ไม่สามารถทราบได้ล่วงหน้าและไม่สามารถสังเกตได้โดยตรง ในทางปฏิบัติจึงต้องมีการวิเคราะห์เชิงสำรวจเพื่อตรวจสอบจำนวนกลุ่มแฝงที่อาจเป็นไปได้ก่อน เรียกว่าการคัดเลือกโมเดล ซึ่งเป็นการพิจารณาเปรียบเทียบความสอดคล้องระหว่างข้อมูลกับโมเดลแข่งขันที่มีจำนวนกลุ่มแฝงแตกต่างกัน จากเกณฑ์สารสนเทศ (information criterion) อาทิ AIC (Akaike, 1974) BIC (Schwarz, 1978), CAIC (Bozdogan, 1987) และ DBIC (Draper, 1995)

ทั้งนี้ เพื่อเป็นการตรวจสอบถึงประสิทธิภาพของโมเดล MMixIRT ที่ประยุกต์จากโมเดล 2PLE และเกณฑ์สารสนเทศที่เหมาะสมสำหรับการคัดเลือกโมเดล งานวิจัยนี้จะนำโมเดลดังกล่าวมาศึกษาเปรียบเทียบกับโมเดล MMix3PL ซึ่งเป็นโมเดลที่มีการคำนึงถึงการเดาเช่นกัน โดยในระยะแรกจะเป็นการศึกษาในสถานการณ์จำลอง เพื่อศึกษาว่าเกณฑ์สารสนเทศใดมีความเหมาะสมสำหรับการคัดเลือกโมเดลมากที่สุด และเพื่อศึกษาว่าขั้นตอนวิธี (algorithm) ที่ใช้ในการประมาณค่าพารามิเตอร์สามารถให้ค่าประมาณที่มีความแม่นยำเพียงใด ส่วนระยะที่สองเป็นการสาธิตการใช้โมเดลกับข้อมูลเชิงประจักษ์ โดยจะนำข้อมูลคะแนนการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET: Ordinary National Educational Test) มาใช้ในการศึกษา เพื่อทดลองหาจำนวนกลุ่มแฝง พารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบในกลุ่มแฝง และความสามารถเฉลี่ยในกลุ่มแฝง ผลจากการวิจัยนี้จะช่วยให้สารสนเทศในการตัดสินใจเลือกใช้เกณฑ์สารสนเทศเพื่อคัดเลือกโมเดลสำหรับโมเดล MMixIRT ที่มีการคำนึงถึงการเดาในการตอบข้อสอบ และจะเป็นต้นแบบในการประยุกต์ใช้โมเดล IRT ทางเลือกกับโมเดลวิเคราะห์อื่น เพื่อหลีกเลี่ยงหรือผ่อนคลายข้อจำกัดของโมเดลวิเคราะห์ต่อไป

1.2 คำถามวิจัย

- 1) โมเดล MMixIRT ที่ประยุกต์จากโมเดล 2PLE (โมเดล MMix2PLE) ควรมีรูปแบบเป็นอย่างไร และควรประมาณค่าพารามิเตอร์ด้วยวิธีใด
- 2) เกณฑ์สารสนเทศใด (AIC, BIC, CAIC และ DBIC) ที่มีความเหมาะสมสำหรับการคัดเลือกโมเดลให้กับโมเดล MMix3PL และ MMix2PLE
- 3) ความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง สำหรับโมเดล MMix3PL และ MMix2PLE มีมากน้อยเพียงใด

- 4) การวิเคราะห์ข้อมูลเชิงประจักษ์ด้วยโมเดล MMix3PL และ MMix2PLE ให้ผลเหมือนหรือต่างกันอย่างไร

1.3 วัตถุประสงค์การวิจัย

- 1) เพื่อศึกษารูปแบบของโมเดล MMix2PLE และวิธีการประมาณค่าพารามิเตอร์
- 2) เพื่อเปรียบเทียบความถูกต้องของการคัดเลือกโมเดลระหว่างเกณฑ์สารสนเทศ AIC, BIC, CAIC และ DBIC ภายใต้สถานการณ์จำลอง
- 3) เพื่อศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง ระหว่างโมเดล MMix3PL กับ MMix2PLE ภายใต้สถานการณ์จำลอง
- 4) เพื่อวิเคราะห์หาจำนวนกลุ่มแฝง พารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง จากข้อมูลเชิงประจักษ์ด้วยโมเดล MMix3PL และ MMix2PLE

1.4 สมมติฐานการวิจัย

- 1) จากการสังเคราะห์งานวิจัยเกี่ยวกับการศึกษาความถูกต้องของการคัดเลือกโมเดลระหว่าง AIC, BIC, CAIC และ DBIC พบว่าในบริบทของโมเดล LC เกณฑ์สารสนเทศ DBIC มีแนวโน้มมีความถูกต้องของการคัดเลือกโมเดลมากกว่าเกณฑ์สารสนเทศตัวอื่น ๆ ในหลายสถานการณ์ การศึกษาเมื่อขนาดตัวอย่างใหญ่ตั้งแต่ 600 (Yang & Yang, 2007) ส่วนในบริบทของโมเดล MixIRT พบว่า BIC มีแนวโน้มมีความถูกต้องของการคัดเลือกโมเดลมากกว่า AIC (Choi, 2014; Li et al., 2009; Sen, 2014) และในบริบทของโมเดล MMixIRT ที่ประยุกต์จากโมเดล Rasch (Cho & Cohen, 2010; Sen et al., 2019) และโมเดล 2PL (Lee et al., 2018) พบว่า BIC และ CAIC มีแนวโน้มมีความถูกต้องของการคัดเลือกโมเดลมากกว่า AIC ดังนั้น ผู้วิจัยจึงขอตั้งสมมติฐานการเปรียบเทียบความถูกต้องของการคัดเลือกโมเดลระหว่าง AIC, BIC, CAIC และ DBIC สำหรับโมเดล MMix3PL และ MMix2PLE ภายใต้สถานการณ์จำลองว่า
“BIC, CAIC และ DBIC น่าจะมีความถูกต้องของการคัดเลือกโมเดลมากกว่า AIC สำหรับโมเดล MMix3PL และ MMix2PLE”
- 2) จากการสังเคราะห์งานวิจัยเกี่ยวกับการศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง พบว่าโมเดล MMixIRT มีแนวโน้มให้ความแม่นยำของการประมาณพารามิเตอร์กลุ่มแฝงค่อนข้างสูงมาก (ร้อยละ 98-100) (Cho & Cohen, 2010) โดยความซับซ้อนของโมเดล (model complexity) ที่เพิ่มขึ้นมีแนวโน้มส่งผลให้ความแม่นยำของการประมาณค่าพารามิเตอร์ของข้อสอบลดลง แตกต่าง

จากการเพิ่มขนาดตัวอย่าง ซึ่งมีแนวโน้มช่วยเพิ่มความแม่นยำของการประมาณค่าพารามิเตอร์ของข้อสอบให้สูงขึ้น (Sen, 2014) อย่างไรก็ตามเงื่อนไขดังกล่าวพบว่าไม่มีผลต่อความแม่นยำของการประมาณค่าความสามารถเฉลี่ยในกลุ่มแฝง (Choi, 2014) ผู้วิจัยจึงขอตั้งสมมติฐานการศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบและความสามารถเฉลี่ยในกลุ่มแฝงสำหรับโมเดล MMix3PL และ MMix2PLE ภายใต้สถานการณ์จำลองว่า

“ความแม่นยำของการประมาณค่าพารามิเตอร์ของข้อสอบสำหรับโมเดล MMix3PL และ MMix2PLE น่าจะมีค่าสูงใกล้เคียงกัน และการเพิ่มขนาดตัวอย่างน่าจะช่วยให้ความแม่นยำของทั้งสองโมเดลสูงขึ้น ส่วนความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง และความสามารถเฉลี่ยในกลุ่มแฝง ทั้งสองโมเดลน่าจะไม่แตกต่างกัน”

1.5 ขอบเขตการวิจัย

- 1) ข้อมูลที่ศึกษามี 2 ระดับ คือ ระดับกลุ่ม (ระดับโรงเรียน) และระดับหน่วย (ระดับนักเรียน)
- 2) ผลการตอบข้อสอบเป็นข้อมูลแบบทวิภาค (dichotomous data) คือ 0 (ผิด) และ 1 (ถูก)
- 3) ข้อมูลจำลองถูกสร้างจากการเขียนชุดคำสั่งด้วยโปรแกรม R ส่วนการประมาณค่าพารามิเตอร์จะใช้วิธีการประมาณค่าแบบเบย์ (Bayesian estimation) ด้วยขั้นตอนวิธี Markov chain Monte Carlo (MCMC) ด้วยโปรแกรม MultiBUGS (Goudie et al., 2017)
- 4) ข้อมูลจำลองถูกสมมติให้มีกลุ่มแฝงระดับโรงเรียน 2 กลุ่ม และระดับนักเรียน 2 กลุ่ม ซึ่งเป็นสถานการณ์ขั้นพื้นฐานสำหรับการศึกษากับโมเดล MMixIRT ส่วนความยาวแบบสอบ (test length) ถูกจำลองให้มี 40 ข้อ เพื่อให้ใกล้เคียงกับแบบสอบ O-NET วิชาคณิตศาสตร์ ระดับ ม.6 (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2562ก) ซึ่งวัด 4 สาระการเรียนรู้ คือ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น โดยสมมติให้แต่ละสาระถูกวัดด้วยข้อสอบ 10 ข้อ
- 5) การเปรียบเทียบความถูกต้องของการคัดเลือกโมเดลและความแม่นยำของการประมาณค่าพารามิเตอร์จะถูกศึกษาภายใต้เงื่อนไข ได้แก่ โมเดลจำลอง 2 เงื่อนไข \times จำนวนกลุ่มแฝง 1 เงื่อนไข \times สัดส่วนของสมาชิกในกลุ่มแฝง 1 เงื่อนไข \times ขนาดตัวอย่างของโรงเรียน 1 เงื่อนไข \times ขนาดตัวอย่างของนักเรียน 2 เงื่อนไข \times ความยาวแบบสอบ 1 เงื่อนไข \times โมเดลวิเคราะห์ 8 เงื่อนไข รวมมีเงื่อนไขการศึกษาทั้งสิ้น 32 เงื่อนไข แต่ละเงื่อนไขจะถูกจำลองซ้ำ 10 รอบ ทำให้มีจำนวนชุดข้อมูลที่ต้องวิเคราะห์รวม 320 ชุด
- 6) ข้อมูลเชิงประจักษ์ที่ใช้คือข้อมูลคะแนนสอบ O-NET วิชาคณิตศาสตร์ ระดับ ม.6 ปีการศึกษา 2562 ซึ่งเป็นข้อมูลทุติยภูมิจากฐานข้อมูลของสถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.)

โดยผู้วิจัยจะสุ่มนักเรียนจากโรงเรียนขนาดต่าง ๆ ในแต่ละสังกัดมาเป็นตัวอย่าง เพื่อสถิติการใช้โมเดล ทั้งนี้เหตุผลในการเลือกข้อมูลดังกล่าวมาศึกษา เนื่องจากผลการรายงานที่ผ่านมาพบว่านักเรียนส่วนใหญ่ของประเทศยังค่อนข้างอ่อนในวิชาคณิตศาสตร์ สังเกตได้จากคะแนนเฉลี่ยระดับประเทศที่มีค่าค่อนข้างต่ำเพียง 25.41 (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2562) อย่างไรก็ตามหากพิจารณาคะแนนตามสังกัดของสถานศึกษากลับพบว่านักเรียนจากโรงเรียนในสังกัดกระทรวงการอุดมศึกษา (อว.) มีคะแนนเฉลี่ยสูงถึง 48.32 โดยมีคะแนนเฉลี่ยในสาระที่ 1-4 เท่ากับ 6.39, 3.67, 23.32 และ 11.56 ตามลำดับ ในขณะที่นักเรียนจากโรงเรียนในสังกัดอื่น เช่น สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) มีคะแนนเฉลี่ยค่อนข้างต่ำเพียง 25.62 โดยมีคะแนนเฉลี่ยในสาระที่ 1-4 เท่ากับ 3.62, 1.90, 13.43 และ 6.84 ตามลำดับ ซึ่งจากผลดังกล่าวจึงอนุมานได้ว่าน่าจะมีภาวะวิวิธพันธุ์ (heterogeneity) ในกลุ่มประชากรผู้สอบ และมีแนวโน้มมีความเหมาะสมสำหรับการนำมาศึกษาด้วยโมเดล MMixIRT ทั้งนี้ข้อมูลดังกล่าวจะมี 2 ระดับคือ ข้อมูลระดับโรงเรียน ได้แก่ สังกัด (สพฐ., สข., อว., กสท.) ขนาดโรงเรียน (เล็ก, กลาง, ใหญ่, ใหญ่พิเศษ) และภูมิภาค (กรุงเทพฯ, กลาง, ใต้, ตะวันออก, เหนือ, ตะวันออกเฉียงเหนือ, ตะวันตก) และข้อมูลระดับนักเรียน ได้แก่ เพศ (ชาย, หญิง) รายได้ผู้ปกครอง (1-50,000 บาท, 50,001-100,000 บาท, 100,001-200,000 บาท, 200,001-300,000 บาท, มากกว่า 300,000 บาท) และผลการตอบข้อสอบ (0, 1)

1.6 นิยามศัพท์ที่ใช้ในการวิจัย

- 1) โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ (MMixIRT) หมายถึง ตัวแบบการวิเคราะห์ที่บูรณาการองค์ความรู้มาจาก 3 โมเดล ได้แก่ โมเดล IRT โมเดล LC และโมเดลพหุระดับ โมเดลจึงสามารถให้สารสนเทศได้ทั้งเชิงปริมาณเกี่ยวกับคุณลักษณะแฝงของผู้สอบ และข้อสอบตามแนวคิดโมเดล IRT และเชิงคุณภาพเกี่ยวกับกลุ่มแฝงตามแนวคิดโมเดล LC รวมถึงมีการคำนึงถึงโครงสร้างพหุระดับของข้อมูลตามแนวคิดโมเดลพหุระดับ
- 2) โมเดล MMix3PL หมายถึง โมเดล MMixIRT ที่มีการนำโมเดล 3PL มาประยุกต์ ทั้งนี้พารามิเตอร์ของโมเดลที่สนใจประมาณค่าในงานวิจัยนี้ ได้แก่ พารามิเตอร์กลุ่มแฝงระดับโรงเรียน (π_g) พารามิเตอร์กลุ่มแฝงระดับนักเรียน ($\pi_{k|g}$) พารามิเตอร์ของข้อสอบในกลุ่มแฝง ($a_{ikg}, b_{ikg}, c_{ikg}$) และพารามิเตอร์ความสามารถเฉลี่ยในกลุ่มแฝง (μ_{kg})
- 3) โมเดล MMix2PLE หมายถึง โมเดล MMixIRT ที่มีการนำโมเดล 2PLE ซึ่งเป็นโมเดล IRT ทางเลือกมาประยุกต์ ทั้งนี้พารามิเตอร์ของโมเดลที่สนใจประมาณค่าในงานวิจัยนี้ ได้แก่ พารามิเตอร์กลุ่มแฝงระดับโรงเรียน (π_g) พารามิเตอร์กลุ่มแฝงระดับนักเรียน ($\pi_{k|g}$) พารามิเตอร์ของข้อสอบในกลุ่มแฝง (a_{ikg}, b_{ikg}) และพารามิเตอร์ความสามารถเฉลี่ยในกลุ่มแฝง (μ_{kg})

- 4) **ความถูกต้องของการคัดเลือกโมเดล** หมายถึง ความตรงกันระหว่างโมเดลที่ได้จากการคัดเลือกด้วยเกณฑ์สารสนเทศกับโมเดลประชากร คำนวณจากค่าร้อยละของจำนวนครั้งที่เกณฑ์สารสนเทศ (AIC, BIC, CAIC, DBIC) เลือกโมเดลได้ตรงกับโมเดลประชากรจากข้อมูลจำลอง 10 ชุด หากโมเดลใดมีเกณฑ์สารสนเทศต่ำที่สุด จะหมายความว่าโมเดลนั้นถูกเลือกให้มีความสอดคล้องกับข้อมูลมากที่สุด และเกณฑ์สารสนเทศใดมีร้อยละของจำนวนครั้งที่เลือกโมเดลได้ตรงกับโมเดลประชากรมากกว่า จะหมายความว่าเกณฑ์สารสนเทศนั้นมีความถูกต้องของการคัดเลือกโมเดลมากกว่า
- 5) **ความแม่นยำของการประมาณค่าพารามิเตอร์** หมายถึง ความสอดคล้องระหว่างค่าพารามิเตอร์ที่ประมาณได้กับค่าพารามิเตอร์จริง ทั้งนี้หากเป็นพารามิเตอร์กลุ่มแฝง (π) ค่าความแม่นยำจะพิจารณาจากร้อยละของความใกล้เคียง (p) แต่ถ้าเป็นพารามิเตอร์ของข้อสอบ (a, b, c) ค่าความแม่นยำจะพิจารณาจากค่าสหสัมพันธ์ (correlation) ความลำเอียง (bias) และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE: root mean square error) ซึ่งเป็นค่าที่แสดงถึงความสัมพันธ์และความคลาดเคลื่อนระหว่างค่าประมาณกับค่าจริง ส่วนพารามิเตอร์ความสามารถเฉลี่ย (μ) ค่าความแม่นยำจะพิจารณาจาก bias และ RMSE ทั้งนี้ค่าสถิติทั้งหมดจะถูกคำนวณจากข้อมูลจำลองจำนวน 10 ชุด หากค่า p สูงเข้าใกล้ 100 ค่า correlation สูงเข้าใกล้ 1 ค่า bias และ RMSE ต่ำเข้าใกล้ 0 จะหมายความว่า การประมาณค่าพารามิเตอร์มีความแม่นยำสูง

1.7 ประโยชน์ที่ได้รับ

ประโยชน์เชิงวิชาการ

- 1) งานวิจัยนี้เป็นต้นแบบของการประยุกต์ใช้โมเดล IRT ทางเลือกที่คำนึงถึงการเดากับโมเดล MMixIRT ซึ่งมีข้อได้เปรียบเหนือกว่าโมเดล MMix3PL เกี่ยวกับการหลีกเลี่ยงปัญหาที่อาจเกิดขึ้นจากการประมาณค่าพารามิเตอร์การเดาโดยตรง และการอธิบายโอกาสในการเดา ถูกที่มีแนวโน้มสอดคล้องกับธรรมชาติของการตอบข้อสอบมากขึ้น เนื่องจากมีการนำคุณลักษณะของผู้สอบและข้อสอบเข้ามาร่วมอธิบายโอกาสในการเดาถูก
- 2) ผลจากการวิจัยทำให้ทราบจุดเด่นของวิธีการประมาณค่าแบบเบสส์สำหรับโมเดล MMixIRT ซึ่งพบว่ามี ความแม่นยำในการประมาณค่า สามารถระบุค่าที่แน่นอนได้เพียงหนึ่งเดียว และให้ค่าประมาณค่าพารามิเตอร์ย้อนกลับได้ค่อนข้างแม่นยำ แม้ว่าขนาดตัวอย่างในกลุ่มแฝงจะมีขนาดเล็ก แตกต่างจากวิธีการประมาณค่าทางสถิติทั่วไปที่จำเป็นต้องอาศัยข้อมูลตัวอย่างที่ต้องมีขนาดใหญ่เพียงพอ เพื่อไม่ให้เกิดการฝ่าฝืนข้อตกลงเกี่ยวกับความเป็นปกติของข้อมูลในการประมาณค่าพารามิเตอร์

- 3) แนวคิดการประยุกต์ใช้โมเดลในงานวิจัยนี้สามารถนำไปบูรณาการต่อยอดเพื่อใช้กับโมเดลที่มีความซับซ้อนขึ้น เช่น โมเดล IRT ที่มีการวัดแบบพหุมิติ (multidimensional IRT) โมเดล IRT ที่มีการตรวจให้คะแนนหลายค่า (polytomous scoring IRT) และโมเดล IRT ทางเลือกฯ อื่น ซึ่งอาจเป็นแนวทางในการช่วยปรับปรุงให้ผลการวัดมีความสอดคล้องกับธรรมชาติของข้อมูล ลักษณะตัวแปร ปรัชญาความเชื่อ หรือบริบทอื่น ๆ ต่อได้ในอนาคต

ประโยชน์เชิงปฏิบัติ

- 1) นักวัดและประเมินผลการศึกษาได้ทราบถึงข้อจำกัดของโมเดล MMix3PL เกี่ยวกับปัญหาความไม่สัมพันธ์กันระหว่างค่าประมาณที่ได้กับค่าจริงของพารามิเตอร์การเดาสำหรับบางสถานการณ์การศึกษา และปัญหาการไม่สามารถระบุลักษณะเด่นของสมาชิกในกลุ่มแฝงเมื่อนำไปวิเคราะห์กับข้อมูลเชิงประจักษ์ ซึ่งสารสนเทศดังกล่าวอาจชี้ให้เห็นถึงความจำเป็นของโมเดลทางเลือก MMix2PLE ในการนำมาใช้แทนโมเดลการวัดแบบดั้งเดิมสำหรับบางสถานการณ์ เพื่อหลีกเลี่ยงข้อจำกัดดังกล่าวที่อาจเกิดขึ้น
- 2) หน่วยงานทางการศึกษาหรือสถาบันการศึกษาได้สารสนเทศเกี่ยวกับจำนวนกลุ่มแฝงของประชากรนักเรียนระดับชั้น ม.6 จากข้อมูลการสอบ O-NET รวมถึงลักษณะเด่นของสมาชิกในกลุ่มแฝงระดับโรงเรียนและระดับนักเรียน ที่พบว่าความสามารถของนักเรียนในโรงเรียนแต่ละกลุ่มมีแนวโน้มมีความเกี่ยวข้องกับที่ตั้งและสังกัดของโรงเรียน และฐานะทางบ้านของนักเรียน ซึ่งสารสนเทศดังกล่าวอาจถูกนำไปใช้ในการพัฒนาหลักสูตรการจัดการเรียนการสอนให้มีความเหมาะสมกับโรงเรียนตามที่ตั้งหรือสังกัด หรืออาจช่วยให้ทราบว่านักเรียนแต่ละกลุ่มยังด้อยในสมรรถภาพทางคณิตศาสตร์เนื้อหาใด ซึ่งจะนำไปสู่การวินิจฉัย วางแผน หรือกำหนดนโยบายเพื่อยกระดับคุณภาพการศึกษาให้มีคุณภาพ
- 3) โมเดล MMix2PLE ที่ประยุกต์ขึ้นในงานวิจัยนี้อาจถูกใช้เป็นโมเดลทางเลือกหนึ่งสำหรับการวิเคราะห์ความเหลื่อมล้ำทางการศึกษา เนื่องจากผลการวิเคราะห์ในส่วนของการศึกษาข้อมูลเชิงประจักษ์พบว่า ความสามารถแฝงโดยเฉลี่ยของนักเรียนในโรงเรียนแต่ละกลุ่มมีความเกี่ยวข้องกับที่ตั้งและสังกัดของโรงเรียน และฐานะทางบ้านของนักเรียน โดยกลุ่มความสามารถสูงส่วนใหญ่เป็นนักเรียนที่มีฐานะทางบ้านดี ศึกษาในโรงเรียนสังกัดสำนักงานคณะกรรมการการอุดมศึกษา (อว.) และอยู่ใน กทม. ส่วนกลุ่มความสามารถต่ำส่วนใหญ่เป็นนักเรียนที่มีฐานะทางบ้านปานกลางถึงต่ำ ศึกษาในโรงเรียนที่ไม่ใช่สังกัด อว. และมีได้อยู่ใน กทม. ซึ่งผลดังกล่าวค่อนข้างสอดคล้องกับงานวิจัยก่อนหน้าหลายงานที่ศึกษาเกี่ยวกับความเหลื่อมล้ำทางการศึกษาโดยใช้ผลการวัดจากคะแนนสอบซึ่งเป็นค่าสังเกตได้

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยเพื่อนำมาสร้างเป็นกรอบแนวคิดในการวิจัย ได้แก่ แนวคิด การเดาคำตอบ โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ การแจกแจงความน่าจะเป็น ของตัวแปรสุ่ม การจำลองด้วยเทคนิคมอนติคาร์โล และข้อมูลการทดสอบ O-NET รายละเอียดมีดังนี้

ตอนที่ 1 แนวคิดการเดาคำตอบในข้อสอบปรนัย

- 1.1 รูปแบบของข้อสอบปรนัย
- 1.2 การเดาคำตอบในข้อสอบปรนัย
- 1.3 ปัจจัยที่ส่งผลต่อการเดาคำตอบ
- 1.4 แนวทางการป้องกันการเดาคำตอบ

ตอนที่ 2 แนวคิดของโมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ

- 2.1 โมเดลทฤษฎีการตอบสนองข้อสอบ (IRT) และโมเดล IRT ทางเลือกที่คำนึงถึงการเดา
- 2.2 โมเดลกลุ่มแฝง (LC)
- 2.3 โมเดลพหุระดับ (multilevel)
- 2.4 โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสม (MixIRT)
- 2.5 โมเดลทฤษฎีการตอบสนองข้อสอบพหุระดับ (multilevel IRT)
- 2.6 โมเดลกลุ่มแฝงพหุระดับ (multilevel LC)
- 2.7 โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ (MMixIRT)

ตอนที่ 3 วิธีการประมาณค่าพารามิเตอร์

- 3.1 วิธีภาวะควรจะเป็นสูงสุด (maximum likelihood)
- 3.2 วิธีการประมาณค่าแบบเบย์ (Bayesian estimation)

ตอนที่ 4 การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo simulation)

ตอนที่ 5 การทดสอบทางการศึกษาระดับชาตินี้พื้นฐาน (O-NET)

ตอนที่ 6 กรอบแนวคิดในการวิจัย

ตอนที่ 1 แนวคิดการเดาคำตอบในข้อสอบปรนัย

ในส่วนนี้เป็นการนำเสนอรูปแบบของข้อสอบปรนัยที่นิยมใช้ในการวัดและประเมินผล รวมถึงแนวคิดเกี่ยวกับการเดาคำตอบ การนำเสนอแบ่งเป็น 4 ประเด็น ได้แก่ 1) รูปแบบของข้อสอบปรนัย 2) การเดาคำตอบในข้อสอบปรนัย 3) ปัจจัยที่ส่งผลต่อการเดาคำตอบ และ 4) แนวทางการป้องกันการเดาคำตอบ รายละเอียดดังนี้

1.1 รูปแบบของข้อสอบปรนัย

ข้อสอบปรนัยเป็นข้อคำถามที่มีการกำหนดรายการตัวเลือกให้กับผู้ตอบ ผู้ตอบสามารถใช้เวลาส่วนใหญ่ในการอ่าน คิด และเลือกคำตอบจากรายการตัวเลือก ข้อสอบปรนัยแบ่งเป็น 3 รูปแบบ ได้แก่ 1) ข้อสอบปรนัยแบบถูก-ผิด (true-false) เป็นข้อสอบที่คำตอบมีความเป็นไปได้สองอย่าง เช่น ถูก-ผิด ใช่-ไม่ใช่ 2) ข้อสอบปรนัยแบบจับคู่ (matching) เป็นข้อสอบที่ให้ผู้สอบจับคู่คำตอบ และ 3) ข้อสอบปรนัยแบบหลายตัวเลือก (multiple-choice) เป็นข้อสอบที่ให้ผู้สอบเลือกคำตอบจากตัวเลือกที่โจทย์กำหนด (ศิริชัย กาญจนวาสี, 2556)

จากรูปแบบของข้อสอบปรนัยข้างต้น ข้อสอบปรนัยแบบหลายตัวเลือกเป็นรูปแบบที่นิยมนำไปทดสอบเพื่อวัดความรู้ของผู้สอบมากที่สุด เนื่องจากข้อสอบดังกล่าวสามารถวัดความรู้ได้หลายระดับ ตั้งแต่ความจำ ความเข้าใจ จนถึงระดับที่สูงขึ้น อีกทั้งยังสามารถใช้เครื่องจักรตรวจให้คะแนนได้อย่างรวดเร็วและแม่นยำ มีค่าใช้จ่ายต่ำ ควบคุมเวลาบริหารจัดการได้ วัดและประเมินผลได้ครอบคลุมตรงตามเนื้อหามากกว่าข้อสอบรูปแบบอื่น (ศิริชัย กาญจนวาสี, 2556; อรรถพรณ แก้วน้อย, 2557)

อย่างไรก็ตามข้อสอบปรนัยอาจมีข้อจำกัดเกี่ยวกับการพัฒนา เนื่องจากการสร้างข้อสอบให้มีคุณภาพยังทำได้ค่อนข้างยาก เพราะจำเป็นต้องอาศัยผู้ทรงคุณวุฒิที่มีความเชี่ยวชาญในเนื้อหาและมีทักษะในการเขียนข้อสอบเป็นอย่างดี นอกจากนี้ยังมีข้อจำกัดเกี่ยวกับการนำไปใช้ในเชิงปฏิบัติคือ 1) ไม่สามารถหลีกเลี่ยงการเดาคำตอบของผู้สอบที่มีความรู้ไม่เพียงพอ ความแปรปรวนของคะแนนความคลาดเคลื่อนที่วัดได้จากแบบสอบจึงอาจมีแนวโน้มสูง ส่งผลทำให้ความเที่ยงและความตรงต่ำ (Budesu & Bo, 2015) และ 2) ไม่สามารถวัดผู้ตอบที่มีความรู้บางส่วน เนื่องจากการตรวจให้คะแนนมีเพียง 2 ค่าคือ 0 (ผิด/ไม่มีความรู้) กับ 1 (ถูก/มีความรู้) จึงไม่เพียงพอต่อการจำแนกผู้สอบกลุ่มนี้ออกจากกลุ่มที่มีความรู้สมบูรณ์ และกลุ่มที่ไม่มีความรู้เลย (Bo et al., 2015)

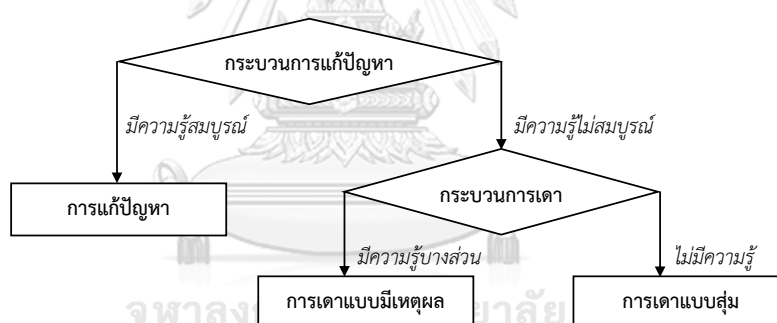
จากข้อจำกัดที่กล่าวมา การเดาจึงเป็นสิ่งที่ส่งผลกระทบต่อการวัดและประเมินผลมากที่สุด เนื่องจากการเปิดช่องให้ผู้สอบที่มีความรู้ไม่เพียงพอมีโอกาสได้คะแนนจากการตอบคำถาม คะแนนที่วัดได้จึงไม่สะท้อนถึงความรู้จริงของผู้สอบ จึงอาจเกิดข้อผิดพลาดในการนำผลคะแนนไปใช้เพื่อประเมินหรือตัดสิน โดยเฉพาะเรื่องที่มีความสำคัญและไม่มีโอกาสติดตามตรวจสอบได้อีก เช่น การสอบเพื่อสำเร็จการศึกษา การสอบคัดเลือกเพื่อเข้าศึกษาต่อหรือทำงาน (ศิริชัย กาญจนวาสี, 2556)

1.2 การเดาคำตอบในข้อสอบปรนัย

จากการศึกษางานวิจัยเกี่ยวกับการเดา พบว่ามีการให้ความหมายและแบ่งรูปแบบไว้ดังนี้ Birnbaum (1968) ให้ความหมายในเชิงการวัดผลว่า “การเดา” หมายถึงโอกาสในการตอบข้อสอบถูกของผู้ตอบที่ไม่มีความรู้ใดเลยเกี่ยวกับข้อคำถาม

Mehrens และ Lehmann (1984) แบ่งการเดาเป็น 2 รูปแบบ ได้แก่ 1) การเดาแบบสุ่ม (blind-, wild-, random guessing) หมายถึง การสุ่มเลือกคำตอบจากบรรดาตัวเลือกทั้งหมด และ 2) การเดาแบบมีเหตุผล (informed-, educated-, logical guessing) หมายถึง การนำความรู้บางส่วนเข้ามาช่วยเลือกคำตอบที่มีแนวโน้มถูกต้อง เช่น การตอบโดยตัดตัวเลือกที่ทราบว่าจะไม่ถูกต้องออกไปก่อน จากนั้นจึงเดาแบบสุ่มจากตัวเลือกที่เหลือ

Hutchinson (1991) อธิบายว่าในการตอบข้อคำถามจะประกอบด้วย 2 กระบวนการ ได้แก่ 1) กระบวนการแก้ปัญหา เป็นการตั้งใจพยายามหาคำตอบ และ 2) กระบวนการเดา เป็นการคาดเดาคำตอบเมื่อผู้ตอบมีความรู้ไม่สมบูรณ์ แบ่งเป็น 2 รูปแบบย่อย คือ 2.1) การเดาแบบมีเหตุผล เกิดขึ้นเมื่อผู้ตอบมีความรู้บางส่วนเกี่ยวกับข้อคำถาม และ 2.2) การเดาแบบสุ่ม เกิดขึ้นเมื่อผู้ตอบไม่มีความรู้ใดเลยเกี่ยวกับข้อคำถาม ดังภาพ 2.1



ภาพ 2.1 การอธิบายการเดาตามแนวคิดของ Hutchinson (1991)

Smith (1993) อธิบายว่ารูปแบบการเดาทั้งสองมีความแตกต่างกันในแง่ของการให้สารสนเทศเกี่ยวกับความรู้ของผู้ตอบ กล่าวคือการเดาแบบสุ่มถือว่าไม่มีการให้สารสนเทศใดเลย เนื่องจากการเดารูปแบบนี้ไม่ได้มาจากการใช้ความรู้ของผู้ตอบ แตกต่างจากการเดาแบบมีเหตุผลที่มีการให้สารสนเทศบางส่วน เนื่องจากการเดารูปแบบนี้มีการนำความรู้บางส่วนเข้ามาช่วยในการเดา

San Martín และคณะ (2006) เรียกการเดาแบบมีเหตุผลว่า “การเดาแบบอิงความสามารถ” เนื่องจากเชื่อว่าโอกาสในการเดาถูกรูปแบบนี้มีความผันแปรตามความสามารถของผู้ตอบเป็นหลัก

Han (2012) ให้ความหมายว่า “การเดา” หมายถึง สิ่งที่เกิดขึ้นเมื่อผู้ตอบไม่ทราบคำตอบที่ถูกต้องของข้อคำถาม แต่ยังคงพยายามที่จะตอบคำถามนั้นเพื่อต้องการเพิ่มโอกาสในการได้คะแนน

ดังนั้น “การเดา” จึงหมายถึงพฤติกรรมในการตอบข้อคำถามของผู้ตอบ เมื่อผู้ตอบมีความรู้ไม่สมบูรณ์ ทั้งนี้การเดาอาจแบ่งได้ 2 รูปแบบ คือ 1) การเดาแบบสุ่ม เกิดขึ้นเมื่อผู้ตอบไม่มีความรู้ใดเลย

เกี่ยวกับข้อคำถาม การเดารูปแบบนี้ถือว่าไม่มีการให้สารสนเทศใดเกี่ยวกับความรู้ของผู้ตอบ และ 2) การเดาแบบอิงความสามารถ เกิดขึ้นเมื่อผู้ตอบมีความรู้บางส่วนเกี่ยวกับข้อคำถาม การเดารูปแบบนี้ถือว่ามีการให้สารสนเทศบางส่วนเกี่ยวกับความรู้ของผู้ตอบ

1.3 ปัจจัยที่ส่งผลต่อการเดาคำตอบ

จากการสังเคราะห์งานวิจัยเกี่ยวกับการเดา พบว่าปัจจัยที่ส่งผลต่อการเดาอาจมาจาก 2 ด้าน คือ 1) คุณลักษณะของข้อสอบ และ 2) คุณลักษณะของผู้สอบ รายละเอียดดังนี้

ก) คุณลักษณะของข้อสอบ

ก.1) *ความยาก* เมื่อข้อสอบมีระดับความยากสูงกว่าความรู้ของผู้ตอบ ผู้ตอบจะเริ่มมีแนวโน้มคาดเดาคำตอบมากขึ้น โดยข้อสอบระดับง่ายมักเปิดช่องให้มีโอกาสในการเดาถูกสูงกว่าปกติ ตรงกันข้ามกับข้อสอบระดับยาก ซึ่งโอกาสในการเดาถูกมักมีแนวโน้มลดต่ำลง (Tsai & Lin, 2015) นอกจากนี้ตำแหน่งของข้อสอบที่มีระดับความยากแตกต่างกันก็อาจส่งผลต่อการเดาเช่นกัน ตัวอย่างเช่น หากข้อสอบในช่วงเริ่มต้นยากเกินไป ผู้ตอบอาจหมดกำลังใจในการสอบและเริ่มเดาคำตอบตั้งแต่ช่วงแรก หรือถ้าข้อสอบในช่วงท้ายยากมากขึ้นและผู้ตอบรู้สึกเหนื่อยล้า ผู้ตอบก็อาจเริ่มเดาคำตอบเช่นกัน (Cao & Stokes, 2008)

ก.2) *อำนาจจำแนก* ข้อสอบที่มีอำนาจจำแนกต่ำหรือไม่สามารถจำแนกผู้ตอบที่มีความสามารถสูงและต่ำออกจากกันได้ มักเปิดช่องให้ผู้ตอบทั้งสองกลุ่มมีโอกาสดูถูกใกล้เคียงกัน และการที่ผู้ตอบกลุ่มต่ำมีโอกาสในการตอบถูกใกล้เคียงกับผู้ตอบกลุ่มสูง ย่อมหมายความว่าข้อสอบข้อนั้นมีโอกาสในการเดาถูกสูง (Hambleton, Swaminathan, & Rogers, 1991)

ก.3) *จำนวนตัวเลือก* หากตัวเลือกในข้อสอบมีมากขึ้น โอกาสในการเดาถูกย่อมลดลงสังเกตได้จากโอกาสในการเดาถูกแบบสุ่ม ซึ่งมีค่า $1/m$ เมื่อ m คือจำนวนตัวเลือก (Han, 2012)

ก.4) *คุณภาพตัวลวง* ตัวลวงที่มีคุณภาพจะมีผลต่อการเดาเป็นอย่างมาก โดยเฉพาะในกลุ่มผู้ตอบความรู้ต่ำ ซึ่งพบว่ามีความสามารถล่อลวงผู้ตอบความรู้ต่ำหรือมีความรู้ไม่สมบูรณ์ได้ดี การเดาของผู้ตอบกลุ่มนี้จึงมักไม่ใช้การเดาแบบสุ่ม (Lord, 1974)

ก.5) *วิชาที่สอบ/ลักษณะตัวเลือก* โดยทั่วไปในวิชาคณิตศาสตร์ ตัวเลือกในข้อสอบมักมีลักษณะเป็นตัวเลข ซึ่งมีจุดเด่นคือผู้ตอบที่มีความรู้สมบูรณ์มักตอบข้อสอบได้ทั้งหมด แต่หากไม่มีความรู้หรือมีความรู้เพียงบางส่วนจะตอบไม่ได้เลย (all-or-none) การเดาในวิชาคณิตศาสตร์จึงมักเป็นการเดาแบบสุ่ม ส่วนในวิชาทางภาษา ตัวเลือกในข้อสอบมักมีลักษณะเป็นข้อความ ผู้ตอบที่มีความรู้ไม่เพียงพอในการตอบสามารถอ่านข้อความในแต่ละตัวเลือก และใช้ความรู้ที่มีตัดตัวเลือกที่ทราบว่าจะไม่ถูกออกไปก่อน แล้วจึงเดาแบบสุ่มจากตัวเลือกที่เหลือ การเดาในวิชาทางภาษาจึงมักเป็นการเดาแบบอิงความสามารถ (San Martín et al., 2006)

ก.6) *วิธีการตอบ/การตรวจให้คะแนน* ข้อสอบปรนัยที่มีตัวถูกหลายตัว (multiple correct options) และมีการตรวจให้คะแนนบางส่วน (partial credit) พบว่าสามารถลดโอกาสในการเดาถูกของผู้ตอบลงไปได้มาก เมื่อเทียบกับข้อสอบปรนัยแบบมีตัวถูกเพียงตัวเดียว (Ben-Simon, Budescu, & Nevo, 1997) ส่วนการตรวจให้คะแนนเฉพาะข้อที่ถูก (rights-only scoring) พบว่าผู้ตอบส่วนใหญ่มีแนวโน้มชอบเสี่ยงเดามากกว่าเว้นคำตอบไว้ (Guo, 2017)

ข) *คุณลักษณะของผู้สอบ*

ข.1) *เพศ* การศึกษาความแตกต่างของพฤติกรรมทดสอบระหว่างเพศชายกับหญิงในบริบทการสอบที่มีความสำคัญสูง (high-stakes testing) เช่น การสอบเก็บคะแนน การสอบเลื่อนชั้น มักพบว่าเพศชายมีพฤติกรรมชอบเสี่ยงเดาคำตอบมากกว่าเว้นคำตอบไว้ ตรงกันข้ามกับเพศหญิงที่ไม่ชอบเสี่ยงเดาคำตอบ แม้จะไม่มี การหักคะแนนเมื่อเดาผิดก็ตาม (Baldiga, 2014) ทั้งนี้ผลดังกล่าวอาจมาจากความแตกต่างทางเพศเกี่ยวกับความเป็นเลิศในการแข่งขันในสถานการณ์ที่มีความเสี่ยง โดยเพศชายอาจคิดว่าการได้ผลงานดีคือการเพิ่มคะแนนให้สูงที่สุดหรือการไม่ยอมรับว่าไม่ทราบคำตอบจึงเสี่ยงเดาคำตอบมากกว่าเว้นคำตอบไว้ ส่วนเพศหญิงอาจคิดว่าการได้ผลงานดีคือการไม่ต้องรับการลงโทษใด จึงเว้นคำตอบมากกว่าเสี่ยงเดา (Croson & Gneezy, 2009)

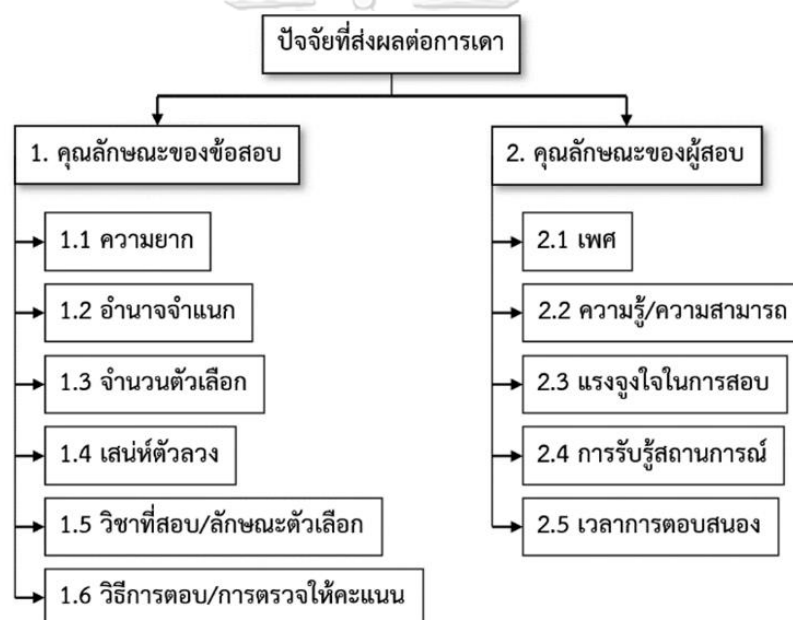
ข.2) *ความรู้/ความสามารถ* ผู้ตอบความสามารถต่ำมักถูกล่อลวงด้วยตัวลวงที่มีคุณภาพโอกาสในการเดาถูกจึงมีแนวโน้มต่ำกว่าโอกาสในการเดาถูกแบบสุ่ม (Lord, 1974) ส่วนผู้ตอบความสามารถสูงมักใช้การตัดตัวเลือกที่ไม่ถูกออกไปก่อน แล้วจึงเดาคำตอบจากตัวเลือกที่เหลือโอกาสในการเดาถูกจึงมีแนวโน้มสูงกว่าโอกาสในการเดาถูกแบบสุ่ม (San Martín et al., 2006)

ข.3) *แรงจูงใจในการสอบ* มักมีความสัมพันธ์กับความสำคัญของการสอบและอาจส่งผลต่อการเดา โดยเฉพาะการสอบที่มีความสำคัญต่ำ (low-stakes testing) เช่น การสอบเพื่อติดตามพัฒนาการผู้เรียน การสอบประเมินระดับชาติที่ไม่มีผลได้ผลเสียต่อตัวผู้สอบ เป็นต้น การสอบในลักษณะนี้ผู้สอบมักให้ความสนใจและพยายามทำข้อสอบในช่วงแรกเท่านั้น เนื่องด้วยความอยากรู้อยากเห็น แต่พอถึงจุดหนึ่งผู้สอบจะละทิ้งความพยายามและเริ่มเดาคำตอบ ประกอบกับการสอบดังกล่าวมักใช้ข้อสอบที่มีจำนวนมากและมีรูปแบบการตอบที่หลากหลาย ผู้สอบจึงเกิดความเหนื่อยล้าได้ง่าย เป็นผลทำให้เกิดการเดามากขึ้นในช่วงท้ายของการทดสอบ (Cao & Stokes, 2008)

ข.4) *การรับรู้สถานการณ์* เมื่อมีการลงโทษหรือการหักคะแนนเมื่อเดาผิด ผู้ตอบที่คาดการณ์ได้ว่ามีคะแนนสูงกว่าคะแนนที่ยอมรับได้ต่ำสุด (minimal acceptable point) หรือคะแนนอ้างอิง มักจะเว้นคำตอบในข้อที่ไม่แน่ใจไว้ เพราะเชื่อมั่นว่าจะได้คะแนนอยู่ในตำแหน่งเดิม จึงไม่จำเป็นต้องเสี่ยงเดาคำตอบ ส่วนผู้ตอบที่คาดการณ์ว่ามีคะแนนต่ำกว่าคะแนนอ้างอิง มักมีแนวโน้มเดาคำตอบมากกว่า เนื่องจากต้องเสี่ยงเพื่อเพิ่มโอกาสให้อยู่ในตำแหน่งที่สูงกว่าคะแนนอ้างอิง (Bereby-Meyer, Meyer, & Flascher, 2002)

ข.5) เวลาการตอบสนอง เมื่อใช้การสอบผ่านคอมพิวเตอร์ (computer-based test) เวลาที่ใช้ในการตอบข้อสอบอาจเข้ามามีบทบาทเป็นตัวจำแนกความแตกต่างระหว่างกระบวนการแก้ปัญหาที่กระบวนการเดา ซึ่งสามารถพิจารณาได้จากขีดแบ่งเวลา (time threshold) ที่ใช้ในการตอบข้อสอบแต่ละข้อ โดยการตอบที่ใช้เวลาน้อยกว่าขีดแบ่งเวลา มักพบว่ามีความโน้มที่จะเป็นการเดาคำตอบ ส่วนการตอบที่ใช้เวลามากกว่าขีดแบ่งเวลา มักมีความโน้มที่จะเป็นการแก้ปัญหามากกว่าการเดา (Wise & Kong, 2005)

โดยสรุปปัจจัยที่ส่งผลต่อการเดาคำตอบในข้อสอบปรนัยอาจมี 2 ด้าน คือ 1) คุณลักษณะของข้อสอบ ได้แก่ ความยาก อำนาจจำแนก จำนวนตัวเลือก คุณภาพตัวเลือก วิชาที่สอบ/ลักษณะตัวเลือก วิธีการตอบ/การตรวจให้คะแนน และ 2) คุณลักษณะของผู้สอบ ได้แก่ เพศ ความรู้/ความสามารถ แรงจูงใจในการสอบ การรับรู้สถานการณ์ และเวลาการตอบสนอง ดังภาพ 2.2



ภาพ 2.2 ปัจจัยที่ส่งผลต่อการเดาคำตอบ

1.4 แนวทางการป้องกันการเดาคำตอบ

จากการศึกษางานวิจัยเกี่ยวกับแนวทางการป้องกันการเดาคำตอบ พบว่าอาจแบ่งได้เป็น 4 วิธี ได้แก่ 1) การปรับวิธีการตรวจให้คะแนน 2) การเปลี่ยนรูปแบบการตอบ 3) การเปลี่ยนโครงสร้างของข้อสอบ และ 4) การหาความเชื่อมั่นของผลการตอบ รายละเอียดดังนี้

ก) การปรับวิธีการตรวจให้คะแนน

ก.1) การตรวจให้คะแนนตามสูตร (formula scoring) เช่น การใช้สูตรปรับแก้การเดา (correction for guessing) หรือ $C = R - W / (m - 1)$ เมื่อ C คือคะแนนปรับแก้ R คือจำนวนข้อที่ตอบถูก W คือจำนวนข้อที่ตอบผิด และ m คือจำนวนตัวเลือกในข้อสอบ วิธีนี้พบว่าสามารถลด

แรงจูงใจในการเดาของผู้สอบกลุ่มที่ไม่ชอบความเสี่ยงได้ ด้วยการเว้นคำตอบแทนการเดา (Holzinger, 1924) ซึ่งอาจช่วยสนับสนุนคุณธรรมและจริยธรรมของผู้สอบ (Lindquist & Hoover, 2015) และอาจช่วยเพิ่มความเที่ยงให้สูงขึ้น (Budescu & Bar-Hillel, 1993) อย่างไรก็ตามวิธีนี้มีข้อจำกัดคือ การลงโทษหรือการหักคะแนนเมื่อเดาผิดจะส่งผลเสียทั้งต่อตัวผู้สอบและผู้ใช้ผลคะแนน เนื่องจากในสถานการณ์จริงอาจมีผู้สอบหลายประเภท เช่น กลุ่มที่ไม่ชอบเสี่ยง กลุ่มที่ไม่ชอบสูญเสีย และกลุ่มอนุรักษ์นิยมการทำข้อสอบแบบดั้งเดิม ผู้สอบแต่ละกลุ่มจึงอาจถูกลงโทษอย่างไม่เป็นสัดส่วน ส่งผลทำให้คะแนนที่ประมาณได้เกิดความลำเอียง (Budescu & Bo, 2015)

ก.2) *การกำหนดน้ำหนักให้ข้อสอบ (item weighting)* เช่น การกำหนดให้ข้อสอบแต่ละข้อมีน้ำหนักคะแนนแตกต่างกัน ซึ่งอาจกำหนดตามระดับความยาก ความตรง ความสามารถในการวินิจฉัย ด้วยเทคนิคการวิเคราะห์ถดถอยหรือการวิเคราะห์องค์ประกอบ (Sykes & Hou, 2003) อย่างไรก็ตามวิธีนี้ยังค่อนข้างมีความซับซ้อนและกำหนดน้ำหนักที่เหมาะสมได้ยาก

ก.3) *การกำหนดน้ำหนักให้ตัวเลือก (option weighting)* เช่น การกำหนดให้ตัวลวงมีความรุนแรงในการลงโทษแตกต่างกันเมื่อผู้สอบเดาผิด ซึ่งสามารถกำหนดได้หลายวิธี เช่น กำหนดตามการตัดสินใจของผู้เชี่ยวชาญจากโครงสร้างความรู้ที่มุ่งวัด (Smith, 1987) กำหนดจากผลการตอบที่ได้จากแบบสอบคู่ขนานที่เป็นมาตรฐานหรือความรู้ก่อนหน้าเกี่ยวกับน้ำหนักคะแนน (Ben-Simon et al., 1997) วิธีดังกล่าวพบว่าอาจช่วยเพิ่มความเที่ยงแบบสอดคล้องภายใน (internal consistency reliability) ให้มีค่าสูงขึ้นได้ (Echternacht, 1976) อย่างไรก็ตามวิธีนี้ไม่ค่อยได้รับความนิยม เนื่องจากมีค่าใช้จ่ายสูงในการวางแผนกำหนดน้ำหนัก และมักพบปัญหาการกำหนดน้ำหนักให้มีความเหมาะสม รวมถึงการคำนวณคะแนนยังมีความซับซ้อน และไม่สะดวกต่อการอธิบายขั้นตอนการตรวจให้ผู้สอบทราบ (Frary, 1989)

ข) การเปลี่ยนรูปแบบการตอบ

ข.1) *การตอบปฏิเสธตัวเลือกที่ผิด* เช่น การสอบแบบตัดทิ้ง (elimination testing) การตอบรูปแบบนี้ผู้สอบต้องระบุว่าตัวเลือกใดเป็นตัวเลือกที่ผิดบ้าง ส่วนตัวเลือกที่คิดว่าเป็นตัวเลือกที่ถูกหรือไม่แน่ใจให้เว้นไว้ กล่าวคือผู้สอบต้องตัดตัวเลือกที่คิดว่าผิดออกไปให้ได้มากที่สุดเท่าที่สามารถทำได้ การตรวจให้คะแนนจะพิจารณาจากทุกตัวเลือกที่ทำเครื่องหมายตอบ และคำนวณคะแนนจากจำนวนตัวเลือกที่ตัดได้ถูกและผิด (Bush, 2015)

ข.2) *การตอบตัวเลือกที่มีแนวโน้มถูก* เช่น การสอบแบบเลือกบางส่วน (subset selection testing) การตอบรูปแบบนี้ผู้สอบสามารถเลือกตัวเลือกที่คาดว่าเป็นตัวเลือกที่ถูกต้องได้หลายตัว แต่ตัวเลือกที่ถูกจะมีเพียงตัวเลือกเดียวเท่านั้น การตรวจให้คะแนนจะมีการลดหลั่นตามจำนวนตัวเลือกที่ผู้สอบเลือกตอบมากขึ้น เนื่องจากเชื่อว่าการตอบตัวเลือกมากขึ้น แสดงถึงความรู้ของผู้สอบที่มีค่าน้อยลง (Dressel & Schmidt, 1953)

รูปแบบการตอบทั้งสองมีความคล้ายคลึงกันในการวัดความรู้บางส่วนของผู้สอบและการลงโทษเมื่อเดาผิด ส่งผลดีต่อการลดโอกาสในการเดาถูก ช่วยจำแนกผู้สอบที่มีระดับความรู้แตกต่างกันได้ชัดเจน และยังเป็น การเพิ่มความเที่ยงได้ทางหนึ่ง (Gibbons, Olkin, & Sobel, 1977) อย่างไรก็ตามการตอบรูปแบบนี้มีข้อจำกัดเกี่ยวกับการตอบและการตรวจให้คะแนน ซึ่งต้องอาศัยเวลานานกว่าการสอบทั่วไป และแม้ว่าการตอบทั้งสองจะมีความเท่าเทียมกันในเชิงคณิตศาสตร์ แต่ในเชิงกลวิธีการคิดเพื่อตอบคำถามยังคงมีความแตกต่างกันอยู่ (Bereby-Meyer, Meyer, & Budescu, 2003)

ค) การเปลี่ยนโครงสร้างข้อสอบ เช่น การปรับข้อสอบให้มีตัวถูกหลายตัว ผู้สอบต้องเลือกตัวเลือกที่ถูกต้องให้ครบทุกตัว โดยไม่ทราบว่าจะตัวเลือกที่ถูกต้องมีจำนวนเท่าใด (Frisbie, 1992) ข้อดีคือโอกาสในการเดาถูกจะลดลงอย่างมากเมื่อเทียบกับข้อสอบที่มีตัวถูกเพียงตัวเดียว และพบว่ามีแนวโน้มให้ความเที่ยงและความตรงสูงกว่าข้อสอบที่มีตัวถูกเพียงตัวเดียวด้วย (Hsu, Moss, & Choosak, 1984) อย่างไรก็ตามข้อสอบดังกล่าวมีข้อจำกัดเกี่ยวกับจำนวนตัวถูกในแต่ละข้อซึ่งอาจมีไม่เท่ากัน คะแนนในแต่ละข้อจึงไม่เท่ากันด้วย หากจะให้คะแนนแต่ละข้อเท่ากันก็ต้องใช้สูตรปรับแก้จึงเป็นการเพิ่มภาระในการตรวจและไม่สะดวกต่อการบริหารจัดการ (Ben-Simon et al., 1997)

ง) การสร้างความเชื่อมั่นให้ผลการตอบ เช่น การเปรียบเทียบระหว่างความรู้จากการประเมินตนเองกับความรู้ที่วัดได้จากแบบสอบ วิธีการนี้เมื่อผู้สอบเลือกคำตอบที่คิดว่าถูกแล้ว ต้องระบุด้วยว่ามั่นใจแค่ไหนว่าผลการตอบนั้นถูกต้อง เพื่อเป็นการสร้างความเชื่อมั่นว่าผลการตอบมีความสอดคล้องกับความรู้จริงของผู้สอบ ข้อดีของวิธีนี้พบว่าอาจให้ความเที่ยงสูงกว่าเมื่อเทียบกับการตรวจให้คะแนนแบบ 0-1 และการตรวจให้คะแนนตามสูตร (Hambleton, Roberts, & Traub, 1970) อย่างไรก็ตามวิธีนี้พบว่าอาจมีปัจจัยที่ไม่เกี่ยวกับความรู้ (non-cognitive factor) เข้ามาเกี่ยวข้องกับการหาความเชื่อมั่น การวัดจึงอาจมีความคลาดเคลื่อนสูง อีกทั้งยังต้องอาศัยเวลาในการบริหารจัดการที่มากขึ้นและมีวิธีการตรวจให้คะแนนที่ซับซ้อนมากกว่าปกติ (Jacobs, 1971) นอกจากนี้ยังควรระมัดระวังหากมีการให้ข้อมูลผิด (misinformation) ตัวอย่างเช่น ผู้สอบมั่นใจว่าคำตอบถูก แต่ปรากฏว่าคำตอบนั้นผิด ซึ่งกรณีดังกล่าวอาจส่งผลทำให้ความเที่ยงต่ำลง (พัชรี มีวรรณ, 2540)

โดยสรุปการป้องกันการเดาทั้ง 4 แนวทางข้างต้น ต่างมีลักษณะที่แตกต่างกัน ซึ่งอาจเปรียบเทียบได้ 3 ประเด็น ได้แก่ 1) สิ่งที่น่าสนใจ 2) ข้อดี และ 3) ข้อจำกัด รายละเอียดดังตาราง 2.1

อย่างไรก็ตามอาจสังเกตได้ว่าวิธีป้องกันการเดาที่กล่าวมา มีข้อจำกัดในเชิงปฏิบัติหลายประการ เช่น การบริหารจัดการที่ต้องใช้เวลานานกว่าปกติ การสร้างเครื่องมือมีค่าใช้จ่ายสูง การคำนวณคะแนนมีความซับซ้อน นอกจากนี้ในสถานการณ์จริงการเดาก็ยังคงเกิดขึ้นอย่างหลีกเลี่ยงไม่ได้ เนื่องจากในการสอบอาจมีปัจจัยต่าง ๆ เข้ามาเกี่ยวข้องทั้งคุณลักษณะของข้อสอบและคุณลักษณะของผู้สอบ จึงเป็นเรื่องยากในเชิงปฏิบัติที่จะป้องกันการเดาด้วยแนวทางดังกล่าว ด้วยเหตุนี้ในเวลาต่อมาจึงมีการพัฒนาโมเดล IRT (เช่น โมเดล 3PL) เพื่อวัดการเดาที่เกิดขึ้นในการทดสอบ

ตาราง 2.1 การเปรียบเทียบการป้องกันการเดาทั้ง 4 แนวทาง

แนวทาง	สิ่งที่สนใจ	ข้อดี	ข้อจำกัด
1. ปรับวิธีการตรวจให้คะแนน			
- ปรับแก้คะแนนตามสูตร	สูตรคำนวณ	- ลดแรงจูงใจในการเดา - อาจเพิ่มความเที่ยงให้ผลการวัด	- อาจเกิดความลำเอียง ในการตรวจให้คะแนนระหว่างกลุ่มผู้สอบ
- กำหนดน้ำหนักให้ข้อสอบ	ความสำคัญของข้อสอบ	- ลดแรงจูงใจในการเดา - อาจเพิ่มความเที่ยงให้ผลการวัด	- กำหนดน้ำหนักที่เหมาะสมยาก - มีขั้นตอนซับซ้อนในการระบุน้ำหนัก
- กำหนดน้ำหนักให้ตัวเลือก	ความสำคัญของตัวเลือก	- ลดแรงจูงใจในการเดา - อาจเพิ่มความเที่ยงให้ผลการวัด	- กำหนดน้ำหนักที่เหมาะสมยาก - คำนวณคะแนนยาก
2. เปลี่ยนรูปแบบการตอบ			
- ตอบตัวเลือกที่ผิด	จำนวนตัวเลือกที่ตอบ	- ลดโอกาสในการเดาถูก - อาจเพิ่มความเที่ยงให้ผลการวัด	- ใช้เวลาบริหารจัดการนานกว่าปกติ - สร้างข้อสอบได้ยาก
- ตอบตัวเลือกที่มีแนวโน้มถูก	จำนวนตัวเลือกที่ตอบ	- ลดโอกาสในการเดาถูก - อาจเพิ่มความเที่ยงให้ผลการวัด	- ใช้เวลาบริหารจัดการนานกว่าปกติ - สร้างข้อสอบได้ยาก
3. เปลี่ยนโครงสร้างข้อสอบ	จำนวนตัวถูก (มากกว่า 1 ตัว)	- ลดโอกาสในการเดาถูก - มีความเที่ยงและความตรงสูงกว่าข้อสอบมาตรฐาน	- ใช้เวลาบริหารจัดการนานกว่าปกติ - มีความซับซ้อนในการคำนวณคะแนนเพราะคะแนนแต่ละข้อไม่เท่ากัน
4. หาความเชื่อมั่นผลการตอบ	ความเชื่อมั่นของคำตอบ	- ให้ความเที่ยงสูงกว่าวิธีการตรวจให้คะแนนแบบ 0-1 และการตรวจให้คะแนนตามสูตร	- มีปัจจัยที่ไม่ใช่ความรู้เข้ามาเกี่ยวข้อง - ใช้เวลาบริหารจัดการนานกว่าปกติ - ควรระวังการให้ข้อมูลผิด

ตอนที่ 2 แนวคิดของโมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ

ในส่วนนี้เป็นการนำเสนอแนวคิดเกี่ยวกับโมเดล IRT แบบผสมพหุระดับ เพื่อทำความเข้าใจถึงองค์ประกอบของโมเดล หลักการและเหตุผลในการบูรณาการโมเดล พารามิเตอร์ในโมเดล รวมถึงงานวิจัยที่เกี่ยวข้อง รายละเอียดดังนี้

2.1 โมเดลทฤษฎีการตอบสนองข้อสอบ (IRT) และโมเดล IRT ทางเลือกที่คำนึงถึงการเดา

ทฤษฎีการทดสอบแบบใหม่ได้เข้ามามีบทบาทในการวัดผลทางการศึกษาด้วยแนวทางที่ต่างไปจากทฤษฎีการทดสอบแบบดั้งเดิม ด้วยการนำโมเดลทฤษฎีการตอบสนองข้อสอบหรือโมเดล IRT เข้ามาช่วยอธิบายความสัมพันธ์ระหว่างคุณลักษณะของข้อสอบและผู้สอบกับความน่าจะเป็นในการตอบข้อสอบถูกในรูปฟังก์ชันของข้อสอบหรือฟังก์ชันการตอบสนองข้อสอบ ซึ่งโดยทั่วไปนิยมเขียนอยู่ในรูปฟังก์ชันโลจิส (ศิริชัย กาญจนวาสี, 2555)

โมเดล IRT มีข้อตกลงเบื้องต้นสำคัญ 3 ประการคือ 1) ความเป็นเอกมิติ (unidimensionality) หมายถึงข้อสอบทุกข้อมุ่งวัดเพียงความสามารถเดียว 2) ความเป็นอิสระ (local independence) หมายถึงผลการตอบข้อสอบข้อเดียวกันของผู้สอบแต่ละคนเป็นอิสระจากกัน (ความเป็นอิสระระหว่างผู้สอบ) และผลการตอบข้อสอบแต่ละข้อของผู้สอบคนเดียวกันเป็นอิสระจากกัน (ความเป็นอิสระระหว่างข้อสอบ) ด้วยคุณสมบัติดังกล่าว พารามิเตอร์ของผู้สอบและข้อสอบจึงไม่แปรเปลี่ยน (invariance) ไปตามกลุ่มผู้สอบหรือชุดข้อสอบที่นำมาใช้ แตกต่างจากทฤษฎีการทดสอบแบบดั้งเดิม

ที่มักให้ผลการวัดเปลี่ยนไปตามกลุ่มผู้สอบหรือชุดข้อสอบ และ 3) การเป็นฟังก์ชันต่อเนื่อง (continuous function) ที่สามารถอธิบายถึงความสัมพันธ์ระหว่างคุณลักษณะแฝงกับความน่าจะเป็นในการตอบข้อสอบได้ถูก (ศิริชัย กาญจนวาสี, 2555) ทั้งนี้โมเดล IRT มีการพัฒนาขึ้นมาหลายรูปแบบ ซึ่งแต่ละรูปแบบอาจมีความแตกต่างกันตามประเภทของข้อมูล การตรวจให้คะแนน ฟังก์ชันทางคณิตศาสตร์ จำนวนพารามิเตอร์ อย่างไรก็ตามโดยทั่วไปโมเดล IRT จะแบ่งเป็น 3 รูปแบบหลักได้แก่ 1) โมเดล IRT ที่มีการตรวจให้คะแนนสองค่า (dichotomous scoring IRT) 2) โมเดล IRT ที่มีการตรวจให้คะแนนหลายค่า (polytomous scoring IRT) และ 3) โมเดล IRT ที่มีการตรวจให้คะแนนแบบต่อเนื่อง (continuous scoring IRT) รายละเอียดดังตาราง 2.2

ตาราง 2.2 รูปแบบของโมเดล IRT

ประเภทของข้อมูล	โมเดล	แหล่งอ้างอิง
ไม่ต่อเนื่อง (มี 2 ค่า) (dichotomous scoring)	1-, 2-, 3-parameter normal ogive (PNO)	Lord (1952)
	1-, 2-, 3-parameter logistic (PL)	Rasch (1960), Birnbaum (1968)
ไม่ต่อเนื่อง (มี >2 ค่า) (polytomous scoring)	nominal response	Bock (1972)
	graded response	Samejima (1969)
	partial credit	Masters (1982)
	generalized partial credit	Muraki (1992)
ต่อเนื่อง	continuous response	Samejima (1972)

(ที่มา: ศิริชัย กาญจนวาสี, 2555; Hambleton & Swaminathan, 1985)

อย่างไรก็ตามงานวิจัยนี้จะนำเสนอโมเดล IRT ที่มีการตรวจให้คะแนนสองค่าเพียง 3 โมเดล ได้แก่ Rasch, 2PL และ 3PL เนื่องจากโมเดลทั้งสามเป็นโมเดลมาตรฐานที่นิยมใช้อย่างแพร่หลายในการวิเคราะห์แบบสอบมาตรฐานทั่วไป รายละเอียดของแต่ละโมเดลมีดังนี้

ก) โมเดลโลจิสต์ 1 พารามิเตอร์ (1PL) หรือโมเดล Rasch

โมเดล Rasch มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (2.1) ซึ่งแสดงถึงความน่าจะเป็นในการตอบข้อสอบ i ได้ถูกต้องของผู้สอบ j ($y_{ij} = 1$ หมายถึงตอบถูก; $y_{ij} = 0$ หมายถึงตอบผิด)

$$P(y_{ij} = 1 | \theta_j) = \frac{1}{1 + \exp[-(\theta_j - b_i)]} \quad (2.1)$$

เมื่อ θ_j คือพารามิเตอร์ความสามารถของผู้สอบ j และ b_i คือพารามิเตอร์ความยากของข้อสอบ i

ข) โมเดลโลจิสต์ 2 พารามิเตอร์ (2PL)

โมเดล 2PL มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (2.2)

$$P(y_{ij} = 1 | \theta_j) = \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]} \quad (2.2)$$

เมื่อ a_i และ b_i คือพารามิเตอร์อำนาจจำแนกและพารามิเตอร์ความยากของข้อสอบ i ตามลำดับ

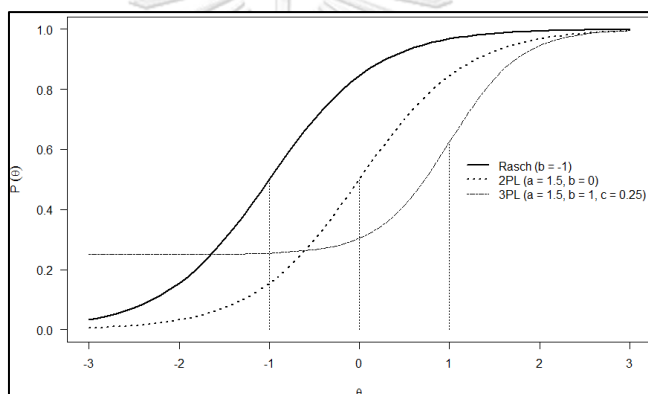
ค) โมเดลโลจิส 3 พารามิเตอร์ (3PL)

โมเดล 3PL มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (2.3)

$$P(y_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]} \quad (2.3)$$

เมื่อ a_i , b_i และ c_i คือพารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก และพารามิเตอร์การเดาของข้อสอบ i ตามลำดับ โดยอาจเรียก c_i ว่าพารามิเตอร์การเดาเทียม (pseudo guessing parameter) เนื่องจากพารามิเตอร์ดังกล่าวคือค่าโอกาสในการเดาถูกซึ่งถือเป็นคุณลักษณะของข้อสอบ มิได้มีความผันแปรตามความสามารถของผู้สอบ ค่าพารามิเตอร์ดังกล่าวจึงยังไม่สอดคล้องกับธรรมชาติของการตอบข้อสอบเท่าที่ควร (Hambleton et al., 1991)

ตัวอย่าง ICCs ของโมเดล Rasch ($b = -1$) โมเดล 2PL ($a = 1.5$, $b = 0$) และโมเดล 3PL ($a = 1.5$, $b = 1$, $c = 0.25$) แสดงดังภาพ 2.3



ภาพ 2.3 ตัวอย่าง ICCs ของโมเดล Rasch, 2PL และ 3PL

ง) โมเดล IRT ทางเลือกที่คำนึงถึงการเดา

จากการศึกษาในหลายงานวิจัยพบว่า โมเดล IRT ที่กล่าวมาข้างต้นอาจมีข้อจำกัดเกี่ยวกับการศึกษาการเดา ตัวอย่างเช่น การนำโมเดล Rasch หรือ 2PL มาวิเคราะห์ข้อสอบปรนัยแบบหลายตัวเลือก ซึ่งข้อสอบรูปแบบนี้เปิดช่องให้ผู้สอบสามารถนำความรู้บางส่วนเข้ามาช่วยในการเดาได้ แต่เนื่องจากโมเดลดังกล่าวไม่มีการคำนึงถึงการเดา สารสนเทศเกี่ยวกับความรู้บางส่วนของผู้สอบที่ใช้ในการเดาจึงอาจถูกละเลยไป (Andrich & Marais, 2014) ส่วนโมเดล 3PL แม้จะมีการคำนึงถึงการเดา แต่ก็อาจมีข้อจำกัดเกี่ยวกับการประมาณค่าพารามิเตอร์ย้อนกลับและการแปลผลการเดา (Lee & Bolt, 2018) โดยพบว่าค่าพารามิเตอร์การเดา (c) ที่ประมาณได้มักมีแนวโน้มไม่สัมพันธ์กับค่าจริงในหลายสถานการณ์ศึกษาเชิงจำลอง ทั้งนี้อาจเป็นเพราะว่า c คือความน่าจะเป็นที่ผู้สอบความสามารถต่ำมากจะตอบข้อสอบได้ถูกต้อง แต่โดยธรรมชาติของข้อมูลทางการศึกษา ผู้สอบที่มีคุณลักษณะดังกล่าวมักมีอยู่น้อยมาก การประมาณค่าโดยอิงจากข้อมูลที่มีอยู่น้อยจึงมักไม่สัมพันธ์กับค่าจริง (Thissen & Wainer, 1982) นอกจากนี้การนิยาม c เป็นพารามิเตอร์ของข้อสอบ ไม่ใช่พารามิเตอร์

ของผู้สอบ ยังเป็นการละเอียดข้อเท็จจริงที่ว่าโอกาสในการเดาถูกอาจแปรผันตามความสามารถของผู้สอบแต่ละคนได้ รวมถึงยังไม่มีการระบุรูปแบบของการเดาว่าเป็นแบบใด การแปลผลการเดาของโมเดล 3PL จึงยังค่อนข้างคลุมเครือและไม่สอดคล้องกับสภาพธรรมชาติ (San Martín et al., 2006)

จากเหตุผลดังกล่าวจึงเป็นที่มาของการพัฒนาโมเดล IRT ทางเลือกที่คำนึงถึงการเดา (Han, 2012; San Martín et al., 2006) ซึ่งเกิดจากการบูรณาการระหว่างปรัชญาความเชื่อการเดาต่าง ๆ กับโมเดล IRT แบบดั้งเดิม (Rasch, 2PL, 3PL) ทำให้โมเดลสามารถพิจารณาโอกาสในการตอบถูกจากกระบวนการแก้ปัญหาและกระบวนการเดาได้พร้อมกัน โดยโอกาสในการตอบถูกจากกระบวนการเดาดังกล่าวจะมีการจำแนกว่าเป็นการเดาถูกแบบสุ่มหรือการเดาถูกแบบอิงความสามารถ ซึ่งคุณสมบัติดังกล่าวอาจมีส่วนช่วยผ่อนคลายหรือหลีกเลี่ยงข้อจำกัดเกี่ยวกับการประมาณค่าและการแปลผลพารามิเตอร์การเดาของโมเดล 3PL

ในงานวิจัยนี้จะขอยกตัวอย่างโมเดล IRT ทางเลือกๆ เพียง 5 โมเดล ได้แก่ 1) โมเดล AG ที่มีปรัชญาความเชื่อที่ว่าโอกาสในการเดาถูกมีความสัมพันธ์กับความสามารถของผู้สอบ 2) โมเดล FG3PL ที่มีปรัชญาความเชื่อที่ว่าโอกาสในการเดาถูกมีความสัมพันธ์กับจำนวนตัวเลือก 3) โมเดล 2PLG ที่มีปรัชญาความเชื่อที่ว่าโอกาสในการเดาถูกมีความสัมพันธ์กับความสามารถของผู้สอบและความยากของข้อสอบ 4) โมเดล 3P-RH ที่มีปรัชญาความเชื่อที่ว่าโอกาสในการเดาถูกมีความสัมพันธ์กับความโน้มเอียงของผลการตอบถูกของผู้สอบแต่ละระดับ และ 5) โมเดล 2PLE ที่มีปรัชญาความเชื่อที่ว่าโอกาสในการเดาถูกมีความสัมพันธ์กับความสามารถของผู้สอบ ความยาก และอำนาจจำแนก รายละเอียดดังนี้

ง.1) โมเดล ability-based guessing (AG)

โมเดล AG (San Martín et al., 2006) ได้รับการขยายแนวคิดมาจากโมเดล Rasch โดยผ่อนคลายให้โอกาสในการเดาถูกมีความสัมพันธ์กับความสามารถของผู้สอบ ดังสมการ (2.4)

$$P(y_{ij} = 1 | \theta_j) = \Lambda[\theta_j - b_i] + (1 - \Lambda[\theta_j - b_i]) \Lambda[\alpha\theta_j + g_i] \quad (2.4)$$

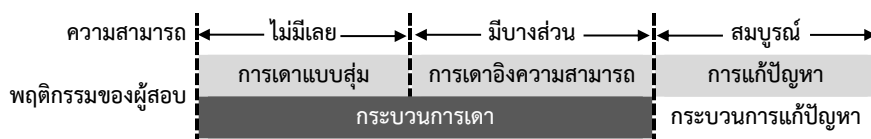
เมื่อ b_i และ g_i คือพารามิเตอร์ความยากและพารามิเตอร์การเดาของข้อสอบ i ; $g_i \in (-\infty, \infty)$; α คือค่าน้ำหนักของความสามารถที่ใช้ในการเดา; $\Lambda[x] = 1/(1 + \exp[-x])$

จากสมการ (2.4) โอกาสในการเดาถูกหรือ $\Lambda[\alpha\theta_j + g_i]$ จะขึ้นอยู่กับพารามิเตอร์การเดา (g) และสัมพันธ์กับความสามารถของผู้สอบ (θ) ด้วยขนาดเท่ากับ α ดังนั้น α และ g จึงทำหน้าที่เป็นตัวกำหนดรูปแบบของการเดาที่เป็นไปได้ 3 กรณี (ดูภาพ 2.4) ดังนี้

กรณี 1: $\alpha = 0$ และ g เข้าใกล้ $-\infty \rightarrow$ ไม่มีการเดาเกิดขึ้น (โมเดล Rasch)

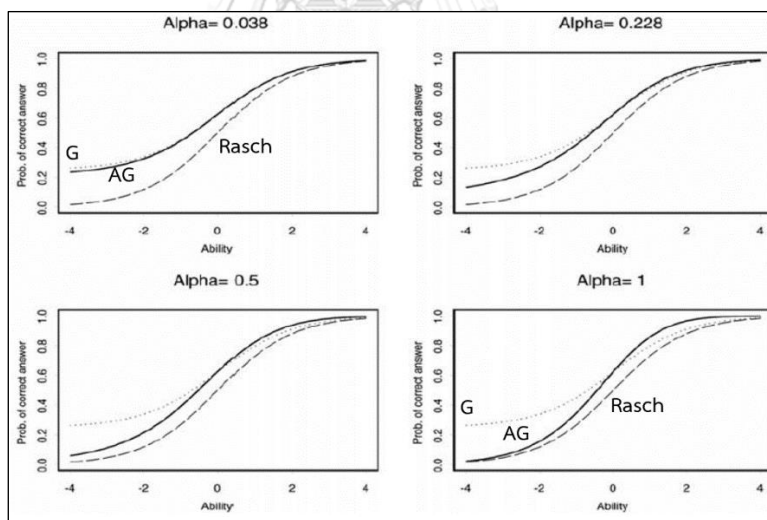
กรณี 2: $\alpha = 0$ และ g มีค่าอื่น ๆ \rightarrow การเดาแบบสุ่ม

กรณี 3: $\alpha > 0$ และ g มีค่าอื่น ๆ \rightarrow การเดาแบบอิงความสามารถ



ภาพ 2.4 มโนทัศน์ของกระบวนการแก้ปัญหาและกระบวนการเดาในโมเดล AG

San Martín และคณะ (2006) ได้เปรียบเทียบโมเดล 3 แบบคือ โมเดล AG (เดาแบบอิงความสามารถ) โมเดล G (เดาแบบสุ่ม) และโมเดล Rasch (ไม่มีการเดา) เพื่อวิเคราะห์ข้อสอบปรนัยแบบ 4 ตัวเลือก พบว่าเมื่อกำหนด $b = 0$, $g = -1.0986$ และ α เข้าใกล้ 0 (ภาพซ้ายบน) ICCs ของโมเดล AG (เส้นทึบ) จะเข้าใกล้ ICCs ของโมเดล G (เส้นจุด) กล่าวคือการเดาที่เกิดขึ้นจะมีแนวโน้มเป็นการเดาแบบสุ่ม แต่เมื่อ α มีค่าเพิ่มขึ้น ทางซ้ายของโมเดล AG จะเข้าใกล้ ICCs ของโมเดล Rasch (เส้นประ) ขณะที่ทางขวาจะสูงกว่า ICCs ของโมเดล Rasch กล่าวคือการเดาที่เกิดขึ้นจะมีแนวโน้มเป็นการเดาแบบอิงความสามารถ ทั้งนี้ San Martín และคณะให้เหตุผลว่าผู้สอบความสามารถต่ำมักมีแนวโน้มถูกล่อลวงด้วยตัวเลือกที่มีคุณภาพได้ง่าย โอกาสในการเดาถูกจึงควรต่ำกว่า .25 (โอกาสในการเดาถูกแบบสุ่ม) ส่วนผู้สอบความสามารถสูงมักอาศัยการตัดตัวเลือกที่ทราบว่าจะไม่ถูกออกไปก่อน แล้วจึงเดาสุ่มจากตัวเลือกที่เหลือ โอกาสในการเดาถูกจึงสูงขึ้นตามความสามารถที่มีมากขึ้นในรูปฟังก์ชันโลจิส และส่งผลทำให้โอกาสในการตอบถูกสูงขึ้นด้วย ดังภาพ 2.5



ภาพ 2.5 ตัวอย่าง ICCs ของโมเดล AG

(ที่มา: San Martín et al., 2006)

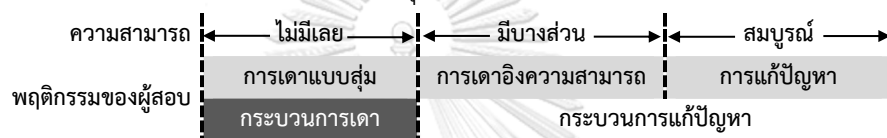
ง.2) โมเดล fixed guessing 3-parameter logistic (FG3PL)

โมเดล FG3PL (Han, 2012) เป็นการปรับนิยามการเดาของโมเดล 3PL ใหม่เป็น “กระบวนการแก้ปัญหาคือแนวทางใดก็ตามที่ช่วยให้ผู้สอบสามารถตอบคำถามได้ ส่วนกระบวนการเดาคือการตอบข้อสอบแบบสุ่มโดยไม่อาศัยความสามารถใดเลย” จากนิยามดังกล่าวโอกาสในการเดาถูกของโมเดลจึงหมายถึงโอกาสในการเดาถูกแบบสุ่มเท่านั้น ดังสมการ (2.5)

$$P(y_{ij} = 1 | \theta_j) = \frac{1}{m_i} + \left(1 - \frac{1}{m_i}\right) \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]} \quad (2.5)$$

เมื่อ a_i และ b_i คือพารามิเตอร์อำนาจจำแนกและพารามิเตอร์ความยากของข้อสอบ i ตามลำดับ; m_i คือจำนวนตัวเลือกของข้อสอบ i

ดังนั้นกระบวนการแก้ปัญหาสำหรับโมเดล FG3PL จึงหมายถึงรวมถึงการแก้ปัญหาของผู้สอบที่มีความรู้สมบูรณ์และการเดาแบบอิงความสามารถของผู้สอบที่มีความรู้บางส่วน ส่วนกระบวนการเดาจะหมายถึงการเดาแบบสุ่มจากผู้สอบที่ไม่มีความรู้ใดเลย (ดูภาพ 2.6) ทั้งนี้การที่การเดาแบบอิงความสามารถถูกพิจารณาให้เป็นกระบวนการแก้ปัญหาไม่ใช่กระบวนการเดา เนื่องจากถือว่าผู้สอบได้พยายามใช้ความสามารถในการแก้ปัญหาเพื่อตอบคำถามแล้ว ส่วน ICCs ของโมเดล FG3PL ยังคงมีลักษณะเหมือนกับโมเดล 3PL ทุกประการ



ภาพ 2.6 มโนทัศน์ของกระบวนการแก้ปัญหาและกระบวนการเดาในโมเดล FG3PL

ง.3) โมเดล 2-parameter logistic guessing (2PLG)

โมเดล 2PLG (Tsai & Lin, 2015) ได้รับการขยายแนวคิดมาจากโมเดล AG โดยปรับปรุงให้โมเดลมีความเป็นนัยทั่วไปขึ้น ด้วยการนำพารามิเตอร์อำนาจจำแนกเข้ามาพิจารณาในกระบวนการแก้ปัญหากลายเป็นโมเดล 2PL และปรับแก้โอกาสในการเดาถูกให้มีความสัมพันธ์กับความสามารถของผู้สอบ ความยาก และจำนวนตัวเลือกของข้อสอบ ดังสมการ (2.6)

$$P(y_{ij} = 1 | \theta_j) = \Lambda[a_i(\theta_j - b_i)] + (1 - \Lambda[a_i(\theta_j - b_i)]) \left(\frac{\exp[w_i \theta_j - b_i]}{1 + (m_i - 1) \exp[w_i \theta_j - b_i]} \right) \quad (2.6)$$

เมื่อ a_i , b_i และ m_i คือพารามิเตอร์อำนาจจำแนก ความยาก และจำนวนตัวเลือกของข้อสอบ i ; w_i คือค่าน้ำหนักของความสามารถของผู้สอบที่ใช้ในการเดาตามความยากของข้อสอบ i

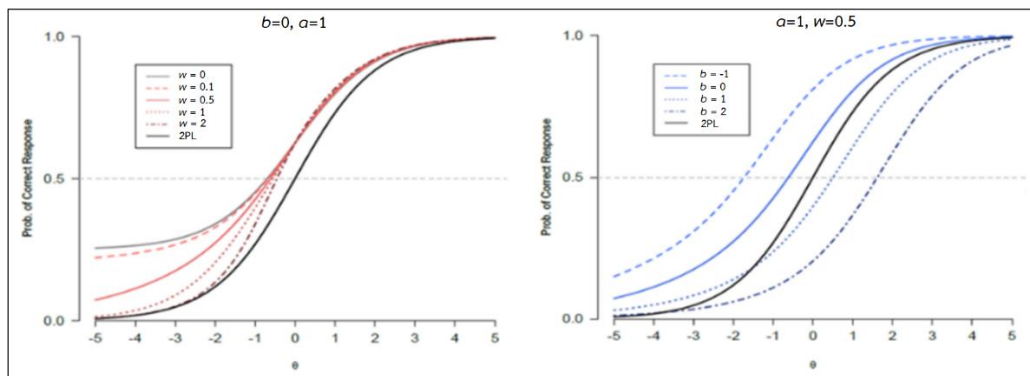
จากสมการ (2.6) โอกาสในการเดาถูกจะขึ้นอยู่กับค่าพารามิเตอร์ความยาก (b) และสัมพันธ์กับความสามารถของผู้สอบ (θ) ด้วยขนาดเท่ากับ w ดังนั้น w และ b จึงทำหน้าที่เป็นตัวกำหนดรูปแบบของการเดาที่เป็นไปได้ 2 กรณี ดังนี้

กรณี 1: $w = 0$ และ $b = 0$ → การเดาแบบสุ่ม

กรณี 2: $w > 0$ และ b มีค่าใด ๆ → การเดาแบบอิงความสามารถ

Tsai และ Lin (2015) ได้ศึกษาเปรียบเทียบโมเดล 2PLG และ 2PL เพื่อวิเคราะห์ข้อสอบปรนัยแบบ 4 ตัวเลือก โดยแบ่งการศึกษาเป็น 2 กรณีคือ 1) การเปลี่ยนค่าน้ำหนัก (w) และ 2) การเปลี่ยนค่าความยาก (b) ผลการศึกษากรณี 1 พบว่าเมื่อกำหนด $a = 1$, $b = 0$ และ $w = 0$ หางซ้ายของโมเดล 2PLG จะเข้าสู่ .25 การเดาที่เกิดขึ้นจึงมีแนวโน้มเป็นการเดาแบบสุ่ม แต่เมื่อ w มีค่า

สูงขึ้น ทางซ้ายของโมเดล 2PLG จะลดต่ำเข้าใกล้ ICCs ของโมเดล 2PL ในขณะที่ทางขวาจะสูงกว่า ICCs ของโมเดล 2PL ซึ่งลักษณะดังกล่าวมีรูปแบบเดียวกับโมเดล AG ดังได้กล่าวไปข้างต้น การเตาที่เกิดขึ้นจึงมีแนวโน้มเป็นการเตาแบบอิงความสามารถ ดังภาพ 2.7 (ซ้าย) ส่วนผลการศึกษาระดับ 2 พบว่าเมื่อ b มีการเปลี่ยนแปลง ความยากของข้อสอบจะสัมพันธ์กับโอกาสในการตอบถูกแตกต่างกัน ตามระดับความสามารถของผู้สอบ กล่าวคือเมื่อข้อสอบยากมากขึ้น ผู้สอบจำเป็นต้องมีความสามารถที่สูงขึ้นด้วยจึงจะมีโอกาสในการตอบถูกเท่าเดิม ดังภาพ 2.7 (ขวา)



ภาพ 2.7 ตัวอย่าง ICCs ของโมเดล 2PLG เมื่อ w (ซ้าย) และ b (ขวา) มีการเปลี่ยนแปลง (ที่มา: Tsai & Lin, 2015)

ง.4) โมเดล 3-parameter residual heteroscedasticity (3P-RH)

โมเดล 3P-RH (Lee & Bolt, 2018) ได้รับการขยายแนวคิดมาจากโมเดล 2PNO (two-parameter normal ogive) โดยผ่อนคลายให้ความแปรปรวนของความคลาดเคลื่อนของการประมาณค่าความสามารถของผู้สอบแต่ละระดับจากผลการตอบรายข้อ ($\sigma_{\epsilon_i|\theta}$) มีความแตกต่างกันได้ (heteroscedasticity) แตกต่างจากโมเดล 2PNO ที่กำหนดว่าความแปรปรวนดังกล่าวมีค่าคงที่สำหรับทุกระดับความสามารถของผู้สอบ การเตาสำหรับโมเดล 3P-RH จึงสามารถพิจารณาได้จากความโน้มเอียงหรือความเบ้ของผลการตอบถูกของผู้สอบแต่ละระดับ ดังสมการ (2.7)

$$P(y_{ij} = 1 | \theta_j) = \Phi \left(\frac{\alpha_i \theta_j + \beta_i}{\sqrt{2} (1 + \exp[-\delta_i \theta_j])^{-1/2}} \right) \quad (2.7)$$

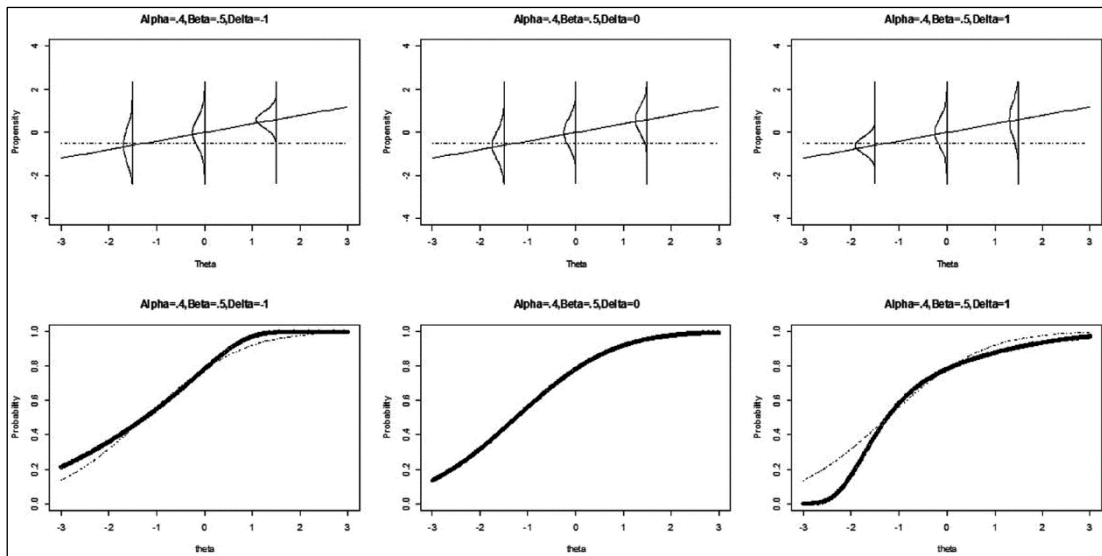
เมื่อ α_i คือพารามิเตอร์อำนาจจำแนกของข้อสอบ i ; $\alpha_i = a_i$ (พารามิเตอร์อำนาจจำแนกบนสเกลโลจิส); β_i คือพารามิเตอร์ความง่ายของข้อสอบ i ; $\beta_i = -a_i b_i$ (b_i คือพารามิเตอร์ความยากบนสเกลโลจิส); δ_i คือพารามิเตอร์ความเบ้ของผลการตอบถูกในข้อสอบ i ; $\delta_i \in (-\infty, \infty)$

จากสมการ (2.7) δ จึงเป็นตัวกำหนดรูปแบบการตอบ 3 กรณี (ดูภาพ 2.8) ดังนี้

กรณี 1: $\delta < 0$ (เบ้ลบ) \rightarrow มีการเตาเกิดขึ้น (แต่ไม่ทราบรูปแบบ)

กรณี 2: $\delta = 0$ (สมมาตร) \rightarrow ไม่มีการเตา (โมเดล 2PNO)

กรณี 3: $\delta > 0$ (เบ้บวก) \rightarrow ข้อสอบซับซ้อนสูง น่าจะใช้การแก้ปัญหาหลายขั้นตอน



ภาพ 2.8 ตัวอย่าง ICCs ของโมเดล 3P-RH (เส้นทึบ) กับ 2PNO (เส้นประ)

(ที่มา: Lee & Bolt, 2018)

ง.5) โมเดล 2-parameter logistic extension (2PLE)

โมเดล 2PLE (Zhu et al., 2019) ได้รับการปรับปรุงมาจากโมเดล 3PL ให้มีคุณสมบัติในการจำแนกการเดาแบบอิงความและการเดาแบบสุ่มออกจากกันได้ ด้วยการกำหนดโอกาสในการเดาถูกให้มีความสัมพันธ์กับความสามารถของผู้สอบและคุณลักษณะของข้อสอบ ได้แก่ ความยากอำนาจจำแนก และจำนวนตัวเลือกของข้อสอบ ดังสมการ (2.8)

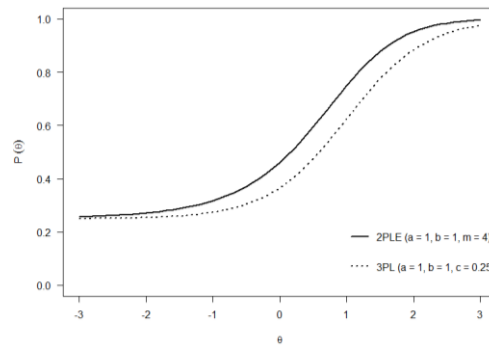
$$P(y_{ij}=1|\theta_j) = g_{ij} + (1-g_{ij}) \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]}; g_{ij} = \left(\frac{1}{m_i}\right)^{\frac{1}{1 + \exp[a_i(\theta_j - b_i)]}} \quad (2.8)$$

เมื่อ a_i , b_i และ m_i คือพารามิเตอร์อำนาจจำแนก ความยาก และจำนวนตัวเลือกของข้อสอบ i

จากสมการ (2.8) โอกาสในการเดาถูก (g) จะขึ้นอยู่กับความสามารถของผู้สอบ (θ) อำนาจจำแนก (a) ความยาก (b) และจำนวนตัวเลือกของข้อสอบ (m) ผลจากการนิยามดังกล่าว โอกาสในการเดาถูกจึงมีความผันแปรตามคุณลักษณะของผู้สอบและข้อสอบ แตกต่างจากโอกาสในการเดาถูกของโมเดล 3PL ที่ถือเป็นคุณลักษณะของข้อสอบเท่านั้น การแปลผลการเดาของโมเดล 2PLE จึงมีแนวโน้มสอดคล้องกับธรรมชาติของการตอบข้อสอบมากกว่า และยังเป็นทางเลือกปัญหาที่อาจเกิดขึ้นจากการประมาณค่าพารามิเตอร์การเดา (c) โดยตรงด้วย ทั้งนี้ θ จะทำหน้าที่เป็นตัวกำหนดรูปแบบของการเดาที่เป็นไปได้ 2 กรณี (ดูภาพ 2.9) ดังนี้

กรณี 1: θ เข้าใกล้ $-\infty \rightarrow$ การเดาแบบสุ่ม ($g \approx 1/m$ ในที่นี้กำหนด $m = 4$)

กรณี 2: $-\infty < \theta < \infty \rightarrow$ การเดาแบบอิงความสามารถ ($1/m < g < 1$)



ภาพ 2.9 ตัวอย่าง ICCs ของโมเดล 2PLE (เส้นทึบ) และ 3PL (เส้นประ)

จากความหลากหลายของปรัชญาความเชื่อเกี่ยวกับการเดา โมเดล IRT ทางเลือกฯ จึงมีรูปแบบและแนวคิดที่แตกต่างกัน โดยสรุปเป็นประเด็นที่น่าสนใจได้ 8 ประเด็น ดังนี้ 1) ปัจจัยที่ส่งผลต่อการเดา 2) พารามิเตอร์ที่ต้องประมาณค่าในโมเดล 3) วิธีการประมาณค่าในงานวิจัย 4) โปรแกรมวิเคราะห์ 5) สถิติที่ใช้เป็นเกณฑ์เปรียบเทียบผลการประมาณค่าและการคัดเลือกโมเดลในงานวิจัย 6) ผลสรุปในงานวิจัย 7) จุดเด่นของโมเดล และ 8) ข้อจำกัดของโมเดล รายละเอียดดังตาราง 2.3

ตาราง 2.3 การเปรียบเทียบระหว่างโมเดล AG, FG3PL, 2PLG, 3P-RH และ 2PLE

ประเด็น	โมเดล AG (2006)	โมเดล FG3PL (2012)	โมเดล 2PLG (2015)	โมเดล 3P-RH (2018)	โมเดล 2PLE (2019)	
1) ปัจจัยที่ส่งผลต่อการเดา	- ความสามารถผู้สอบ	- จำนวนตัวเลือก	- ความสามารถผู้สอบ - ความยากข้อสอบ - จำนวนตัวเลือก	- ความเข้มของการตอบถูก	- ความสามารถผู้สอบ - ความยากข้อสอบ - อำนาจจำแนกข้อสอบ - จำนวนตัวเลือก	
2) พารามิเตอร์ที่ต้องประมาณค่า	b = ความยาก g = การเดา α = ค่าน้ำหนักของความสามารถที่ใช้ในการเดา	b = ความยาก a = อำนาจจำแนก	b = ความยาก a = อำนาจจำแนก w = ค่าน้ำหนักของความสามารถที่ใช้ในการเดา	β = ความง่าย α = อำนาจจำแนก δ = ความเข้มของการตอบถูก	b = ความยาก a = อำนาจจำแนก	
3) วิธีการประมาณค่า	Bayes (MCMC), NLMIXED	ML (EM)	Bayes (MCMC)	Bayes (MCMC)	ML (Hermite-Gauss quadrature)	
4) โปรแกรม	SAS (NLMIXED), WINBUGS	PARSCALE, PARAM-3PL, WinGen	R (R2jags)	R (optim), Mx, SAS	ไม่ระบุ	
5) สถิติที่ใช้พิจารณา	- การประมาณค่า - การคัดเลือกโมเดล	Confidence Band -2LL, AIC, BIC	Bias, RMSE χ^2	Credible interval ไม่ระบุ	Correlation -2LL, AIC, BIC, DIC	Bias, RMSE AIC, BIC
6) ผลสรุปในงานวิจัย	- โมเดล AG ประมาณค่าพารามิเตอร์ได้ดี และสอดคล้องกับข้อมูลเชิงประจักษ์	- โมเดล FG3PL ประมาณค่าพารามิเตอร์ได้ดี และใช้ง่ายกว่าโมเดล 3PL	- โมเดล 2PLG ประมาณค่าพารามิเตอร์ได้ดี	- โมเดล 3P-RH ประมาณค่าพารามิเตอร์ได้ดี และมีแนวโน้มสอดคล้องกับข้อมูลมากกว่าโมเดล 3PL	- โมเดล 2PLE ประมาณค่าพารามิเตอร์ได้ดี และมีแนวโน้มสอดคล้องกับข้อมูลมากกว่าโมเดล 3PL	
7) จุดเด่น	- เป็นโมเดลขยายของ Rasch - ระบุรูปแบบการเดาได้	- เป็นโมเดลเฉพาะของ 3PL - มีพารามิเตอร์น้อยกว่า 3PL แต่ให้ผลการวิเคราะห์แทบจะไม่แตกต่างจาก 3PL	- เป็นโมเดลขยายของ 2PL - ระบุรูปแบบการเดาได้	- เป็นโมเดลขยายของ 2PNO - แปลผลการเดาได้อย่างอิสระ - บอกความซับซ้อนในการทำข้อสอบแต่ละข้อได้	- เป็นโมเดลขยายของ 2PL - เป็นโมเดลเฉพาะของ 3PL - ระบุรูปแบบการเดาได้	
8) ข้อจำกัด	- ไม่มีค่าอำนาจจำแนก - พารามิเตอร์ g อยู่บนสเกลโลจิส จึงนำมาเทียบกับพารามิเตอร์ c ในโมเดล 3PL ไม่ได้โดยตรง	- การเดาเป็นแบบสุ่มเท่านั้น จึงไม่สอดคล้องกับสภาพธรรมชาติเท่าที่ควร	- โมเดลซับซ้อน - การเดาแบบสุ่มเกิดขึ้นเมื่อข้อสอบยากปานกลาง ($b = 0$) จึงไม่สอดคล้องกับสภาพธรรมชาติเท่าที่ควร	- โมเดลซับซ้อน เนื่องจากอยู่ในรูป normal ogive - ระบุรูปแบบการเดาไม่ได้	- ไม่สามารถอธิบายกรณี $g < 1/m$ (โอกาสการเดาถูกน้อยกว่าโอกาสการเดาถูกแบบสุ่ม) ซึ่งอาจเกิดขึ้นเมื่อผู้สอบความสามารถต่ำเจอตัวลวงที่มีคุณภาพ	

หมายเหตุ: MCMC=Markov chain Monte Carlo; ML=Maximum likelihood; EM=Expectation-maximization; LL=Log-likelihood; AIC=Akaike information criterion; BIC=Bayesian information criterion; DIC=Deviance information criterion

2.2 โมเดลกลุ่มแฝง (LC)

โมเดลกลุ่มแฝง (LC: latent class) มีแนวคิดคล้ายกับการวิเคราะห์องค์ประกอบ (FA: factor analysis) แต่มีจุดแตกต่างตรงที่ตัวแปรแฝงในโมเดล FA เป็นตัวแปรต่อเนื่อง ส่วนตัวแปรแฝงในโมเดล LC เป็นตัวแปรจัดกลุ่ม และอาจมองความแตกต่างในแง่ของการจัดกลุ่มได้ว่าโมเดล FA เป็นการจัดกลุ่มให้กับตัวแปร (variables) ในขณะที่โมเดล LC เป็นการจัดกลุ่มให้กับบุคคล (persons)

เป้าหมายของโมเดล LC คือการจัดกลุ่มให้กับบุคคลที่มีแบบแผนการตอบคล้ายคลึงกัน เพื่อจำแนกบุคคลออกเป็นกลุ่มย่อย ผลการวิเคราะห์จะให้สารสนเทศเกี่ยวกับจำนวนกลุ่มแฝงของข้อมูล สัดส่วนของสมาชิกในกลุ่มแฝง และอาจรวมถึงปัจจัยที่ใช้ในการจำแนกกลุ่ม โดยมีข้อตกลงสำคัญ 2 ประการคือ 1) ความเป็นอิสระ กล่าวคือความน่าจะเป็นของแบบแผนการตอบข้อสอบทั้งหมด I ข้อของผู้สอบ j ซึ่งอยู่ในกลุ่มแฝง k จะเท่ากับผลคูณของความน่าจะเป็นของผลการตอบข้อสอบรายข้อของผู้สอบในกลุ่มแฝงนั้น หรือ $P(\underline{y}_j | k) = \prod_{i=1}^I P(y_{ij} | k)$ และ 2) สมาชิกในกลุ่มแฝงต้องมาจากประชากรกลุ่มเดียวกัน กล่าวคือผลรวมของสัดส่วนของสมาชิกในแต่ละกลุ่มแฝงต้องมีค่าเป็น 1 หรือ $\sum_{k=1}^K \pi_k = 1$ เมื่อ π_k คือสัดส่วนของสมาชิกในกลุ่มแฝง k และ $0 < \pi_k < 1$

รูปแบบทั่วไปของโมเดล LC (Tay et al., 2011) สามารถแสดงได้ดังสมการ (2.9)

$$P(\underline{y}_j) = \sum_{k=1}^K \pi_k P(\underline{y}_j | k) \quad (2.9)$$

เมื่อ \underline{y}_j คือเวกเตอร์แบบแผนการตอบข้อสอบทั้งหมด I ข้อของผู้สอบ j ; $\underline{y}_j = (y_{1j} \ y_{2j} \ \dots \ y_{Ij})$; π_k คือสัดส่วนของสมาชิกในกลุ่มแฝง k ; $P(\underline{y}_j | k)$ คือความน่าจะเป็นแบบมีเงื่อนไขของแบบแผนการตอบของผู้สอบ j ซึ่งอยู่ในกลุ่มแฝง k เมื่อ $P(\underline{y}_j | k) = \prod_{i=1}^I P(y_{ij} | k)$

การประมาณค่าพารามิเตอร์สำหรับโมเดล LC โดยทั่วไปนิยมใช้วิธีภาวะควรจะเป็นสูงสุด (maximum likelihood) ด้วยขั้นตอนวิธี Newton-Raphson หรือ expectation-maximization (EM) ซึ่งเป็นขั้นตอนวิธีทวนซ้ำ (iterative algorithm) เหมือนกัน อย่างไรก็ตามวิธี EM สามารถให้ค่าประมาณที่คงเส้นคงวามากกว่าเมื่อข้อมูลมีการสูญหาย (Wilks, 2011) จึงได้รับความนิยมใช้อย่างแพร่หลายมากกว่า ทั้งนี้ขั้นตอนวิธีทวนซ้ำดังกล่าวจะประกอบด้วย 3 ขั้นตอน คือ 1) แบ่งบุคคลออกเป็นกลุ่มแฝงโดยใช้การสุ่ม 2) จัดกลุ่มแฝงใหม่นำมาเปรียบเทียบกับกลุ่มเก่า โดยทำการคัดเลือกโมเดลจากเกณฑ์สารสนเทศ เช่น AIC หรือ BIC หากโมเดลกลุ่มแฝงใดมี AIC หรือ BIC ต่ำกว่า จะหมายความว่าโมเดลกลุ่มแฝงนั้นมีความสอดคล้องกับข้อมูลมากกว่า และ 3) จัดกลุ่มแฝงใหม่ซ้ำจนกระทั่งได้ผลการจัดกลุ่มแฝงที่ดีที่สุด (Magidson & Vermunt, 2004)

โปรแกรมคอมพิวเตอร์ที่รองรับการวิเคราะห์โมเดล LC เช่น โปรแกรม Latent Gold (Vermunt & Magidson, 2016) โปรแกรม Mplus (Muthén & Muthén, 1998-2015) และโปรแกรม R แพคเกจ polCA (Linzer & Lewis, 2014)

2.3 โมเดลพหุระดับ (multilevel)

โมเดลพหุระดับหรือโมเดลเชิงเส้นแบบลดหลั่น (hierarchical linear model) เป็นโมเดลวิเคราะห์ที่มีการแบ่งความแปรปรวนของความคลาดเคลื่อนจากการวัดในระดับต่าง ๆ ออกจากกัน ถูกพัฒนาขึ้นเพื่อใช้แทนโมเดลถดถอยทั่วไป ซึ่งเป็นการวิเคราะห์ระดับเดียว (single-level analysis) ที่มีค่าผิดพลาดของข้อมูลเกี่ยวกับความเป็นอิสระของตัวอย่างเมื่อนำไปวิเคราะห์กับข้อมูลทางการศึกษา (Raudenbush & Bryk, 2002) เนื่องจากธรรมชาติของข้อมูลดังกล่าวมักมีโครงสร้างแบบพหุระดับ กล่าวคือข้อมูลในระดับล่างจะซ้อนอยู่ภายใต้ข้อมูลในระดับบน เช่น ข้อมูลของนักเรียนซ้อนอยู่ภายใต้โรงเรียน โรงเรียนซ้อนอยู่ภายใต้สังกัดสถานศึกษา เป็นต้น หากนำข้อมูลดังกล่าวมาวิเคราะห์ด้วยโมเดลถดถอยทั่วไปอาจทำให้เกิดข้อจำกัดคือ 1) ผลการประมาณค่าพารามิเตอร์อาจคลาดเคลื่อนไปจากค่าจริง เนื่องจากการวิเคราะห์แบบรวมหน่วยตัวอย่างทุกกลุ่มเข้าด้วยกัน มักก่อให้เกิดข้อผิดพลาดในการสรุปผลระหว่างระดับ (aggregation bias) และ 2) ผลการวิเคราะห์อาจถูกรบกวนจากตัวแปรในระดับกลุ่ม ทำให้การประมาณค่าความคลาดเคลื่อนมาตรฐานของค่าพารามิเตอร์ในโมเดลมีแนวโน้มต่ำกว่าค่าจริง (ไพรัตน์ วงษ์นาม, 2555; ศิริชัย กาญจนวาสิ, 2554)

ด้วยเหตุนี้โมเดลพหุระดับจึงถูกพัฒนาขึ้นเพื่อผ่อนคลายข้อจำกัดดังกล่าว โดยในงานวิจัยนี้จะขอเสนอเฉพาะโมเดลพหุระดับที่ไม่มีตัวแปรทำนายใด ๆ หรือโมเดลปราศจากเงื่อนไขอย่างสมบูรณ์ (fully unconditional model) ที่มี 2 ระดับคือ ระดับหน่วย (ระดับนักเรียน) และระดับกลุ่ม (ระดับโรงเรียน) เพื่อจะนำแนวคิดไปบูรณาการกับโมเดล IRT และโมเดล LC ต่อไป รายละเอียดดังนี้

โมเดลระดับที่ 1 หรือโมเดลระดับนักเรียน (student-level model) เป็นการแสดงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรต้นในระดับนักเรียน ดังสมการ (2.10)

$$y_{js} = \beta_{0s} + r_{js} \quad (2.10)$$

เมื่อ y_{js} คือตัวแปรตามซึ่งเป็นค่าสังเกตของนักเรียน j ในโรงเรียน s ; r_{js} คือค่าความคลาดเคลื่อนในระดับนักเรียน เมื่อ $r_{js} \sim N(0, \sigma^2)$; σ^2 คือความแปรปรวนภายในกลุ่ม; β_{0s} คือค่าเฉลี่ยของค่าสังเกตของนักเรียนในโรงเรียน s ซึ่งถือว่าเป็นพารามิเตอร์สุ่ม (random parameter) ที่ต้องมีการกำหนดโมเดลเพิ่มเติมเพื่ออธิบายความแปรปรวนที่เกิดขึ้นต่อไป เรียกโมเดลดังกล่าวว่าโมเดลระดับที่ 2 หรือโมเดลระดับโรงเรียน (school-level model) ดังสมการ (2.11)

$$\beta_{0s} = \gamma_{00} + u_{0s} \quad (2.11)$$

เมื่อ γ_{00} คือค่าเฉลี่ยของค่าสังเกตของนักเรียนทุกโรงเรียน; u_{0s} คือค่าความคลาดเคลื่อนในระดับโรงเรียน เมื่อ $u_{0s} \sim N(0, \tau)$; τ คือความแปรปรวนระหว่างกลุ่ม ทั้งนี้เมื่อนำสมการ (2.10) และ (2.11) มาเขียนรวมกันจะเรียกว่าโมเดลอิทธิพลผสม (mixed-effect model) ดังสมการ (2.12)

$$y_{js} = \gamma_{00} + r_{js} + u_{0s} \quad (2.12)$$

โดยอาจเรียก γ_{00} ว่าอิทธิพลคงที่ (fixed effect) และ r_{js} , u_{0s} ว่าอิทธิพลสุ่ม (random effect)

การวิเคราะห์ด้วยโมเดลพหุระดับข้างต้นจะให้สารสนเทศว่า “ข้อมูลมีความเหมาะสมสำหรับการนำมาวิเคราะห์แบบพหุระดับหรือไม่” ด้วยการพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (ICC : intraclass correlation coefficient) ซึ่งมีสูตรคำนวณดังสมการ (2.13)

$$ICC = \frac{\tau}{\tau + \sigma^2} \quad (2.13)$$

เมื่อ ICC คือค่าสถิติที่บอกว่าความแปรปรวนทั้งหมดของข้อมูลสามารถถูกอธิบายได้ด้วยข้อมูลในระดับโรงเรียนเพียงใด หาก $ICC > .05$ จะถือว่าข้อมูลมีความเหมาะสม (Hox & Maas, 2001)

โปรแกรมคอมพิวเตอร์ที่รองรับการวิเคราะห์โมเดลพหุระดับ เช่น โปรแกรม HLM (Raudenbush et al., 2019) โปรแกรม Mplus และโปรแกรม R แพคเกจ lme4 (Bates et al., 2019)

2.4 โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสม (MixIRT)

เนื่องจากโมเดล IRT ทั่วไปมีข้อตกลงเกี่ยวกับความไม่แปรเปลี่ยน ซึ่งมีสมมติฐานว่าผู้สอบมาจากประชากรที่มีภาวะเอกพันธ์กันเพียงกลุ่มเดียว อย่างไรก็ตามสมมติฐานดังกล่าวอาจเป็นการฝ่าฝืนธรรมชาติของข้อมูล เพราะโดยทั่วไปแม้ผู้สอบจะมาจากประชากรกลุ่มเดียวกัน แต่ก็อาจมีคุณลักษณะที่แตกต่างกันได้ (Tay et al., 2011) เช่น นักเรียนจากโรงเรียนเดียวกันอาจมีแบบแผนการตอบข้อสอบที่แตกต่างกัน และความแตกต่างนี้อาจมีมากจนเชื่อได้ว่าผู้สอบน่าจะมาจากประชากรกลุ่มย่อยหลายกลุ่มมากกว่า ด้วยเหตุนี้จึงมีการบูรณาการโมเดล IRT เข้ากับโมเดล LC กลายเป็นโมเดล mixture IRT (MixIRT) โดยผ่อนคลายข้อตกลงเกี่ยวกับความไม่แปรเปลี่ยนลง พารามิเตอร์ในโมเดล MixIRT จึงมีค่าผันแปรได้ตามกลุ่มย่อยหรือกลุ่มแฝงของประชากร (Rost, 1990)

เป้าหมายของโมเดล MixIRT จึงเป็นการจัดกลุ่มให้กับบุคคลที่มีแบบแผนการตอบคล้ายคลึงกัน เช่นเดียวกับโมเดล LC แต่มีจุดแตกต่างตรงที่โมเดล MixIRT มีการให้สารสนเทศเชิงปริมาณเกี่ยวกับคุณลักษณะแฝงของผู้สอบและข้อสอบตามแนวคิดของโมเดล IRT ด้วย (Mislevy & Verhelst, 1990) จากความครอบคลุมของการให้สารสนเทศ โมเดล MixIRT จึงมักถูกนำไปใช้เพื่อแก้ปัญหาทางการทดสอบ เช่น การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Cohen & Bolt, 2005) การศึกษาทวิวิธีการตอบข้อสอบ (Bolt et al., 2001) การตรวจสอบความเร็วในการทำแบบสอบ (Bolt et al., 2002) ทั้งนี้การแบ่งรูปแบบของโมเดล MixIRT สามารถพิจารณาได้ในทำนองเดียวกับโมเดล IRT ดังนี้

ก) โมเดล mixture Rasch (MixRasch)

โมเดล MixRasch เป็นการบูรณาการระหว่างโมเดล Rasch กับโมเดล LC โดยมีข้อตกลงว่า 1) สมาชิกในกลุ่มแฝงต้องมาจากประชากรกลุ่มเดียวกัน และ 2) โมเดล Rasch มีความสอดคล้องกับข้อมูลในกลุ่มแฝง โดยกลุ่มแฝงแต่ละกลุ่มจะมีพารามิเตอร์ความยากของข้อสอบเพียงชุดเดียว เพื่อแสดงนัยว่าสมาชิกในกลุ่มแฝงเดียวกันมีแบบแผนการตอบที่คล้ายคลึงกัน (Cohen et al., 2002)

จากการบูรณาการระหว่างโมเดล LC และ IRT สามารถกำหนดความน่าจะเป็นของแบบแผนการตอบข้อสอบได้เป็น $P(y_j) = \sum_{k=1}^K \pi_k \left(\prod_{i=1}^I P(y_{ij} | k, \theta_{jk}) \right)$ เมื่อ $P(y_{ij} | k, \theta_{jk})$ คือความน่าจะเป็นแบบมีเงื่อนไขของผลการตอบข้อสอบ i ของผู้สอบ j ที่อยู่ในกลุ่มแฝง k และมีความสามารถ θ_{jk} ซึ่งค่าความน่าจะเป็นดังกล่าวสามารถอธิบายได้ด้วยโมเดล Rasch ดังนั้นความน่าจะเป็นของแบบแผนการตอบข้อสอบสำหรับโมเดล MixRasch จึงมีรูปแบบดังสมการ (2.14)

$$P(y_j) = \sum_{k=1}^K \pi_k \prod_{i=1}^I \frac{1}{1 + \exp[-y_{ij} (\theta_{jk} - b_{ik})]} \quad (2.14)$$

ดังนั้นฟังก์ชันการตอบสนองข้อสอบสำหรับโมเดล MixRasch จึงมีรูปแบบดังสมการ (2.15)

$$P(y_{ij} = 1 | \theta_{jk}) = \sum_{k=1}^K \pi_k \frac{1}{1 + \exp[-(\theta_{jk} - b_{ik})]} \quad (2.15)$$

เมื่อ π_k คือสัดส่วนของสมาชิกในกลุ่มแฝง k ; θ_{jk} คือพารามิเตอร์ความสามารถของผู้สอบ j ในกลุ่มแฝง k เมื่อ $\theta_{jk} \sim N(\mu_k, \sigma_k^2)$; μ_k และ σ_k^2 คือค่าเฉลี่ยและความแปรปรวนของความสามารถในกลุ่มแฝง k ตามลำดับ; b_{ik} คือพารามิเตอร์ความยากของข้อสอบ i ในกลุ่มแฝง k

ข) โมเดล mixture 2PL (Mix2PL)

โมเดล Mix2PL ได้รับการขยายแนวคิดมาจากโมเดล MixRasch โดยผ่อนคลายให้พารามิเตอร์อำนาจจำแนกของข้อสอบมีค่าแตกต่างกันได้ตามกลุ่มแฝง ภายใต้ข้อตกลงว่าโมเดล 2PL มีความสอดคล้องกับข้อมูลในกลุ่มแฝง โมเดล Mix2PL จึงมีรูปแบบดังสมการ (2.16)

$$P(y_{ij} = 1 | \theta_{jk}) = \sum_{k=1}^K \pi_k \frac{1}{1 + \exp[-a_{ik}(\theta_{jk} - b_{ik})]} \quad (2.16)$$

เมื่อ a_{ik} คือพารามิเตอร์อำนาจจำแนกของข้อสอบ i ในกลุ่มแฝง k

ค) โมเดล mixture 3PL (Mix3PL)

โมเดล Mix3PL ได้รับการขยายแนวคิดมาจากโมเดล Mix2PL ด้วยการเพิ่มพารามิเตอร์การเดาให้มีค่าแตกต่างกันได้ตามกลุ่มแฝง ภายใต้ข้อตกลงว่าโมเดล 3PL มีความสอดคล้องกับข้อมูลในกลุ่มแฝง โมเดล Mix3PL จึงมีรูปแบบดังสมการ (2.17)

$$P(y_{ij} = 1 | \theta_{jk}) = \sum_{k=1}^K \pi_k \left(c_{ik} + (1 - c_{ik}) \frac{1}{1 + \exp[-a_{ik}(\theta_{jk} - b_{ik})]} \right) \quad (2.17)$$

เมื่อ c_{ik} คือพารามิเตอร์การเดาของข้อสอบ i ในกลุ่มแฝง k

การประมาณค่าพารามิเตอร์สำหรับโมเดล MixIRT โดยทั่วไปจะใช้วิธีการประมาณเพื่อหาสัดส่วนของสมาชิกในกลุ่มแฝงและคุณลักษณะแฝงไปพร้อมกัน (simultaneous estimation) ซึ่งสามารถทำได้ 2 วิธีคือ 1) วิธีภาวะควรจะเป็นสูงสุดด้วยขั้นตอนวิธี EM (Rost, 1990) และ 2) วิธีการประมาณค่าแบบเบย์ส์ด้วยขั้นตอนวิธี MCMC ซึ่งเป็นการสุ่มตัวอย่างจากการแจกแจงความน่าจะเป็น

ของลูกโซ่มาร์คอฟ ซึ่งมีคุณสมบัติในการให้การแจกแจงภายหลัง (posterior distribution) มีลักษณะแบบเดียวกับการแจกแจงก่อนหน้า (prior distribution) ตัวอย่างการกำหนดการแจกแจงก่อนหน้าของพารามิเตอร์สำหรับโมเดล Mix3PL (Choi, 2014) เช่น

$$\text{พารามิเตอร์ผู้สอบ: } \theta_{jk} \sim N(\mu_k, 1); \mu_k \sim N(0, 1)$$

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

$$\text{พารามิเตอร์กลุ่มแฝง: } (\pi_1, \dots, \pi_K) \sim \text{Dir}(0.5, \dots, 0.5)$$

เมื่อ $j = 1, 2, \dots, J$; $k = 1, 2, \dots, K$; $i = 1, 2, \dots, I$

การคัดเลือกโมเดลสำหรับโมเดล MixIRT โดยทั่วไปนิยมพิจารณาจากเกณฑ์สารสนเทศ BIC เนื่องจากผลการศึกษาในหลายงานวิจัยพบว่า BIC มีแนวโน้มสามารถคัดเลือกโมเดลที่มีจำนวนกลุ่มแฝงได้สอดคล้องกับข้อมูลมากที่สุด (Choi, 2014; Li et al., 2009; Sen, 2014)

โปรแกรมคอมพิวเตอร์ที่รองรับการวิเคราะห์โมเดล MixIRT เช่น โปรแกรม WINMIRA (von Davier, 2001) และโปรแกรม R แพคเกจ mixRasch (Willse, 2014) สำหรับโมเดล MixRasch โปรแกรม Latent Gold และโปรแกรม Mplus สำหรับโมเดล MixIRT ทั่วไป

2.5 โมเดลทฤษฎีการตอบสนองข้อสอบพหุระดับ (multilevel IRT)

จากที่กล่าวไปในตอนต้นแล้วว่าธรรมชาติของข้อมูลทางการศึกษามักเป็นข้อมูลพหุระดับ แต่เนื่องจากโมเดล IRT ทั่วไปยังไม่มีคำนึงถึงโครงสร้างพหุระดับของข้อมูล ผลการวัดในระดับหน่วย (นักเรียน) จึงอาจถูกรบกวนจากตัวแปรในระดับกลุ่ม (โรงเรียน) ทำให้เกิดความคลาดเคลื่อนขึ้นในการประมาณค่า ด้วยเหตุนี้โมเดล multilevel IRT จึงถูกพัฒนาขึ้นเพื่อผ่อนคลายปัญหาดังกล่าว

แนวความคิดพัฒนาโมเดล multilevel IRT มีผู้เสนอไว้หลายวิธี เช่น Fox และ Glas (2001) Kamata (2001) และ Maier (2001) อย่างไรก็ตามในงานวิจัยนี้ขอยกตัวอย่างแนวความคิดพัฒนาโมเดล multilevel IRT แบบ 3 ระดับของ Kamata (2001) เท่านั้น รายละเอียดดังนี้

ก) โมเดล multilevel Rasch

โมเดล multilevel Rasch จะมีการเพิ่มโมเดล Rasch เข้ามาในโมเดลพหุระดับอีกหนึ่งระดับ เรียกโมเดลนี้ว่าโมเดลระดับข้อสอบ (item-level model) ดังสมการ (2.18)

$$P(y_{ijs} = 1 | \theta_{js}) = \frac{1}{1 + \exp[-(\theta_{js} - b_{ijs})]} \quad (2.18)$$

เมื่อ y_{ijs} คือผลการตอบข้อสอบ i ของนักเรียน j ในโรงเรียน s ; θ_{js} คือพารามิเตอร์ความสามารถของนักเรียน j ในโรงเรียน s ; b_{ijs} คือพารามิเตอร์ความยากของข้อสอบ i ซึ่งแปรผันตามหน่วยตัวอย่างในระดับนักเรียนและโรงเรียน ทั้งนี้ θ_{js} และ b_{ijs} จะถือว่าเป็นพารามิเตอร์สุ่ม ซึ่งต้องกำหนดโมเดลเพื่ออธิบายความแปรปรวนที่เกิดขึ้นต่อไป เรียกว่าโมเดลระดับนักเรียน ดังสมการ (2.19)

$$\theta_{js} = \gamma_{0s} + r_{js} \quad \text{และ} \quad b_{ijs} = b_{is} \quad (2.19)$$

เมื่อ γ_{0s} คือความสามารถเฉลี่ยของนักเรียนในโรงเรียน s ; r_{js} คือค่าความคลาดเคลื่อนในระดับนักเรียน เมื่อ $r_{js} \sim N(0, \tau)$; τ คือความแปรปรวนของความสามารถในระดับนักเรียน ส่วน b_{ijs} จะกลายเป็น b_{is} เนื่องจากพารามิเตอร์ความยากเป็นคุณลักษณะของข้อสอบ จึงมีค่าคงที่สำหรับนักเรียนทุกคน ทั้งนี้ γ_{0s} และ b_{is} จะถือว่าเป็นพารามิเตอร์สุ่ม ซึ่งต้องกำหนดโมเดลเพื่ออธิบายความแปรปรวนที่เกิดขึ้นต่อไป เรียกว่าโมเดลระดับโรงเรียน ดังสมการ (2.20)

$$\gamma_{0s} = \gamma_{00} + u_{0s} \quad \text{และ} \quad b_{is} = b_i \quad (2.20)$$

เมื่อ γ_{00} คือความสามารถเฉลี่ยของนักเรียนทุกโรงเรียน; u_{0s} คือค่าความคลาดเคลื่อนในระดับโรงเรียน เมื่อ $u_{0s} \sim N(0, \zeta)$; ζ คือความแปรปรวนของความสามารถในระดับโรงเรียน ส่วน b_{is} จะกลายเป็น b_i เนื่องจากพารามิเตอร์ความยากเป็นคุณลักษณะของข้อสอบ จึงมีค่าคงที่สำหรับทุกโรงเรียน และอาจพิจารณาให้ $\gamma_{0s} = u_{0s}$ เมื่อ $u_{0s} \sim N(\gamma_{00}, \zeta)$ โมเดล multilevel Rasch จึงมีรูปแบบดังสมการ (2.21)

$$P(y_{ijs} = 1 | r_{js}) = \frac{1}{1 + \exp\left[-\left(u_{0s} + r_{js}\right) - b_i\right]} \quad (2.21)$$

เมื่อ u_{0s} คือค่าความคลาดเคลื่อนในระดับโรงเรียน เมื่อ $u_{0s} \sim N(\gamma_{00}, \zeta)$; r_{js} คือค่าความคลาดเคลื่อนในระดับนักเรียน เมื่อ $r_{js} \sim N(0, \tau)$; b_i คือพารามิเตอร์ความยากของข้อสอบ i

ข) โมเดล multilevel 2PL

โมเดล multilevel 2PL มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (2.22)

$$P(y_{ijs} = 1 | r_{js}) = \frac{1}{1 + \exp\left[-a_i \left(u_{0s} + r_{js}\right) - b_i\right]} \quad (2.22)$$

เมื่อ a_i คือพารามิเตอร์อำนาจจำแนกของข้อสอบ i

ค) โมเดล multilevel 3PL

โมเดล multilevel 3PL มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (2.23)

$$P(y_{ijs} = 1 | r_{js}) = c_i + (1 - c_i) \frac{1}{1 + \exp\left[-a_i \left(u_{0s} + r_{js}\right) - b_i\right]} \quad (2.23)$$

เมื่อ c_i คือพารามิเตอร์การเดาของข้อสอบ i

ส่วน ICC สำหรับโมเดล multilevel IRT มีสูตรคำนวณดังสมการ (2.24)

$$ICC = \frac{\zeta}{\zeta + \tau} \quad (2.24)$$

เมื่อ τ และ ζ คือความแปรปรวนของความสามารถในระดับนักเรียนและระดับโรงเรียน ตามลำดับ ส่วนความแปรปรวนของคลาดเคลื่อนจากการวัดในระดับข้อสอบจะไม่ถูกนำมาพิจารณา เนื่องจากค่าดังกล่าวได้ถูกรวมอยู่ในโมเดล IRT ซึ่งเป็นโมเดลการวัดคุณลักษณะแฝงแล้ว (Fox & Glas, 2001)

โปรแกรมคอมพิวเตอร์ที่รองรับการวิเคราะห์โมเดล multilevel IRT เช่น โปรแกรม Mplus โปรแกรม R แพคเกจ lme4 และโปรแกรม MultiBUGS

2.6 โมเดลกลุ่มแฝงพหุระดับ (multilevel LC)

ข้อจำกัดหนึ่งของโมเดล LC คือมีการละเลยโครงสร้างพหุระดับของข้อมูล ด้วยเหตุนี้โมเดล multilevel LC (Vermunt, 2003) จึงถูกพัฒนาขึ้นเพื่อผ่อนคลายข้อจำกัดดังกล่าว

จากการบูรณาการระหว่างโมเดล LC และโมเดลพหุระดับ สามารถกำหนดความน่าจะเป็นของแบบแผนการตอบข้อสอบในระดับโรงเรียนได้เป็น $P(\underline{y}_s) = \sum_{g=1}^G \pi_g \left(\prod_{j=1}^{n_s} P(y_{js} | g) \right)$ เมื่อ $P(y_{js} | g)$ คือความน่าจะเป็นแบบมีเงื่อนไขของแบบแผนการตอบข้อสอบของนักเรียน j ในโรงเรียน s ซึ่งอยู่ในกลุ่มแฝงระดับโรงเรียน g และในทำนองเดียวกันสามารถกำหนดความน่าจะเป็นของแบบแผนการตอบข้อสอบในระดับนักเรียนได้เป็น $P(\underline{y}_{js} | g) = \sum_{k=1}^K \pi_{k|g} \left(\prod_{i=1}^I P(y_{ijs} | k, g) \right)$ เมื่อ $P(y_{ijs} | k, g)$ คือความน่าจะเป็นแบบมีเงื่อนไขของผลการตอบข้อสอบ i ของนักเรียน j ในโรงเรียน s ซึ่งอยู่ในกลุ่มแฝงระดับนักเรียน k ในกลุ่มแฝงระดับโรงเรียน g

รูปแบบทั่วไปของโมเดล multilevel LC จึงมีรูปแบบดังสมการ (2.25)

$$P(\underline{y}_s) = \sum_{g=1}^G \pi_g \left[\prod_{j=1}^{n_s} \left\{ \sum_{k=1}^K \pi_{k|g} \left(\prod_{i=1}^I P(y_{ijs} | k, g) \right) \right\} \right] \quad (2.25)$$

เมื่อ \underline{y}_s คือเวกเตอร์แบบแผนการตอบข้อสอบทั้งหมด I ข้อ ของนักเรียน n_s คน ในโรงเรียน s ; π_g คือสัดส่วนของสมาชิกในกลุ่มแฝงระดับโรงเรียน g ; $\pi_{k|g}$ คือสัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียน k ในกลุ่มแฝงระดับโรงเรียน g ; $P(y_{ijs} | k, g)$ คือความน่าจะเป็นแบบมีเงื่อนไขของผลการตอบข้อสอบ i ของนักเรียน j ในโรงเรียน s ซึ่งอยู่ในกลุ่มแฝงต่าง ๆ

การคัดเลือกโมเดลสำหรับโมเดล multilevel LC โดยทั่วไปนิยมพิจารณาจากเกณฑ์สารสนเทศ BIC เช่นเดียวกับโมเดล LC (Henry & Muthén, 2010)

โปรแกรมคอมพิวเตอร์ที่รองรับการวิเคราะห์โมเดล multilevel LC เช่น โปรแกรม Latent Gold โปรแกรม Mplus และโปรแกรม MultiBUGS

2.7 โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ (MMixIRT)

จากการบูรณาการระหว่างสองโมเดลทำให้เกิดโมเดลใหม่ ได้แก่ 1) MixIRT 2) multilevel IRT และ 3) multilevel LC อย่างไรก็ตามโมเดลเหล่านี้อาจมีข้อจำกัดบางประการดังนี้

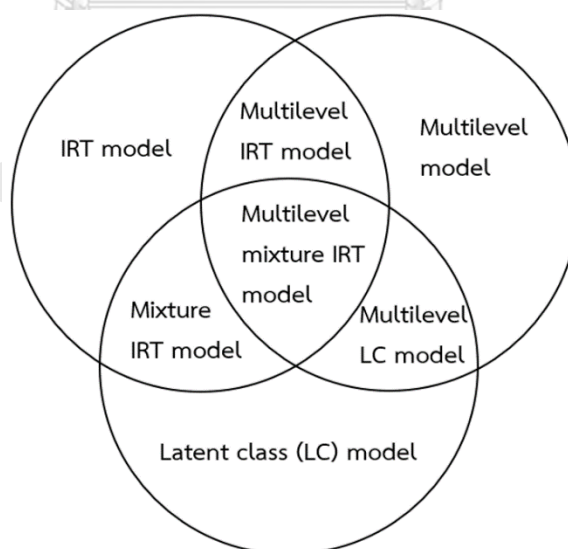
ข้อจำกัดของโมเดล MixIRT คือโมเดลมีการละเลยต่อโครงสร้างพหุระดับของข้อมูล ซึ่งอาจส่งผลให้การคัดเลือกโมเดลของเกณฑ์สารสนเทศบางตัวขาดความถูกต้อง และการประมาณค่าพารามิเตอร์อาจคลาดเคลื่อนไปจากค่าจริง (Lee et al., 2018; Sen et al., 2019)

ข้อจำกัดของโมเดล multilevel IRT คือโมเดลยังไม่มีคำนึงถึงความวิวิธพันธุ์ของประชากร ซึ่งอาจเกิดขึ้นได้ตามธรรมชาติของข้อมูล ดังนั้นผลการวิเคราะห์จึงอาจขาดสารสนเทศเกี่ยวกับสมาชิกในกลุ่มประชากรย่อยหรือกลุ่มแฝงไป (Cho & Cohen, 2010)

ข้อจำกัดของโมเดล multilevel LC คือโมเดลมีการวัดความสามารถของผู้สอบเป็นกลุ่มแฝง โดยในกลุ่มแฝงนี้ผู้สอบทุกคนจะถูกพิจารณาให้มีความสามารถเท่าเทียมกัน ผลการวิเคราะห์จึงขาดการให้สารสนเทศเกี่ยวกับความผันแปรของความสามารถระหว่างผู้สอบ (Cho, 2007)

ด้วยเหตุนี้โมเดล MMixIRT (Cho, 2007; Vermunt, 2007) จึงถูกพัฒนาขึ้นเพื่อผ่อนคลายข้อจำกัดข้างต้น โดยมีเป้าหมายคือการจัดกลุ่มแฝงให้กับนักเรียนในแต่ละโรงเรียนที่มีแบบแผนการตอบคล้ายคลึงกันและมีการให้สารสนเทศเชิงปริมาณเกี่ยวกับคุณลักษณะแฝงเช่นเดียวกับโมเดล MixIRT แต่มีจุดแตกต่างตรงที่โมเดล MMixIRT มีการพิจารณาให้กลุ่มแฝงระดับหน่วย (นักเรียน) ซ้อนอยู่ภายใต้กลุ่มแฝงระดับกลุ่ม (โรงเรียน) เพื่อแสดงนัยว่าโรงเรียนอาจมีผลต่อคุณลักษณะของนักเรียนตามแนวคิดของโมเดลพหุระดับ ด้วยความครอบคลุมของการให้สารสนเทศที่สอดคล้องกับธรรมชาติของข้อมูลทางการศึกษา โมเดล MMixIRT จึงถูกนำไปใช้อย่างแพร่หลายในการวิจัยทางการศึกษา เช่น การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเพื่อจำแนกกลุ่มโรงเรียนตามกลุ่มนักเรียนที่มีแบบแผนการตอบคล้ายกัน (Cho & Cohen, 2010; Finch & Finch, 2013) การจำแนกกลุ่มประเทศตามกลุ่มบุคคลที่มีภาวะอารมณ์คล้ายกัน (Tay et al., 2011) การจำแนกกลุ่มหลักสูตรตามกลุ่มนักศึกษาที่มีความพึงพอใจคล้ายกัน (Bacci & Ginaldi, 2015)

จากการบูรณาการองค์ความรู้ระหว่าง 3 โมเดล ได้แก่ โมเดล IRT โมเดล LC และโมเดลพหุระดับ องค์ประกอบของโมเดล MMixIRT จึงสามารถแสดงได้ดังภาพ 2.10 และอาจนำไปเปรียบเทียบกับ การวิเคราะห์ตัวแปรแฝงรูปแบบอื่น ๆ ได้ดังตาราง 2.4



ภาพ 2.10 แผนภาพองค์ประกอบของโมเดล MMixIRT
(ที่มา: Cho, 2007)

ตาราง 2.4 โมเดลการวิเคราะห์ตัวแปรแฝง

โครงสร้างข้อมูล	ตัวแปรแฝง	ตัวแปรสังเกตได้	
		ต่อเนื่อง (continuous)	จัดกลุ่ม (categorical)
ระดับเดียว (single-level)	ต่อเนื่อง (continuous)	การวิเคราะห์องค์ประกอบ (FA: factor analysis)	ทฤษฎีการตอบสนองข้อสอบ (IRT: item response theory)
	จัดกลุ่ม (categorical)	การวิเคราะห์โปรไฟล์แฝง (LP: latent profile analysis)	การวิเคราะห์กลุ่มแฝง (LC: latent class analysis)
	ต่อเนื่อง/จัดกลุ่ม (mixed)	การวิเคราะห์องค์ประกอบแบบผสม (MixFA: mixture FA)	ทฤษฎีการตอบสนองข้อสอบแบบผสม (MixIRT: mixture IRT)
พหุระดับ (multilevel)	ต่อเนื่อง (continuous)	การวิเคราะห์องค์ประกอบพหุระดับ (multilevel FA)	ทฤษฎีการตอบสนองข้อสอบพหุระดับ (multilevel IRT)
	จัดกลุ่ม (categorical)	การวิเคราะห์โปรไฟล์แฝงพหุระดับ (multilevel LP)	การวิเคราะห์กลุ่มแฝงพหุระดับ (multilevel LC)
	ต่อเนื่อง/จัดกลุ่ม (mixed)	การวิเคราะห์องค์ประกอบแบบผสมพหุระดับ (MMixFA: multilevel mixture FA)	ทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ (MMixIRT: multilevel mixture IRT)

(ที่มา: Lubke & Muthén, 2005; Tay et al., 2011)

การแบ่งรูปแบบของโมเดล MMixIRT สามารถพิจารณาได้ในทำนองเดียวกับโมเดล IRT ดังนี้

ก) โมเดล multilevel mixture Rasch (MMixRasch)

โมเดล MMixRasch (Cho, 2007; Cho & Cohen, 2010) ได้รับการบูรณาการมาจากโมเดล Rasch โมเดล LC และโมเดลพหุระดับ จึงมีข้อตกลงเบื้องต้นคล้ายกับโมเดล MixIRT แต่มีข้อตกลงเพิ่มเติมว่าข้อมูลในระดับโรงเรียนอาจมีผลต่อความน่าจะเป็นของการเป็นสมาชิกในกลุ่มแฝงระดับนักเรียนด้วย (Tay et al., 2011)

จากโมเดล multilevel LC ในสมการ (2.25) ทราบว่าความน่าจะเป็นของแบบแผนการตอบข้อสอบในระดับโรงเรียนคือ $P(\underline{y}_s) = \sum_{g=1}^G \pi_g \left[\prod_{j=1}^{n_s} \left\{ \sum_{k=1}^K \pi_{k|g} \left(\prod_{i=1}^I P(y_{ijs} | k, g) \right) \right\} \right]$ เมื่อ $P(y_{ijs} | k, g)$ คือความน่าจะเป็นแบบมีเงื่อนไขของผลการตอบข้อสอบรายข้อ ซึ่งสามารถอธิบายได้ด้วยโมเดล Rasch ดังนั้นความน่าจะเป็นของแบบแผนการตอบข้อสอบในระดับโรงเรียนสำหรับโมเดล MMixRasch จึงเป็นดังสมการ (2.26)

$$P(\underline{y}_s) = \sum_{g=1}^G \pi_g \left[\prod_{j=1}^{n_s} \left\{ \sum_{k=1}^K \pi_{k|g} \left(\prod_{i=1}^I \frac{1}{1 + \exp[-y_{ijs} (\theta_{jskg} - b_{ikg})]} \right) \right\} \right] \quad (2.26)$$

ดังนั้นฟังก์ชันการตอบสนองข้อสอบสำหรับโมเดล MMixRasch จึงมีรูปแบบดังสมการ (2.27)

$$P(y_{ijs} = 1 | \theta_{jskg}) = \sum_{g=1}^G \sum_{k=1}^K \pi_g \cdot \pi_{k|g} \left(\frac{1}{1 + \exp[-(\theta_{jskg} - b_{ikg})]} \right) \quad (2.27)$$

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; b_{ikg} คือพารามิเตอร์ความยากของข้อสอบ i ในกลุ่มแฝงระดับนักเรียน k และระดับโรงเรียน g โดยโครงสร้างสัดส่วนของสมาชิกในกลุ่มแฝงระดับโรงเรียน (π_g) และสัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียนในกลุ่มแฝงระดับโรงเรียน ($\pi_{k|g}$) สามารถแสดงได้ดังตาราง 2.5

ตาราง 2.5 โครงสร้างสัดส่วนของสมาชิกในกลุ่มแฝงสำหรับโมเดล MMixIRT

กลุ่มแฝงระดับ นักเรียนที่ k	กลุ่มแฝงระดับโรงเรียนที่ g				$\sum_{g=1}^G \pi_g = 1$
	$g = 1$	$g = 2$	\dots	$g = G$	
$k = 1$	$\pi_{1 1}$	$\pi_{1 2}$	\dots	$\pi_{1 G}$	
$k = 2$	$\pi_{2 1}$	$\pi_{2 2}$	\dots	$\pi_{2 G}$	
\vdots	\vdots	\vdots	\dots	\vdots	
$k = K$	$\pi_{K 1}$	$\pi_{K 2}$	\dots	$\pi_{K G}$	
	$\sum_{k=1}^K \pi_{k 1} = 1$	$\sum_{k=1}^K \pi_{k 2} = 1$	\dots	$\sum_{k=1}^K \pi_{k G} = 1$	

(ที่มา: Cho & Cohen, 2010)

ทั้งนี้อาจเรียกโมเดลในสมการ (2.27) ว่า MMixRasch รูปแบบทั่วไป เนื่องจากมีการรวมความแปรปรวนของความคลาดเคลื่อนจากการวัดในระดับนักเรียนและระดับโรงเรียนไว้ในโมเดลแล้ว จึงไม่สามารถแยกความแปรปรวนออกเป็นระดับได้เหมือนโมเดล multilevel IRT (Cho, 2007) อย่างไรก็ตามต่อมาได้มีการพัฒนาโมเดล MMixRasch รูปแบบเฉพาะซึ่งมีการแยกความแปรปรวนดังกล่าวในระดับนักเรียนและระดับโรงเรียนออกจากกัน โดยมีการกำหนดเงื่อนไขบังคับ (constraint) เกี่ยวกับกลุ่มแฝงเพิ่มเติม เพื่อประโยชน์สำหรับการศึกษาในบริบทเฉพาะ (Cho & Cohen, 2010) ซึ่งแบ่งได้เป็น 3 รูปแบบย่อย รายละเอียดดังนี้

ก.1) โมเดล MMixRasch รูปแบบเฉพาะ 1

โมเดล MMixRasch รูปแบบเฉพาะ 1 มีการกำหนดเงื่อนไขบังคับให้สัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียนมีค่าเท่ากันสำหรับทุกกลุ่มแฝงระดับโรงเรียน ($\pi_{k|1} = \pi_{k|2} = \dots = \pi_{k|G}$) ผลดังกล่าวจึงทำให้พารามิเตอร์ความสามารถและความยากของข้อสอบถูกแยกเป็นพารามิเตอร์ระดับนักเรียนและระดับโรงเรียนคือ $\theta_{jskg} = \theta_{jsk} + \theta_{jsg}$ และ $b_{ikg} = b_{ik} + b_{ig}$ ดังนั้นฟังก์ชันการตอบสนองข้อสอบสำหรับโมเดล MMixRasch รูปแบบเฉพาะ 1 จึงมีรูปแบบดังสมการ (2.28)

$$P(y_{ijs} = 1 | \theta_{jsk}, \theta_{jsg}) = \sum_{g=1}^G \sum_{k=1}^K \pi_g \cdot \pi_k \left(\frac{1}{1 + \exp[-((\theta_{jsk} + \theta_{jsg}) - (b_{ik} + b_{ig}))]} \right) \quad (2.28)$$

ก.2) โมเดล MMixRasch รูปแบบเฉพาะ 2

โมเดล MMixRasch รูปแบบเฉพาะ 2 มีการกำหนดเงื่อนไขบังคับให้พารามิเตอร์ความสามารถและความยากของข้อสอบไม่มีความผันแปรระหว่างกลุ่มแฝงระดับโรงเรียน ผลดังกล่าวจึงทำให้ $\theta_{jsk} = \theta_{jsk} + \theta_{js}$ และ $b_{ikg} = b_{ik}$ (Asparouhov & Muthén, 2008) ดังนั้นฟังก์ชันการตอบสนองข้อสอบสำหรับโมเดล MMixRasch รูปแบบเฉพาะ 2 จึงมีรูปแบบดังสมการ (2.29)

$$P(y_{ijs} = 1 | \theta_{jsk}, \theta_{js}) = \sum_{k=1}^K \pi_k \left(\frac{1}{1 + \exp[-((\theta_{jsk} + \theta_{js}) - b_{ik})]} \right) \quad (2.29)$$

ก.3) โมเดล MMixRasch รูปแบบเฉพาะ 3

โมเดล MMixRasch รูปแบบเฉพาะ 3 มีการกำหนดเงื่อนไขบังคับให้พารามิเตอร์ความสามารถและความยากของข้อสอบไม่มีความผันแปรระหว่างกลุ่มแฝงระดับนักเรียน ผลดังกล่าวจึงทำให้ $\theta_{jsk} = \theta_{js} + \theta_{jsg}$ และ $b_{ikg} = b_{ig}$ (Vermunt, 2007) ดังนั้นฟังก์ชันการตอบสนองข้อสอบสำหรับโมเดล MMixRasch รูปแบบเฉพาะ 3 จึงมีรูปแบบดังสมการ (2.30)

$$P(y_{ijs} = 1 | \theta_{js}, \theta_{jsg}) = \sum_{g=1}^G \pi_g \left(\frac{1}{1 + \exp[-((\theta_{js} + \theta_{jsg}) - b_{ig})]} \right) \quad (2.30)$$

ทั้งนี้อาจเปรียบเทียบโมเดล MMixRasch รูปแบบต่าง ๆ ได้ดังตาราง 2.6

ตาราง 2.6 การเปรียบเทียบโมเดล MMixRasch รูปแบบทั่วไปและรูปแบบเฉพาะ

โมเดล	เงื่อนไขบังคับ	ระดับ	ความสามารถ	กลุ่มแฝง	ความยาก
รูปแบบทั่วไป	สัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียนในกลุ่มแฝงระดับโรงเรียนมีค่าแตกต่างกัน	นักเรียนและโรงเรียน	$\theta_{jsk} \sim N(\mu_k, \sigma_k^2)$	$\pi_{k g}, \pi_g$	b_{ikg}
รูปแบบเฉพาะ 1	สัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียนในกลุ่มแฝงระดับโรงเรียนมีค่าเท่ากัน	นักเรียน	$\theta_{jsk} \sim N(\mu_k, \sigma_k^2)$	π_k	b_{ik}
		โรงเรียน	$\theta_{jsg} \sim N(\mu_g, \sigma_g^2)$	π_g	b_{ig}
รูปแบบเฉพาะ 2	พารามิเตอร์ไม่มีความผันแปรระหว่างกลุ่มแฝงระดับโรงเรียน	นักเรียน	$\theta_{jsk} \sim N(\mu_k, \sigma_k^2)$	π_k	b_{ik}
		โรงเรียน	$\theta_{js} \sim N(0, 1)$	-	-
รูปแบบเฉพาะ 3	พารามิเตอร์ไม่มีความผันแปรระหว่างกลุ่มแฝงระดับนักเรียน	นักเรียน	$\theta_{js} \sim N(0, 1)$	-	-
		โรงเรียน	$\theta_{jsg} \sim N(\mu_g, \sigma_g^2)$	π_g	b_{ig}

(ที่มา: Cho & Cohen, 2010)

อย่างไรก็ตามสำหรับโมเดล MMix2PL และ MMix3PL ในงานวิจัยนี้จะขอเสนอเฉพาะโมเดลรูปแบบทั่วไป เนื่องจากเป็นรูปแบบที่สอดคล้องกับธรรมชาติของข้อมูลทางการศึกษา ซึ่งโดยทั่วไปพบว่าค่าสัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียนและระดับโรงเรียนมีค่าแตกต่างกัน

ข) โมเดล multilevel mixture 2PL (MMix2PL)

โมเดล MMix2PL มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (2.31)

$$P(y_{ijs} = 1 | \theta_{jskg}) = \sum_{g=1}^G \sum_{k=1}^K \pi_g \cdot \pi_{k|g} \left(\frac{1}{1 + \exp[-a_{ikg} (\theta_{jskg} - b_{ikg})]} \right) \quad (2.31)$$

เมื่อ a_{ikg} คือพารามิเตอร์อำนาจจำแนกของข้อสอบ i ในกลุ่มแฝงระดับนักเรียน k และระดับโรงเรียน g

ค) โมเดล multilevel mixture 3PL (MMix3PL)

โมเดล MMix3PL มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (2.32)

$$P(y_{ijs} = 1 | \theta_{jskg}) = \sum_{g=1}^G \sum_{k=1}^K \pi_g \cdot \pi_{k|g} \left(c_{ikg} + (1 - c_{ikg}) \frac{1}{1 + \exp[-a_{ikg} (\theta_{jskg} - b_{ikg})]} \right) \quad (2.32)$$

เมื่อ c_{ikg} คือพารามิเตอร์การเดาของข้อสอบ i ในกลุ่มแฝงระดับนักเรียน k และระดับโรงเรียน g

การประมาณค่าพารามิเตอร์สำหรับโมเดล MMixIRT โดยทั่วไปนิยมใช้วิธีการประมาณเพื่อหาสัดส่วนของสมาชิกในกลุ่มแฝงและคุณลักษณะแฝงไปพร้อมกันเช่นเดียวกับโมเดล MixIRT ซึ่งทำได้ 2 วิธีคือ 1) วิธีภาวะควรจะเป็นสูงสุด และ 2) วิธีการประมาณค่าแบบเบส์ โดยตัวอย่างการกำหนดการแจกแจงก่อนหน้าของพารามิเตอร์สำหรับโมเดล MMixRasch (Cho & Cohen, 2010) เช่น

$$\text{พารามิเตอร์ผู้สอบ: } \theta_{jskg} \sim N(\mu_{kg}, \sigma_{kg}^2); \theta_{jskg} = \sigma_{kg} \cdot \eta_{jskg}; \sigma_{kg} \sim N(0, 1) T(0,)$$

$$\eta_{jskg} \sim N(\mu_{kg}, 1); \mu_{kg} \sim N(0, 1) \text{ เมื่อ } \mu_{11} = 0$$

$$\text{พารามิเตอร์ข้อสอบ: } b_{ikg} \sim N(0, 1)$$

$$\text{พารามิเตอร์กลุ่มแฝง: } g \sim \text{Mult}(1, \pi_g); k \sim \text{Mult}(1, \pi_{k|g})$$

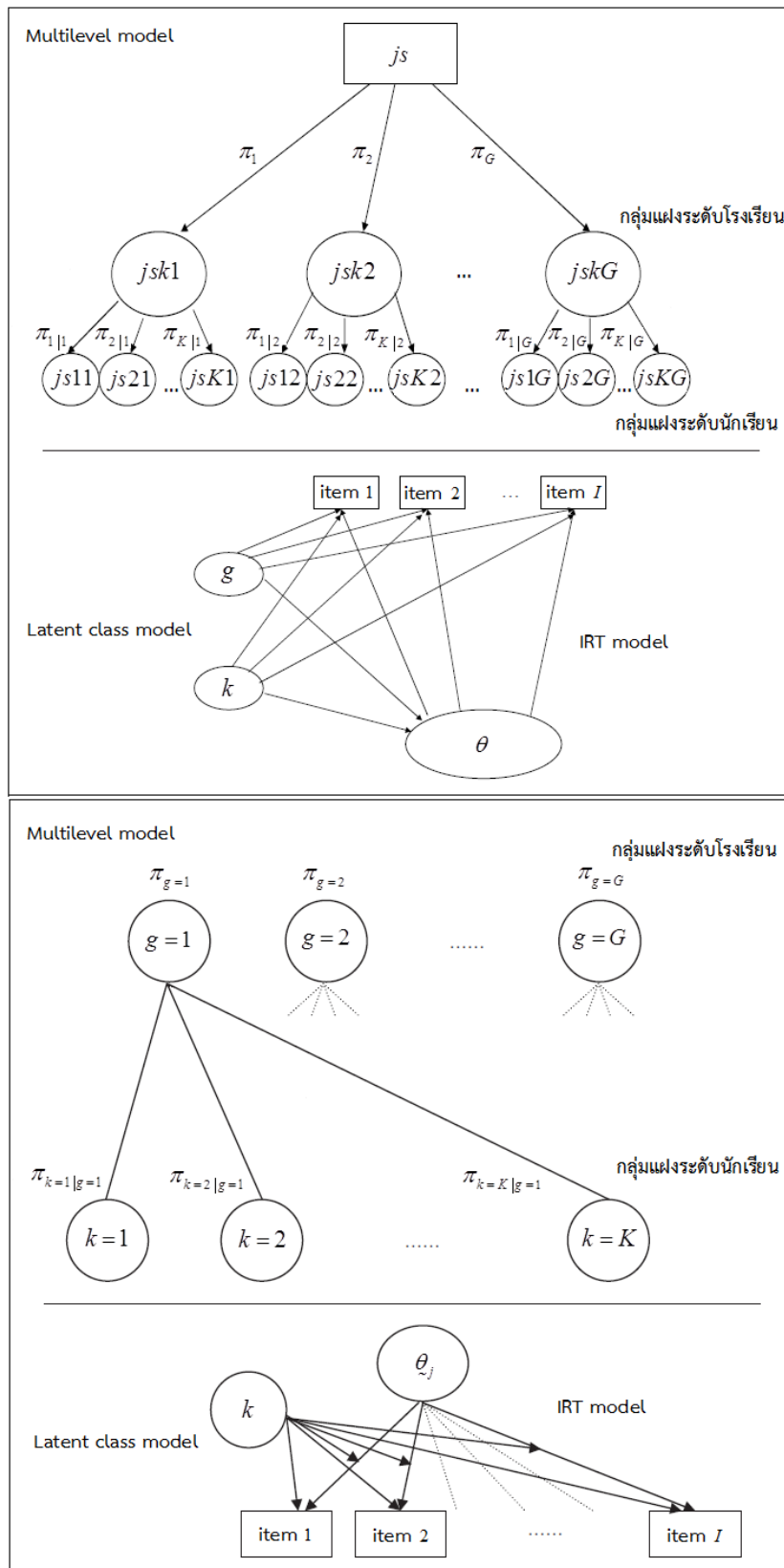
$$(\pi_1, \dots, \pi_G) \sim \text{Dir}(1, \dots, 1); (\pi_{1|1}, \dots, \pi_{K|G}) \sim \text{Dir}(1, \dots, 1)$$

ส่วน ICC สำหรับโมเดล MMixIRT รูปแบบทั่วไปจะไม่มีสูตรคำนวณโดยตรง เนื่องจากความแปรปรวนของความคลาดเคลื่อนในระดับนักเรียนและระดับโรงเรียนได้ถูกรวมไว้ในโมเดลแล้ว อย่างไรก็ตามถ้าต้องการหา ICC สำหรับโมเดล MMixIRT รูปแบบทั่วไป Cho และ Cohen (2010) ได้เสนอว่าสามารถนำ ICC จากโมเดล multilevel IRT มาพิจารณาแทนได้ เนื่องจากโมเดลทั้งสองมีการคำนึงถึงโครงสร้างพหุระดับของข้อมูลและมีการวัดคุณลักษณะแฝงเป็นค่าต่อเนื่องเหมือนกัน

การคัดเลือกโมเดลสำหรับโมเดล MMixIRT โดยทั่วไปนิยมพิจารณาจากเกณฑ์สารสนเทศ BIC เนื่องจากผลการศึกษาพบว่า BIC มีแนวโน้มสามารถคัดเลือกโมเดลที่มีจำนวนกลุ่มแฝงได้สอดคล้องกับข้อมูลมากที่สุด (Sen & Cohen, 2020; Sen et al., 2019)

โปรแกรมคอมพิวเตอร์ที่รองรับการวิเคราะห์โมเดล MMixIRT เช่น โปรแกรม Latent Gold โปรแกรม Mplus และโปรแกรม MultiBUGS

จากที่ได้กล่าวไปในตอนต้นแล้วว่า โมเดล MMixIRT เกิดจากการบูรณาการองค์ความรู้ระหว่าง 3 โมเดล ได้แก่ 1) โมเดล IRT 2) โมเดลกลุ่มแฝง และ 3) โมเดลพหุระดับ ในส่วนนี้ผู้วิจัยจึงขอยกตัวอย่างโครงสร้างของโมเดล MMixRasch และ MMix2PL ที่มีการนำเสนอไว้ในงานวิจัยก่อนหน้า มาอธิบายเป็นกรณีศึกษาดังภาพ 2.11



ภาพ 2.11 โครงสร้างของโมเดล MMixRasch (บน) และ MMix2PL (ล่าง)
(ที่มา: Cho, 2007; Tay et al., 2011)

โครงสร้างของโมเดล MMixRasch (ดูภาพ 2.11 (บน)) สามารถอธิบายได้ดังนี้

- โมเดลพหุระดับ: ค่าสังเกตของนักเรียน j ในโรงเรียน s จะถูกจัดให้อยู่ในกลุ่มแฝงระดับโรงเรียนตามค่าสัดส่วนของสมาชิกในกลุ่มแฝงระดับโรงเรียน (π_s) และถูกจัดให้อยู่ในกลุ่มแฝงระดับนักเรียนตามค่าสัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียนภายใต้กลุ่มแฝงระดับโรงเรียน ($\pi_{k|s}$)

- โมเดล LC: การจำแนกกลุ่มแฝงระดับนักเรียนและระดับโรงเรียนจะถูกพิจารณาจากแบบแผนการตอบข้อสอบและความสามารถแฝงของผู้สอบ (θ)

- โมเดล IRT: ความสามารถแฝงสามารถวัดได้จากผลการตอบข้อสอบรายข้อ

ส่วนโครงสร้างของโมเดล MMix2PL (ดูภาพ 2.11 (ล่าง)) สามารถอธิบายได้ดังนี้

- โมเดลพหุระดับ: โรงเรียนแต่ละแห่งจะถูกจัดให้อยู่ในกลุ่มแฝงระดับโรงเรียนตามค่าสัดส่วนของสมาชิกในกลุ่มแฝงระดับโรงเรียน (π_s) และนักเรียนในแต่ละโรงเรียนก็จะถูกจัดให้อยู่ในกลุ่มแฝงระดับนักเรียนตามค่าสัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียนภายใต้กลุ่มแฝงระดับโรงเรียน ($\pi_{k|s}$)

- โมเดล LC: การจำแนกกลุ่มแฝงระดับนักเรียนจะพิจารณาจากแบบแผนการตอบข้อสอบและค่าน้ำหนัก (loading) ระหว่างความสามารถแฝงกับผลการตอบข้อสอบรายข้อ เพื่อแสดงนัยว่าพารามิเตอร์ความยาก (b) และพารามิเตอร์อำนาจจำแนก (a) มีค่าแปรเปลี่ยนได้ตามกลุ่มแฝง

- โมเดล IRT: ความสามารถแฝงสามารถวัดได้จากผลการตอบข้อสอบรายข้อ

ทั้งนี้การสังเคราะห์งานวิจัยเกี่ยวกับโมเดล MMixIRT พบว่าโมเดลที่นิยมนำมาศึกษามีเพียงโมเดล MMixRasch และ MMix2PL ส่วนการประมาณค่าพบว่าวิธีภาวะควรจะเป็นสูงสุด (ML) มีการนำมาใช้มากกว่าวิธีแบบเบส์ โดยถ้าเป็นวิธี ML โปรแกรมที่ใช้จะเป็น Latent Gold หรือ Mplus แต่หากเป็นวิธีแบบเบส์ โปรแกรมที่ใช้จะเป็นโปรแกรม WinBUGS รายละเอียดดังตาราง 2.7

ตาราง 2.7 การสังเคราะห์งานวิจัยที่ศึกษาเกี่ยวกับโมเดล MMixIRT

แหล่งอ้างอิง	โมเดล	ประเด็นศึกษา	วิธีประมาณค่า	เกณฑ์สารสนเทศ	โปรแกรม
Vermunt (2008)	MMixRasch MMix2PL	- เสนอวิธีประมาณค่า ML สำหรับโมเดล MMixIRT - นำโมเดลไปทดลองใช้กับข้อมูลทางการศึกษา	ML (EM)	BIC	- Latent Gold
Cho และ Cohen (2010)	MMixRasch	- ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและจำแนกกลุ่มโรงเรียนตามกลุ่มนักเรียนที่มีคุณลักษณะคล้ายกัน	Bayes (MCMC)	AIC, BIC	- WinBUGS
Tay และคณะ (2011)	MMM-IRT (MMix2PL พหุมิติ)	- นำเสนอโมเดล MMM-IRT - จำแนกกลุ่มประเทศตามกลุ่มคนที่มีภาวะอารมณ์คล้ายกัน	ML (EM)	BIC	- Latent Gold
Finch และ Finch (2013)	MMixRasch	- ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ	ML (EM)	AIC, ABIC	- Mplus
Bacci และ Ginaldi (2015)	MMix2PL	- จำแนกกลุ่มหลักสูตรระดับมหาวิทยาลัยตามกลุ่มนักศึกษาที่มีความพึงพอใจคล้ายกัน	ML (EM)	BIC	- Latent Gold - R (MultiLCIRT)
Lee และคณะ (2018)	MMix2PL	- เปรียบเทียบความแม่นยำของการประมาณค่าพารามิเตอร์และความถูกต้องของการตรวจสอบจำนวนกลุ่มแฝง เมื่อมีการละเลยโครงสร้างพหุระดับ	ML (EM)	AIC, BIC	- Mplus
Sen และคณะ (2019)	MMixRasch	- เปรียบเทียบ ความถูกต้อง ของเกณฑ์สารสนเทศ ในการตรวจสอบจำนวนกลุ่มแฝง	Bayes (MCMC)	AIC, BIC, CAIC, SABIC	- WinBUGS
Sen และ Cohen (2020)	MMix2PL	- เปรียบเทียบความแม่นยำของการประมาณค่าพารามิเตอร์และความถูกต้องของการตรวจสอบจำนวนกลุ่มแฝง เมื่อคุณลักษณะของแบบสอบและขนาดตัวอย่างแตกต่างกัน	ML (EM)	AIC, BIC, CAIC, SABIC	- Mplus

โดยสรุปเมื่อเปรียบเทียบโมเดลทั้ง 7 รูปแบบข้างต้น อาจแบ่งโมเดลตามการวัดความสามารถแฝงได้เป็น 4 ประเภท ได้แก่ 1) โมเดลที่ไม่มีการวัดความสามารถแฝง คือ โมเดลพหุระดับ 2) โมเดลที่วัดความสามารถแฝงเป็นค่าไม่ต่อเนื่อง ได้แก่ โมเดล LC และ multilevel LC 3) โมเดลที่วัดความสามารถแฝงเป็นค่าต่อเนื่อง ได้แก่ โมเดล IRT และ multilevel IRT และ 4) โมเดลที่วัดความสามารถแฝงเป็นทั้งค่าต่อเนื่องและไม่ต่อเนื่อง (แบบผสม) ได้แก่ โมเดล MixIRT และ MMixIRT โดยอาจสังเกตได้ว่าการวัดพารามิเตอร์ของข้อสอบจะมีเฉพาะในโมเดลที่มีการบูรณาการมาจากโมเดล IRT ในขณะที่การวัดพารามิเตอร์กลุ่มแฝงจะมีเฉพาะในโมเดลที่มีการบูรณาการมาจากโมเดล LC เช่นกัน ส่วนโปรแกรมคอมพิวเตอร์ที่รองรับการวิเคราะห์สำหรับทุกโมเดล เช่น โปรแกรม Mplus, WinBUGS, OpenBUGS, MultiBUGS รายละเอียดดังตาราง 2.8

ตาราง 2.8 การเปรียบเทียบพารามิเตอร์ของโมเดลและโปรแกรมวิเคราะห์

	พารามิเตอร์			โปรแกรมที่รองรับการวิเคราะห์
	ความสามารถ	ข้อสอบ	กลุ่มแฝง	
1. IRT	$\theta_j \sim N(0,1)$ * Continuous	a_i, b_i, c_i	-	- IRTPRO - Mplus - R package e.g. <i>ltm, mirt, irtoys</i> - WinBUGS/OpenBUGS/MultiBUGS
2. LC	$\theta_j = \sum_{k=1}^K \pi_k \cdot \theta_k$ * Categorical	-	π_k	- Latent Gold - Mplus - R package e.g. <i>randomLCA, poLCA</i> - WinBUGS/OpenBUGS/MultiBUGS
3. Multilevel	-	-	-	- HLM - Mplus - R package e.g. <i>lme4, nlme, rstan</i> - WinBUGS/OpenBUGS/MultiBUGS
4. MixIRT	$\theta_{jk} \sim N(\mu_k, \sigma_k^2)$ * Mixed	a_{ik}, b_{ik}, c_{ik}	π_k	- Latent Gold - Mplus - R package e.g. <i>psychomix, mixRasch</i> - WinBUGS/OpenBUGS/MultiBUGS - WINMIRA
5. Multilevel IRT	$\theta_{js} = u_{0s} + r_{js};$ $u_{0s} \sim N(\gamma_{00}, \zeta), r_{js} \sim N(0, \tau)$ * Continuous	a_i, b_i, c_i	-	- Mplus - R package e.g. <i>lme4, nlme, rstan</i> - WinBUGS/OpenBUGS/MultiBUGS
6. Multilevel LC	$\theta_{js} = \sum_{g=1}^G \sum_{k=1}^K \pi_g \pi_{kjg} \cdot \theta_{kjg}$ * Categorical	-	π_g, π_{kjg}	- Latent Gold - Mplus - WinBUGS/OpenBUGS/MultiBUGS
7. MMixIRT	$\theta_{jkg} \sim N(\mu_{kg}, \sigma_{kg}^2)$ * Mixed	$a_{ikg}, b_{ikg}, c_{ikg}$	π_g, π_{kjg}	- Latent Gold - Mplus - WinBUGS/OpenBUGS/MultiBUGS

ตอนที่ 3 วิธีการประมาณค่าพารามิเตอร์

จากที่นำเสนอไปในส่วนที่แล้วว่า การประมาณค่าพารามิเตอร์ในโมเดลโดยทั่วไปสามารถทำได้ 2 วิธีคือ 1) วิธีภาวะควรจะเป็นสูงสุด และ 2) วิธีการประมาณค่าแบบเบย์ ในส่วนนี้จึงขอกกล่าวถึงแนวคิดในการประมาณค่าของทั้งสองวิธีโดยสังเขป รายละเอียดดังนี้

3.1 วิธีภาวะควรจะเป็นสูงสุด (maximum likelihood)

วิธีภาวะควรจะเป็นสูงสุดเป็นเทคนิคการคัดเลือกเซตของค่าประมาณพารามิเตอร์ที่จะทำให้ฟังก์ชันภาวะควรจะเป็นมีค่าสูงสุด ซึ่งคำนวณได้จากการแก้สมการเชิงอนุพันธ์อันดับหนึ่งของฟังก์ชันล็อกภาวะควรจะเป็น (log-likelihood function) ที่มีการกำหนดค่าเป็น 0 อย่างไรก็ตามในทางปฏิบัติการพิสูจน์สูตรสมการเชิงอนุพันธ์ให้อยู่ในรูปปิดเป็นเรื่องที่ยากมาก เนื่องจากโมเดลมักมีโครงสร้างที่ซับซ้อนและไม่ได้อยู่ในรูปแบบเชิงเส้น การประมาณค่าด้วยวิธีนี้จึงมักจะต้องอาศัยขั้นตอนวิธีทวนซ้ำ เช่น ขั้นตอนวิธี EM ซึ่งเป็นเทคนิคที่นำความรู้เกี่ยวกับพารามิเตอร์ส่วนหนึ่งมาประมาณค่าพารามิเตอร์อีกส่วนหนึ่ง และทำซ้ำจนกว่าค่าประมาณของพารามิเตอร์จะลู่เข้าสู่ค่าที่เหมาะสม (สิวะโชติ ศรีสุทธิยากร, 2555) ทั้งนี้ผู้สนใจสามารถศึกษารายละเอียดของการประมาณค่าพารามิเตอร์สำหรับโมเดล MMixIRT ด้วยวิธีภาวะควรจะเป็นสูงสุดได้จาก Vermunt (2008)

3.2 วิธีการประมาณค่าแบบเบย์ (Bayesian estimation)

วิธีการประมาณค่าแบบเบย์เป็นเทคนิคการประมาณค่าที่มีแนวคิดที่ว่า “พารามิเตอร์คือตัวแปรสุ่มที่ทราบลักษณะการแจกแจงอยู่ก่อนแล้ว” แตกต่างจากการประมาณค่าทางสถิติทั่วไปซึ่งมีแนวคิดที่ว่า “พารามิเตอร์คือค่าคงที่ซึ่งไม่ทราบค่า” ทั้งนี้การแจกแจงที่ทราบอยู่ก่อนแล้วจะถูกเรียกว่า การแจกแจงก่อนหน้า (prior) เนื่องจากเป็นการแจกแจงที่กำหนดขึ้นก่อนทำการเก็บรวบรวมข้อมูล ซึ่งอาจทราบมาจากงานวิจัยที่เกี่ยวข้อง จากนั้นเมื่อเก็บรวบรวมข้อมูลค่าสังเกตมาได้ก็จะนำมาสร้างฟังก์ชันภาวะควรจะเป็น และนำฟังก์ชันดังกล่าวไปปรับปรุงการแจกแจงก่อนหน้า โดยการแจกแจงหลังการปรับปรุงจะถูกเรียกว่า การแจกแจงภายหลัง (posterior) ดังแสดงด้วยความสัมพันธ์ (2.33)

$$P(\theta | y) \propto L(\theta) P(\theta) \quad (2.33)$$

เมื่อ θ คือเวกเตอร์ของพารามิเตอร์ที่สนใจศึกษา; y คือเวกเตอร์ของข้อมูลค่าสังเกต; $P(\theta)$ คือความน่าจะเป็นของการแจกแจงก่อนหน้า; $L(\theta)$ คือฟังก์ชันภาวะควรจะเป็นที่ได้จากข้อมูลค่าสังเกต; $P(\theta | y)$ คือความน่าจะเป็นของการแจกแจงภายหลัง (Lim & Drasgow, 1990)

จากความสัมพันธ์ข้างต้นอาจกล่าวได้ว่า วิธีการประมาณค่าแบบเบย์เป็นการนำสารสนเทศจากข้อมูลเชิงประจักษ์เข้ามาช่วยปรับปรุงสารสนเทศที่มาจากความรู้ก่อนหน้า ทำให้ได้สารสนเทศใหม่ที่สมบูรณ์มากขึ้น ดังนั้นหากมีการกำหนดการแจกแจงก่อนหน้าได้อย่างเหมาะสม ผลการประมาณค่าแบบเบย์ย่อมสมบูรณ์มากกว่าการประมาณค่าด้วยวิธีทั่วไป

อย่างไรก็ตามในโมเดลที่มีความซับซ้อนมักพบว่าจำนวนพารามิเตอร์ที่ต้องประมาณค่ามีอยู่เป็นจำนวนมาก ความน่าจะเป็นของการแจกแจงภายหลังจึงมีจำนวนมิติมากตามไปด้วย ประกอบกับยังเป็นที่ยากในการพิสูจน์สูตรให้อยู่ในรูปปิด จากปัญหาดังกล่าวจึงมีการนำวิธีเชิงจำลองเข้ามาช่วยในการประมาณค่าความน่าจะเป็นของการแจกแจงภายหลัง โดยวิธีที่นิยมใช้มากที่สุดก็คือขั้นตอนวิธี Markov chain Monte Carlo (MCMC)

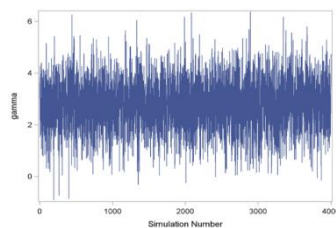
ขั้นตอนวิธี MCMC เป็นเทคนิคการจำลองตัวอย่างเพื่อประมาณค่าความน่าจะเป็นของการแจกแจงภายหลังของพารามิเตอร์ที่สนใจ โดยอาศัยการสุ่มตัวอย่างจากลูกโซ่มาร์คอฟ ซึ่งมีคุณสมบัติลดทอนไม่ได้ (irreducible) และไม่เป็นคาบ (aperiodic) ดังนั้นหากจำนวนรอบที่ทำการสุ่มมีมากพอ การแจกแจงของลูกโซ่มาร์คอฟจะลู่เข้าสู่การแจกแจงสถานะคงตัว (stationary distribution) ซึ่งมีค่าเทียบเท่ากับการแจกแจงภายหลังของพารามิเตอร์ที่สนใจ ทั้งนี้ประเด็นที่ควรพิจารณาสำหรับการประมาณค่าแบบเบย์ในขั้นตอนวิธี MCMC (สิวะโชติ ศรีสุทธิยากร, 2555) มีดังนี้

ข.1) *ค่าเริ่มต้น (initial value)* การกำหนดค่าเริ่มต้นที่ดีจะทำให้ลูกโซ่มาร์คอฟสามารถลู่เข้าสู่การแจกแจงสถานะคงตัวได้รวดเร็ว โดยหลักในการเลือกค่าเริ่มต้นที่ดีควรเลือกค่าให้ใกล้เคียงกับจุดศูนย์กลางของการแจกแจงภายหลังที่ต้องการให้มากที่สุด

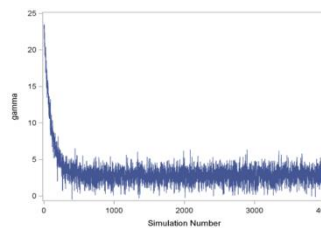
ข.2) *คาบระยะทดลอง (burn-in period)* ลูกโซ่มาร์คอฟที่สร้างขึ้นจำเป็นต้องใช้ระยะเวลาช่วงหนึ่งเพื่อที่จะเข้าสู่สถานะคงตัว อย่างไรก็ตามค่าสังเกตในช่วงแรกของการสร้างลูกโซ่จะยังไม่ได้ถูกสุ่มมาจากการแจกแจงภายหลังที่ต้องการ ดังนั้นจึงจำเป็นต้องตัดค่าสังเกต (burn-in) ในช่วงแรกทิ้ง เพื่อไม่ให้ค่าสังเกตในช่วงที่ลูกโซ่ยังไม่เข้าสู่สถานะคงตัวเข้ามามีอิทธิพลต่อการประมาณค่าความน่าจะเป็นของการแจกแจงภายหลัง

ข.3) *การลู่เข้าของลูกโซ่ (chain convergence)* ในทางปฏิบัติผู้วิเคราะห์ไม่สามารถตรวจสอบได้ว่าลูกโซ่มีการลู่เข้าสู่การแจกแจงสถานะคงตัวตามที่ต้องการแล้ว ทำได้แต่เพียงตรวจสอบว่าลูกโซ่ยังไม่ลู่เข้าสู่การแจกแจงสถานะคงตัวเท่านั้น แนวทางที่ใช้สำหรับตรวจสอบแบ่งได้เป็น 2 แนวทางคือ การวินิจฉัยเชิงกราฟิก และการวินิจฉัยเชิงสถิติ รายละเอียดดังนี้

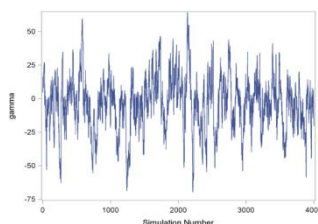
ข.3.1) *การวินิจฉัยเชิงกราฟิก* เช่น trace plot ซึ่งแสดงรอยการเปลี่ยนแปลงของค่าสังเกตที่สุ่มได้ในแต่ละรอบของลูกโซ่ หากกราฟมีลักษณะคงที่หรือไม่มีการเปลี่ยนแปลงนักก็อาจอนุมานได้ว่าลูกโซ่เข้าสู่สถานะคงตัวแล้ว ตัวอย่าง trace plot ในกรณีต่าง ๆ แสดงดังภาพ 2.12-2.15 อย่างไรก็ตามวิธีนี้อาจยังไม่สามารถสรุปถึงการลู่เข้าของลูกโซ่ได้อย่างแน่ชัด เนื่องจากการลู่เข้าของลูกโซ่เป็นสิ่งที่ไม่สามารถทราบได้ จึงควรวินิจฉัยเชิงสถิติประกอบเพื่อเพิ่มความน่าเชื่อถือด้วย



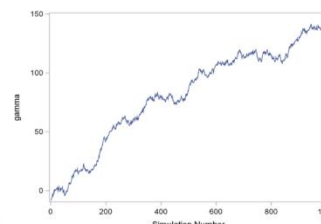
ภาพ 2.12 ลูกโซ่มีแนวโน้มเข้าสู่สถานะคงตัว



ภาพ 2.13 ลูกโซ่ควรถูกตัดทิ้งก่อนประมาณค่า



ภาพ 2.14 ลูกโซ่มีจำนวนน้อยเกินไป



ภาพ 2.15 ลูกโซ่มีปัญหาการเข้าสู่สถานะคงตัว

(ที่มา: สิวะโชติ ศรีสุทธิยากร, 2555)

ข.3.2) *แนวทางการวินิจฉัยเชิงสถิติ* แบ่งเป็น 2 กลุ่มตามจำนวนลูกโซ่ที่ใช้ประมาณค่า ได้แก่ 1) การวินิจฉัยเมื่อประมาณค่าด้วยลูกโซ่เดียว เช่น 1.1) วิธีของ Geweke (1992) เป็นการเปรียบเทียบความแตกต่างของค่าเฉลี่ยระหว่างลูกโซ่ช่วงแรกและช่วงหลัง หากไม่พบความแตกต่างในการทดสอบก็อาจอนุมานได้ว่าลูกโซ่ได้เข้าสู่สถานะคงตัวแล้ว 1.2) วิธีของ Heidelberger และ Welch (1983) ประกอบด้วยการทดสอบ 2 ส่วน ส่วนแรกเรียกว่า stationary test เป็นการทดสอบเพื่อหาว่าลูกโซ่ที่สุ่มมาเข้าสู่สถานะคงตัวแล้วหรือไม่ การปฏิเสธสมมติฐานหลักจะหมายความว่าลูกโซ่ดังกล่าวต้องการจำนวนรอบในการสร้างเพิ่มขึ้นอีกถึงจะทำให้เข้าสู่สถานะคงตัว ส่วนที่สองเรียกว่า half-width test เป็นการตรวจสอบว่าจำนวนรอบในการสร้างลูกโซ่เพียงพอต่อการประมาณค่าเฉลี่ยของลูกโซ่ได้อย่างแม่นยำหรือไม่ การปฏิเสธสมมติฐานหลักจะหมายความว่าลูกโซ่ดังกล่าวต้องการจำนวนรอบในการสร้างเพิ่มขึ้นอีกถึงจะทำให้มีความแม่นยำตามที่กำหนด และ 1.3) วิธีของ Raftery และ Lewis (1992) เป็นการตรวจสอบการลู่เข้าโดยพิจารณาจากระดับความแม่นยำตามค่าควอนไทล์ที่กำหนด ซึ่งโดยทั่วไปกำหนดไว้ที่เปอร์เซ็นต์ไทม์ที่ 2.5 ที่ระดับความแม่นยำ .005 ด้วยความน่าจะเป็น .95 ผลการตรวจสอบจะให้สารสนเทศเกี่ยวกับจำนวนรอบที่น้อยที่สุดในการสร้างลูกโซ่ เพื่อให้ค่าสังเกตที่สุ่มในลูกโซ่ไม่มีความสัมพันธ์กันเองหรือมีอัตตสหสัมพันธ์ศูนย์ (zero autocorrelation) และ 2) การวินิจฉัยเมื่อประมาณค่าด้วยหลายลูกโซ่ ได้แก่ วิธีของ Gelman และ Rubin (1992) เป็นการพิจารณาจากค่าสถิติ shrink factor ซึ่งถูกคำนวณมาจากความแปรปรวนภายในลูกโซ่และความแปรปรวนระหว่างลูกโซ่ตามหลักการวิเคราะห์ความแปรปรวน (ANOVA) โดยหากค่าสถิติเข้าใกล้ 1 จะอนุมานได้ว่าลูกโซ่ได้เข้าสู่สถานะคงตัวแล้ว ทั้งนี้การวินิจฉัยเชิงสถิติที่กล่าวมาสามารถตรวจสอบได้จากโปรแกรม MultiBUGS และโปรแกรม R แพคเกจ coda (Plummer et al., 2019)

ข.4) *อัตสหสัมพันธ์ (autocorrelation)* เนื่องจากค่าสังเกตที่สุ่มจากลูกโซ่อาจมีความสัมพันธ์กันเองตามจำนวนรอบจำลอง ดังนั้นในการสร้างลูกโซ่ควรตรวจสอบอัตสหสัมพันธ์ระหว่างค่าสังเกตในช่วงเวลา (lag) ที่แตกต่างกัน เพื่อเป็นสารสนเทศช่วยตัดสินใจว่าควรเลือกเก็บค่าสังเกตเป็นจำนวนเท่าใด จึงจะเพียงพอทำให้ค่าเหล่านั้นมีความใกล้เคียงที่จะเป็นอิสระจากกัน

ข.5) *ค่าสถิติสรุปผล (summary statistics)* ในขั้นตอนวิธี MCMC สามารถนำตัวอย่างค่าสังเกตที่สุ่มจากลูกโซ่มาประมาณค่าพารามิเตอร์ที่สนใจได้โดยตรง กล่าวคือถ้ากำหนด $\Theta = \{\theta_1, \theta_2, \dots, \theta_p\}$ เป็นแถวลำดับ (array) สำหรับเวกเตอร์ของพารามิเตอร์ที่สนใจ p ตัว โดยเวกเตอร์ของพารามิเตอร์แต่ละตัวมีค่าสังเกตจากการสุ่มตัวอย่างด้วยลูกโซ่จำนวน n ค่า หรือ $\theta_i = \{\theta_i^t \mid t=1, 2, \dots, n\}$ แล้วค่าสถิติสรุปผลที่สำคัญมีดังนี้

ข.5.1) *ค่าเฉลี่ยภายหลัง (posterior mean)*

$$E(\theta_i | y) \approx \bar{\theta}_i = \frac{1}{n} \sum_{t=1}^n \theta_i^t \quad (2.34)$$

ข.5.2) *ความแปรปรวนภายหลัง (posterior variance)*

$$\text{Var}(\theta_i | y) \approx s_i^2 = \frac{1}{n-1} \sum_{t=1}^n (\theta_i^t - \bar{\theta}_i)^2 \quad (2.35)$$

ทั้งนี้จากวิธีการประมาณค่าแบบเบย์ที่กล่าวมา อาจสังเกตได้ว่าการแจกแจงความน่าจะเป็นมีบทบาทสำคัญเป็นอย่างยิ่งต่อการกำหนดการแจกแจงก่อนหน้าเพื่อใช้ในการประมาณค่าพารามิเตอร์ ในที่นี้จึงขอยกตัวอย่างการแจกแจงความน่าจะเป็นที่สำคัญ ซึ่งอาจนำมาใช้ในการประมาณค่าแบบเบย์กับโมเดล MMixIRT ได้แก่ 1) การแจกแจงความน่าจะเป็นแบบไม่ต่อเนื่อง และ 2) การแจกแจงความน่าจะเป็นแบบต่อเนื่อง (กัลยา วานิชย์บัญชา, 2552; ดำรง ทิพย์โยธา, 2544) รายละเอียดดังนี้

1) การแจกแจงความน่าจะเป็นแบบไม่ต่อเนื่อง

1.1) การแจกแจงเบอร์นูลลี (Bernoulli distribution)

การแจกแจงเบอร์นูลลีเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ที่เกิดจากการทดลองเพียง 1 ครั้ง และในการทดลอง 1 ครั้ง มีผลได้ 2 แบบคือ ผลสำเร็จ (success) หรือผลล้มเหลว (failure) ด้วยความน่าจะเป็นที่เกิดผลสำเร็จเท่ากับ p และความน่าจะเป็นที่เกิดผลล้มเหลวเท่ากับ $1-p$ เขียนสัญลักษณ์แทนด้วย $X \sim \text{Ber}(p)$ มีฟังก์ชันความน่าจะเป็นดังสมการ (2.36)

$$P(X = x) = p^x (1-p)^{1-x} ; x = 0, 1 \quad (2.36)$$

เมื่อค่าคาดหวังและความแปรปรวนของ X คือ p และ $p(1-p)$ ตามลำดับ

1.2) การแจกแจงทวินาม (binomial distribution)

การแจกแจงทวินามเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ที่เกิดจากการทดลองจำนวน n ครั้ง ซึ่งเป็นอิสระจากกัน และในการทดลองแต่ละครั้งมีผลได้ 2 แบบคือ ผลสำเร็จ

หรือผลล้มเหลว ด้วยความน่าจะเป็นที่เกิดผลสำเร็จเท่ากับ p และความน่าจะเป็นที่เกิดผลล้มเหลวเท่ากับ $1-p$ เขียนสัญลักษณ์แทนด้วย $X \sim B(n, p)$ มีฟังก์ชันความน่าจะเป็นดังสมการ (2.37)

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = 0, 1, \dots, n \quad (2.37)$$

เมื่อค่าคาดหวังและความแปรปรวนของ X คือ np และ $np(1-p)$ ตามลำดับ

1.3) การแจกแจงพหุนาม (multinomial distribution)

การแจกแจงพหุนามเป็นการแจกแจงความน่าจะเป็นของเวกเตอร์ตัวแปรสุ่ม \underline{X} ที่เกิดจากการทดลองจำนวน n ครั้ง ซึ่งเป็นอิสระจากกัน และในการทดลองแต่ละครั้งมีผลได้ k แบบคือ E_1, E_2, \dots, E_k ด้วยความน่าจะเป็น p_1, p_2, \dots, p_k ตามลำดับ โดยในการทดลองมี E_1 เกิดขึ้น x_1 ครั้ง, E_2 เกิดขึ้น x_2 ครั้ง, \dots , E_k เกิดขึ้น x_k ครั้ง ดังนั้นถ้ากำหนด $\underline{X} = (X_1, X_2, \dots, X_k)$ และ $\underline{p} = (p_1, p_2, \dots, p_k)$ สามารถเขียนสัญลักษณ์แทนด้วย $\underline{X} \sim Mult(n, \underline{p})$ มีฟังก์ชันความน่าจะเป็นดังสมการ (2.38)

$$P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}; \quad \sum_{i=1}^k x_i = n, \quad \sum_{i=1}^k p_i = 1 \quad (2.38)$$

เมื่อค่าคาดหวังและความแปรปรวนของ X_i คือ np_i และ $np_i(1-p_i)$ ตามลำดับ

2) การแจกแจงความน่าจะเป็นแบบต่อเนื่อง

2.1) การแจกแจงยูนิฟอร์ม (uniform distribution)

การแจกแจงยูนิฟอร์มเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง X ที่มีค่าได้ทุกค่าจริงในช่วง $[a, b]$ ด้วยความน่าจะเป็นเท่า ๆ กัน เขียนสัญลักษณ์แทนด้วย $X \sim U(a, b)$ มีฟังก์ชันความหนาแน่นดังสมการ (2.39) หรือแสดงในเชิงกราฟได้ดังภาพ 2.16

$$f(x) = \frac{1}{b-a}; \quad x \in [a, b] \quad \text{และ} \quad f(x) = 0; \quad x \text{ มีค่าอื่น ๆ} \quad (2.39)$$

เมื่อค่าเฉลี่ยและความแปรปรวนของ X คือ $(a+b)/2$ และ $(b-a)^2/12$ ตามลำดับ

2.2) การแจกแจงปกติตัวแปรเดียว (univariate normal distribution)

การแจกแจงปกติตัวแปรเดียวเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง X ที่มีค่าส่วนใหญ่ใกล้เคียงกับค่าเฉลี่ยของตัวแปรชุดนั้น และบางส่วนมีค่ามากกว่าหรือน้อยกว่าค่าเฉลี่ย เขียนสัญลักษณ์แทนด้วย $X \sim N(\mu, \sigma^2)$ มีฟังก์ชันความหนาแน่นดังสมการ (2.40) หรือแสดงในเชิงกราฟได้ดังภาพ 2.17

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]; \quad x \in (-\infty, \infty) \quad (2.40)$$

เมื่อค่าเฉลี่ยและความแปรปรวนของ X คือ μ และ σ^2 ตามลำดับ ทั้งนี้ถ้า $\mu=0$ และ $\sigma^2=1$ จะเรียกว่า การแจกแจงปกติมาตรฐาน (standard normal distribution) หรือ $X \sim N(0, 1)$

2.3) การแจกแจงปกติหลายตัวแปร (multivariate normal distribution)

การแจกแจงปกติหลายตัวแปรเป็นการแจกแจงความน่าจะเป็นของเวกเตอร์ตัวแปรสุ่มต่อเนื่อง X ซึ่งเป็นรูปแบบทั่วไปของการแจกแจงปกติตัวแปรเดียว โดยมีการขยายมิติของค่าเฉลี่ยและความแปรปรวนเป็นเวกเตอร์ค่าเฉลี่ย $\underline{\mu}$ ขนาด $n \times 1$ และเมทริกซ์ความแปรปรวน Σ ขนาด $n \times n$ เขียนสัญลักษณ์แทนด้วย $X \sim MVN(\underline{\mu}, \Sigma)$ มีฟังก์ชันความหนาแน่นดังสมการ (2.41) หรือแสดงในเชิงกราฟได้ดังภาพ 2.18

$$f(x) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x - \underline{\mu})' \Sigma^{-1}(x - \underline{\mu})\right] \quad (2.41)$$

ทั้งนี้หากเวกเตอร์สุ่ม $X = (X_1, X_2, \dots, X_n)$ มีการแจกแจงแบบปกติหลายตัวแปร แล้วตัวแปรสุ่มแต่ละตัว X_1, X_2, \dots, X_n จะมีการแจกแจงแบบปกติด้วย แต่ในทางกลับกันอาจไม่เป็นจริงเสมอไป

2.4) การแจกแจงแกมมา (gamma distribution)

การแจกแจงแกมมาเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง X ที่มีพารามิเตอร์ $\alpha > 0$ และ $\beta > 0$ เขียนสัญลักษณ์แทนด้วย $X \sim \text{Gamma}(\alpha, \beta)$ มีฟังก์ชันความหนาแน่นดังสมการ (2.42) หรือแสดงในเชิงกราฟได้ดังภาพ 2.19

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp[-\beta x]; \quad x \in (0, \infty) \quad (2.42)$$

เมื่อค่าเฉลี่ยและความแปรปรวนของ X คือ α/β และ α/β^2 ตามลำดับ และ Γ คือฟังก์ชันที่มีคุณสมบัติดังนี้ $\Gamma(n) = (n-1)\Gamma(n-1)$, $n > 1$; $\Gamma(n) = (n-1)!$, $n \in \mathbb{N}$; $\Gamma(1/2) = \sqrt{\pi}$

2.5) การแจกแจงบีตา (beta distribution)

การแจกแจงบีตาเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง X ที่มีพารามิเตอร์ $\alpha > 0$ และ $\beta > 0$ เขียนสัญลักษณ์แทนด้วย $X \sim \text{Beta}(\alpha, \beta)$ มีฟังก์ชันความหนาแน่นดังสมการ (2.43) หรือแสดงในเชิงกราฟได้ดังภาพ 2.20

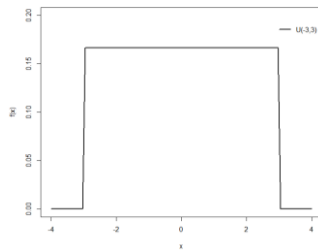
$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}; \quad x \in (0, 1) \quad (2.43)$$

เมื่อค่าเฉลี่ยและความแปรปรวนของ X คือ $\alpha/(\alpha + \beta)$ และ $\alpha\beta/[(\alpha + \beta)^2(\alpha + \beta + 1)]$ ตามลำดับ

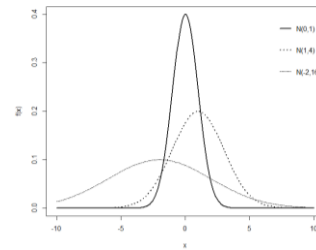
2.6) การแจกแจงดีริเชล (Dirichlet distribution)

การแจกแจงดีริเชลเป็นการแจกแจงความน่าจะเป็นของเวกเตอร์ตัวแปรสุ่มต่อเนื่อง X ซึ่งเป็นรูปแบบทั่วไปของการแจกแจงบีตา โดยมีการขยายมิติค่า α กลายเป็นเวกเตอร์ $\underline{\alpha}$ ขนาด $k \times 1$ เมื่อ $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k)$ และ $\alpha_i > 0$ เขียนสัญลักษณ์แทนด้วย $X \sim \text{Dir}(\underline{\alpha})$ มีฟังก์ชันความหนาแน่นดังสมการ (2.44) หรือแสดงในเชิงกราฟได้ดังภาพ 2.21

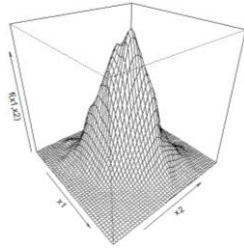
$$f(x) = \frac{1}{B(\underline{\alpha})} \prod_{i=1}^k x_i^{\alpha_i-1} \quad \text{และ} \quad B(\underline{\alpha}) = \frac{\prod_{i=1}^k \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^k \alpha_i\right)}; \quad x \in (0, 1), \quad \sum_{i=1}^k x_i = 1 \quad (2.44)$$



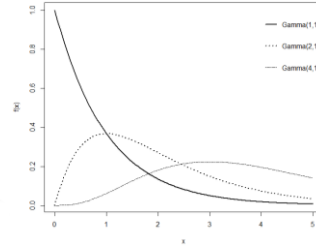
ภาพ 2.16 ตัวอย่างการแจกแจงยูนิฟอร์ม



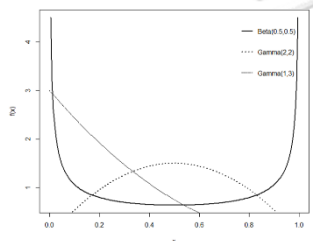
ภาพ 2.17 ตัวอย่างการแจกแจงปกติตัวแปรเดียว



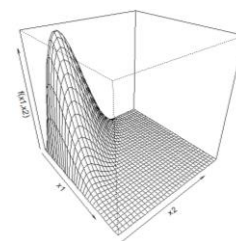
ภาพ 2.18 ตัวอย่างการแจกแจงปกติสองตัวแปร



ภาพ 2.19 ตัวอย่างการแจกแจงแกมมา



ภาพ 2.20 ตัวอย่างการแจกแจงบีตา



ภาพ 2.21 ตัวอย่างการแจกแจงดีกรีเคล

ตอนที่ 4 การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo simulation)

ในส่วนนี้เป็นการศึกษาเกี่ยวกับความเป็นมา ขั้นตอนการจำลองข้อมูล รวมถึงข้อดีและข้อจำกัดของการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ซึ่งผู้วิจัยจะนำเทคนิคดังกล่าวมาสร้างข้อมูลจำลองภายใต้สถานการณ์ต่าง ๆ รายละเอียดมีดังนี้

1) ความเป็นมาของการจำลองด้วยเทคนิคมอนติคาร์โล

การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลเป็นเทคนิคที่สร้างขึ้นจากทฤษฎีความน่าจะเป็น (probability theory) และทฤษฎีการสุ่ม (theory of random) โดยพัฒนาการของเทคนิคนี้สามารถแบ่งได้เป็น 3 ช่วง ได้แก่ ช่วงบุกเบิกเป็นช่วงที่นิยมใช้เทคนิคมอนติคาร์โลในแก้ปัญหาทางสถิติ เช่น การทดลองสุ่มตัวอย่างจากการแจกแจงทางสถิติในรูปแบบต่าง ๆ อย่างไรก็ตามเทคนิคนี้ยังไม่เป็นที่ยอมรับสำหรับการวิจัยในช่วงเวลาดังกล่าว ช่วงที่สองเป็นช่วงที่มีคำว่ามอนติคาร์โลเกิดขึ้น โดย Metropolis และ Ulam เป็นผู้เริ่มใช้ศัพท์คำนี้ โดยนำมาใช้เป็นเครื่องมือวิจัยในช่วงสงครามโลกครั้งที่ 2 เพื่อศึกษาปัญหาเกี่ยวกับการระเบิดปรมาณู จากการศึกษาดังกล่าวทำให้เทคนิคมอนติคาร์โลได้รับความนิยมมากขึ้น และช่วงที่สามเป็นช่วงที่นักวิจัยนิยมใช้เทคนิคมอนติคาร์โลในการจำลองข้อมูลด้วยเครื่องคอมพิวเตอร์เพื่อแก้ปัญหาทางการวิจัยและสถิติ

2) ขั้นตอนการจำลองข้อมูล

การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล เป็นวิธีการจำลองข้อมูลโดยอาศัยการสร้างโมเดล และการใช้เลขสุ่ม (random number) เข้ามาช่วยแก้ปัญหาหรือหาคำตอบ การศึกษาด้วยวิธีการจำลองข้อมูลโดยทั่วไปแบ่งเป็น 4 ขั้นตอนคือ 1) ระบุระบบหรือปัญหา 2) สร้างโมเดลเชิงคณิตศาสตร์ 3) จำลองข้อมูล และ 4) วิเคราะห์ผลลัพธ์ ดังภาพ 2.22



ภาพ 2.22 ขั้นตอนการจำลองข้อมูล

(ที่มา: เสกสรร เกียรติสุไพบูลย์, 2555)

จากขั้นตอนของการจำลองข้อมูลดังภาพ 2.22 สามารถอธิบายรายละเอียดได้ดังนี้

ก) การระบุระบบหรือปัญหา เป็นการนิยามปัญหาที่สนใจ ผู้วิจัยจะต้องทำความเข้าใจระบบหรือปัญหาที่ต้องการศึกษาเป็นอย่างดี สามารถแยกปัญหาที่ซับซ้อนออกเป็นส่วนย่อย และเข้าใจถึงความสัมพันธ์ระหว่างส่วนย่อยเหล่านั้น เพื่อนำไปสร้างโมเดลทางคณิตศาสตร์ที่ถูกต้องครบถ้วนตามนิยาม

ข) การสร้างโมเดลทางคณิตศาสตร์ เป็นการออกแบบโมเดลตามการนิยามปัญหา ผู้วิจัยต้องกำหนดตัวแปรต่าง ๆ ที่เกี่ยวข้อง รวมถึงเงื่อนไข ข้อกำหนด และความสัมพันธ์ระหว่างตัวแปรในรูปแบบสมการหรือสมการทางคณิตศาสตร์เพื่อใช้เป็นแบบจำลองในการอธิบายปัญหาที่ต้องการศึกษา

ค) การจำลอง เป็นการนำโมเดลทางคณิตศาสตร์มาจำลองบนคอมพิวเตอร์ ซึ่งในการจำลองข้อมูลแบ่งได้เป็น 3 ขั้นตอน ดังนี้

ค.1) การสร้างเลขสุ่ม เลขสุ่มที่สร้างขึ้นมักเป็นค่าตัวแปรสุ่มที่มีการแจกแจงยูนิฟอร์มบนช่วง $[0,1]$ และมีความอิสระจากกัน แบ่งเป็น 2 รูปแบบคือ 1) เลขสุ่มแท้ (true random number) ซึ่งมีวิธีสร้างได้หลายวิธี เช่น การจับสลาก การทอดลูกเต๋า การโยนเหรียญ การใช้วงล้อ อย่างไรก็ตามในทางปฏิบัติวิธีเหล่านี้ยังสร้างเลขสุ่มได้น้อยและสิ้นเปลืองเวลามาก และ 2) เลขสุ่มเทียม (pseudo-random number) ซึ่งมีวิธีการคำนวณมาจากสูตรทางคณิตศาสตร์ เลขสุ่มรูปแบบนี้จึงไม่มีคุณสมบัติความเป็นอิสระ และไม่ใชเลขสุ่มแท้ อย่างไรก็ตามแม้ว่าเลขสุ่มเทียมเหล่านี้จะมีรูปแบบที่แน่นอน แต่ในทางปฏิบัติเมื่อไม่ทราบถึงสูตรที่มา ก็จะเสมือนว่าไม่มีรูปแบบและไม่สามารถคาดเดาได้

ค.2) การประยุกต์ปัญหาที่ต้องการศึกษาเข้ามาใช้กับตัวเลขสุ่ม เป็นขั้นตอนที่ผู้วิจัยนำข้อมูลที่สร้างขึ้นตามปัญหาที่กำหนดไว้มาใช้กับตัวเลขสุ่ม

ค.3) การทดลองสุ่ม เป็นการนำโมเดลทางคณิตศาสตร์ที่กำหนดขึ้นตามแบบแผนการทดลองมาเข้าสู่กระบวนการสุ่มและทำในลักษณะซ้ำ ๆ กัน เพื่อประมาณค่าพารามิเตอร์ในโมเดล

ง) การวิเคราะห์ผลลัพธ์ เป็นการนำผลลัพธ์ที่ได้จากการจำลองมาวิเคราะห์ เนื่องจากการจำลองเป็นการจำลองสุ่ม ดังนั้นผลลัพธ์ที่ได้จากการจำลองจะมีลักษณะสุ่มด้วย ผู้วิจัยจึงควรเลือกใช้วิธีการทางสถิติที่เหมาะสมสำหรับการวิเคราะห์ผลลัพธ์ เพื่อให้การตีความมีความถูกต้องและน่าเชื่อถือ

3) ข้อดีและข้อจำกัดของการจำลองด้วยเทคนิคมอนติคาร์โล

ข้อดีและข้อจำกัดของการจำลองด้วยเทคนิคดังกล่าวมีดังนี้ (ชนะศึก นิชานนท์, 2553)

ก) ข้อดี

- เหมาะสมต่อการวิเคราะห์ปัญหาหรือสถานการณ์ที่ไม่สามารถกระทำได้ด้วยวิธีการเก็บรวบรวมข้อมูลในทางปฏิบัติ เนื่องจากปัญหาหรือสถานการณ์จริงอาจมีความซับซ้อนและมีค่าใช้จ่ายสูง

- สรุปอ้างอิงผลสู่ประชากรเป้าหมายได้ หากการจำลองอยู่ภายใต้สถานการณ์ใกล้เคียงกัน
- กำหนดและจัดกระทำค่าพารามิเตอร์ พร้อมทั้งศึกษาอิทธิพลต่าง ๆ ได้ในคราวเดียว
- กำหนดระยะเวลาที่แน่นอนในการดำเนินการทดลองได้

ข) ข้อจำกัด

- ผลจากการจำลองข้อมูลอาจมีข้อจำกัดและไม่ครอบคลุมทุกกรณี
- การนำผลจากการจำลองไปใช้ ต้องคำนึงถึงความสอดคล้องกับสถานการณ์
- การใช้วิธีที่ไม่เหมาะสมในการจำลอง อาจทำให้ผลการวิจัยไม่สามารถนำไปใช้ประโยชน์ได้
- ความถูกต้องและน่าเชื่อถือของผลการจำลองขึ้นอยู่กับจำนวนรอบการทำซ้ำและความถูกต้องของตัวเลขที่ได้จากการจำลองข้อมูล

ตอนที่ 5 การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET)

ในส่วนนี้เป็นการนำเสนอข้อมูลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ซึ่งเป็นข้อมูลเชิงประจักษ์ที่ผู้วิจัยต้องการนำมาศึกษา การนำเสนอแบ่งเป็น 4 ประเด็น ได้แก่ 1) ความเป็นมาและความสำคัญของการทดสอบ 2) ข้อจำกัดในการใช้ประโยชน์จากผลสอบ 3) รูปแบบของการทดสอบวิชาคณิตศาสตร์ และ 4) ตัวแปรที่ใช้ในการจำแนกกลุ่มโรงเรียนและนักเรียน รายละเอียดดังนี้

1) ความเป็นมาและความสำคัญของการทดสอบ

กระทรวงศึกษาธิการได้กำหนดให้มีการวัดและประเมินผลการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ไว้ 4 ระดับ (กระทรวงศึกษาธิการ, 2552) ดังนี้

ก) การประเมินระดับชั้นเรียน เป็นการวัดและประเมินกระบวนการจัดการเรียนรู้ ซึ่งมีผู้สอนดำเนินการจัดการเรียนการสอนและทำหน้าที่ประเมินผลเอง โดยอาจเปิดโอกาสให้ผู้เรียนประเมินตนเอง เพื่อนประเมินเพื่อน หรือผู้ปกครองร่วมประเมิน เพื่อตรวจสอบพัฒนาการการเรียนรู้ของผู้เรียน

ข) การประเมินระดับสถานศึกษา เป็นการตรวจสอบผลการเรียนรู้ของผู้เรียนเป็นรายปีหรือรายภาค เพื่อประเมินการจัดการศึกษาของสถานศึกษาตามเป้าหมายที่กำหนด ซึ่งจะใช้เป็น

สารสนเทศในการนำไปปรับปรุงหลักสูตร วิธีการจัดการเรียนการสอน รวมถึงการจัดทำแผนพัฒนาคุณภาพการศึกษาตามแนวทางการประกันคุณภาพการศึกษา

ค) การประเมินระดับเขตพื้นที่ เป็นการประเมินคุณภาพผู้เรียนในระดับเขตพื้นที่การศึกษา เพื่อใช้เป็นข้อมูลในการพัฒนาคุณภาพการศึกษาของเขตพื้นที่การศึกษาตามภาระรับผิดชอบ

ง) การประเมินระดับชาติ เป็นการประเมินคุณภาพผู้เรียนในระดับชาติ โดยสถานศึกษาต้องจัดให้ผู้เรียนทุกคนที่เรียนในระดับชั้น ป.6 ม.3 และ ม.6 เข้ารับการประเมิน ผลจากการประเมินจะถูกใช้เพื่อเทียบเคียงคุณภาพการศึกษาในระดับต่าง ๆ และนำไปวางแผนยกระดับคุณภาพการจัดการศึกษา รวมถึงเป็นข้อมูลสนับสนุนการตัดสินใจในระดับนโยบายของประเทศ

สถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.) รับผิดชอบเป็นองค์กรกลางในการศึกษา วิจัย และให้บริการเกี่ยวกับการทดสอบและการประเมินผลทางการศึกษา โดยที่ผ่านมา สทศ. ได้มีการจัดทดสอบหลายประเภท อาทิ การทดสอบวิชาความถนัดทั่วไป (GAT: general aptitude test) เพื่อประเมินความพร้อมในการเข้าเรียนระดับมหาวิทยาลัย การทดสอบวิชาความถนัดทางวิชาการและวิชาชีพ (PAT: professional and academic aptitude test) เพื่อประเมินแนวเข้าสู่เส้นทางอาชีพ ในคณะที่เลือก และการทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน (O-NET) เพื่อประเมินความรู้และความคิดของนักเรียน อย่างไรก็ตามงานวิจัยนี้จะขอกล่าวถึงเฉพาะ O-NET เนื่องจากเป็นการทดสอบที่นักเรียนส่วนใหญ่ต้องเข้ารับการประเมิน ผลการสอบจึงมีความสำคัญต่อการตัดสินใจสมัครสอบทางการเรียนรู้อันของผู้เรียน และอาจนำไปใช้เป็นสารสนเทศในการประเมินคุณภาพสถานศึกษา

การทดสอบ O-NET ถือเป็นประเมินระดับชาติ แบบสอบที่นำมาใช้จึงเป็นชุดเดียวกันหรือคู่ขนานกัน และจัดสอบในวันเดียวกันทั่วทั้งประเทศ ภายใต้การกำกับดูแลรับผิดชอบจาก สทศ. โดยผลการสอบจะถูกนำไปประเมินการผ่านชั้นเรียนหรือเป็นส่วนหนึ่งของการคัดเลือกเพื่อเข้าศึกษาต่อในระดับ ม.1 และ ม.4 ในโรงเรียนที่มีอัตราการแข่งขันสูง และเป็นส่วนหนึ่งของการพิจารณาเพื่อรับนักเรียนเข้าศึกษาต่อในระดับมหาวิทยาลัย นอกจากนี้ยังอาจใช้เป็นเกณฑ์ในการประกันคุณภาพภายในและภายนอกของสถานศึกษา

เนื้อหาของ การทดสอบ O-NET จะครอบคลุมตามข้อกำหนดการวัดและประเมินผลการเรียนรู้ระดับชาติตามมาตรฐานการเรียนรู้ของหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ประกอบด้วย 5 วิชา ได้แก่ 1) คณิตศาสตร์ 2) วิทยาศาสตร์ 3) ภาษาอังกฤษ 4) ภาษาไทย และ 5) สังคมศึกษา ศาสนาและวัฒนธรรม (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2558)

2) ข้อกำหนดในการวัดและประเมินผลการสอบ

การวัดและประเมินผลการทดสอบ O-NET โดยทั่วไปพบว่ายังคงยึดตามแนวทางการทดสอบแบบดั้งเดิมเป็นหลัก อาทิ การวัดผลโดยอิงจากคะแนนรวมของแบบสอบทั้งฉบับ การตัดสินความรู้ของผู้เรียนหรือการประเมินคุณภาพสถานศึกษาโดยเทียบจากคะแนนเฉลี่ยระดับประเทศหรือคะแนน

เฉลี่ยตามสังกัดของสถานศึกษา เป็นต้น ซึ่งการวัดและประเมินลักษณะดังกล่าวมักให้ผลแปรเปลี่ยนไปตามกลุ่มผู้สอบหรือชุดข้อสอบที่ใช้ (ศิริชัย กาญจนวาสี, 2555)

เพื่อให้การวัดมีความสอดคล้องกับธรรมชาติของข้อมูลทางการศึกษา และได้สารสนเทศที่สามารถนำไปประเมินเพื่อวินิจฉัย วางแผน หรือปรับปรุงหลักสูตรการเรียนการสอนในโรงเรียน ซึ่งอาจช่วยให้นักเรียนมีผลสัมฤทธิ์ทางการเรียนที่สูงขึ้น งานวิจัยนี้จึงต้องการนำข้อมูลคะแนนสอบ O-NET วิชาคณิตศาสตร์ ม.6 มาศึกษาตามแนวทางการทดสอบแบบใหม่ด้วยโมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ (MMixIRT) โดยเหตุผลในการเลือกข้อมูลดังกล่าวมาศึกษา เนื่องจากผลการรายงานที่ผ่านมามีพบว่านักเรียนส่วนใหญ่ของประเทศยังคงอ่อนข้างอ่อนในวิชาคณิตศาสตร์ สังเกตได้จากคะแนนเฉลี่ยระดับประเทศที่มีค่าค่อนข้างต่ำเพียง 25.41 (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2562ข) อย่างไรก็ตามหากพิจารณาตามสังกัดของสถานศึกษากลับพบว่านักเรียนจากสังกัดกระทรวงการอุดมศึกษา (อว.) มีคะแนนเฉลี่ยสูงถึง 48.32 โดยมีคะแนนเฉลี่ยในสาระที่ 1-4 เท่ากับ 6.39, 3.67, 23.32 และ 11.56 ตามลำดับ ขณะที่นักเรียนจากโรงเรียนในสังกัดอื่น เช่น สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) มีคะแนนเฉลี่ยค่อนข้างต่ำเพียง 25.62 โดยมีคะแนนเฉลี่ยในสาระที่ 1-4 เท่ากับ 3.62, 1.90, 13.43 และ 6.84 ตามลำดับ จากผลดังกล่าวจึงอนุมานได้ว่าน่าจะมีความวิวิธพันธุ์เกิดขึ้นในกลุ่มประชากรผู้สอบ และอาจแบ่งประชากรออกเป็นกลุ่มย่อยหรือกลุ่มแฝงตามแบบแผนการตอบข้อสอบที่แตกต่างกันได้ ดังนั้นข้อมูลจึงมีความเหมาะสมสำหรับการนำมาศึกษาด้วยโมเดล MMixIRT

3) รูปแบบของการทดสอบวิชาคณิตศาสตร์

การทดสอบ O-NET วิชาคณิตศาสตร์ ปีการศึกษา 2562 ระดับ ม.6 มีรูปแบบดังนี้

- 1) แบบสอบประกอบด้วยข้อสอบปรนัยแบบ 5 ตัวเลือกที่มีตัวถูกเพียงตัวเดียว 32 ข้อ และข้อสอบอัตนัยแบบเติมตัวเลข 8 ข้อ อย่างไรก็ตามงานวิจัยนี้จะใช้ข้อมูลจากข้อสอบปรนัยเท่านั้น
- 2) ข้อสอบทุกข้อมีคะแนนเท่ากันคือข้อละ 2.5 คะแนน
- 3) ข้อสอบครอบคลุม 4 สาระการเรียนรู้ คือ สาระที่ 1 จำนวนและการดำเนินการ (5 ข้อ) สาระที่ 2 การวัด (2 ข้อ) สาระที่ 4 พีชคณิต (16 ข้อ) และสาระที่ 5 การวิเคราะห์ข้อมูลและความน่าจะเป็น (9 ข้อ) รายละเอียดดังตาราง 2.9

ตาราง 2.9 จำนวนข้อสอบส่วนปรนัย O-NET วิชาคณิตศาสตร์ ระดับ ม.6

สาระการเรียนรู้/มาตรฐาน	จำนวนข้อ	คะแนน
สาระที่ 1 จำนวนและการดำเนินการ		
- มาตรฐาน ค 1.1 เข้าใจถึงความหลากหลายของการแสดงจำนวนและการใช้จำนวนในชีวิตจริง	2	5.0
- มาตรฐาน ค 1.2 เข้าใจถึงผลที่เกิดขึ้นจากการดำเนินการของจำนวนและความสัมพันธ์ระหว่างการดำเนินการต่าง ๆ และสามารถใช้อการดำเนินการในการแก้ปัญหา	1	2.5
- มาตรฐาน ค 1.4 เข้าใจระบบจำนวนและนำเสนอสมบัติเกี่ยวกับจำนวนไปใช้	2	5.0
สาระที่ 2 การวัด		
- มาตรฐาน ค 2.2 แก้ปัญหาเกี่ยวกับการวัด	2	5.0
สาระที่ 4 พีชคณิต		
- มาตรฐาน ค 4.1 เข้าใจและวิเคราะห์แบบรูป (pattern) ความสัมพันธ์และฟังก์ชัน	7	17.5
- มาตรฐาน ค 4.2 ใช้นิพจน์ สมการ อสมการ กราฟ และตัวแบบเชิงคณิตศาสตร์อื่น ๆ แทนสถานการณ์ต่าง ๆ ตลอดจนแปลความหมายและนำไปใช้แก้ปัญหา	9	22.5
สาระที่ 5 การวิเคราะห์ข้อมูลและความน่าจะเป็น		
- มาตรฐาน ค 5.1 เข้าใจและใช้วิธีการทางสถิติในการวิเคราะห์ข้อมูล	5	12.5
- มาตรฐาน ค 5.2 ใช้วิธีการทางสถิติและความรู้เกี่ยวกับความน่าจะเป็นในการคาดการณ์ได้อย่างสมเหตุสมผล	4	10.0
รวม	32	80.0

(ที่มา: สถาบันทดสอบทางการศึกษาแห่งชาติ, 2562ก)

4) ตัวแปรที่ใช้ในการจำแนกกลุ่มโรงเรียนและนักเรียน

ภายหลังจากการทดสอบ O-NET ดำเนินการเสร็จสิ้น สทศ. จะดำเนินการบันทึกข้อมูลคะแนนสอบรายข้อ (0, 1) ข้อมูลของโรงเรียน (รหัสโรงเรียน, สังกัด, ขนาดโรงเรียน, เขตโรงเรียน, ภูมิภาค, จังหวัด, อำเภอ) และข้อมูลของนักเรียน (รหัสนักเรียน, เพศ, รายได้ผู้ปกครอง) จัดเก็บเป็นฐานข้อมูลเพื่อใช้ในการทำรายงานสรุปผลการทดสอบและเผยแพร่ทางเว็บไซต์ของ สทศ. ต่อไป อย่างไรก็ตามในงานวิจัยนี้จะขอนำข้อมูลเพียงบางส่วนจากฐานข้อมูลดังกล่าวมากำหนดเป็นตัวแปรเพื่อใช้ในการพิจารณาร่วมกับการจำแนกกลุ่มโรงเรียนและระดับโรงเรียนและระดับนักเรียนในส่วนของการศึกษาจากข้อมูลเชิงประจักษ์ต่อไป รายละเอียดดังตาราง 2.10

ตาราง 2.10 ตัวแปรที่ใช้ในการจำแนกกลุ่มโรงเรียนและนักเรียน

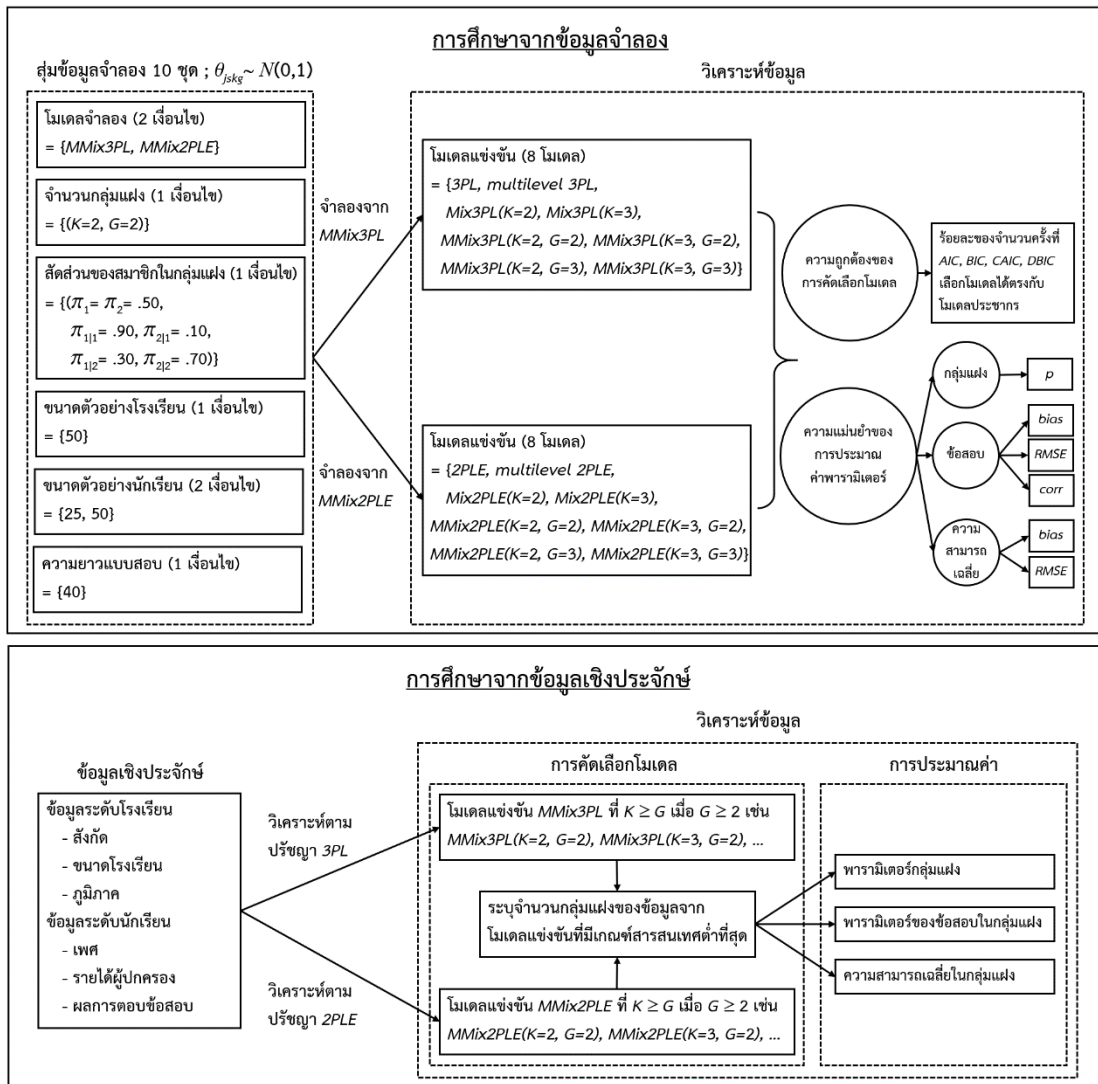
ตัวแปรระดับโรงเรียน		ตัวแปรระดับนักเรียน	
1. สังกัด	3. ภูมิภาค	1. เพศ	3. ผลการตอบข้อสอบ
- สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.)	- ใต้	- ชาย	- ถูก (1)
- สำนักงานคณะกรรมการส่งเสริมการศึกษาเอกชน (สช.)	- กรุงเทพฯ	- หญิง	- ผิด (0)
- สำนักงานคณะกรรมการการอุดมศึกษา (อว.)	- กลาง	2. รายได้ผู้ปกครอง	
- กรมส่งเสริมการปกครองท้องถิ่น (กสท.)	- ตะวันออก	- 1-50,000 บาท	
2. ขนาดโรงเรียน	- เหนือ	- 50,001-100,000 บาท	
- เล็ก	- ตะวันออกเฉียงเหนือ	- 100,001-200,000 บาท	
- กลาง	- ตะวันตก	- 200,001-300,000 บาท	
- ใหญ่		- มากกว่า 300,000 บาท	
- ใหญ่พิเศษ			

ตอนที่ 6 กรอบแนวคิดในการวิจัย

ผู้วิจัยได้พัฒนากรอบแนวคิดในการวิจัยเพื่อศึกษาเกณฑ์สารสนเทศที่เหมาะสมสำหรับการคัดเลือกโมเดล ความแม่นยำของการประมาณค่าพารามิเตอร์ และเพื่อหาวิธีการใช้โมเดลกับข้อมูลเชิงประจักษ์ด้วยโมเดล MMix3PL และ MMix2PLE การศึกษาแบ่งเป็น 2 ส่วน ได้แก่ 1) การศึกษาภายใต้สถานการณ์จำลอง และ 2) การศึกษาจากข้อมูลเชิงประจักษ์ รายละเอียดดังนี้

1) การศึกษาภายใต้สถานการณ์จำลอง เป็นการนำโมเดล MMix3PL และ MMix2PLE มาศึกษาภายใต้สถานการณ์จำลองด้วยเทคนิคมอนติคาร์โล เพื่อเปรียบเทียบความถูกต้องของการคัดเลือกโมเดลระหว่างเกณฑ์สารสนเทศ AIC, BIC, CAIC และ DBIC และเพื่อศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง โดยพิจารณาจากค่าร้อยละของความใกล้เคียง (p) ค่าสหสัมพันธ์ (corr) ค่า bias และ RMSE ทั้งนี้ ข้อมูลจะถูกจำลองจากเงื่อนไขต่าง ๆ ได้แก่ โมเดลจำลอง จำนวนกลุ่มแฝง สัดส่วนสมาชิกในกลุ่มแฝง ขนาดตัวอย่างโรงเรียนและนักเรียน ความยาวแบบสอบ และโมเดลวิเคราะห์ รวม 32 เงื่อนไข แต่ละเงื่อนไขจะถูกจำลองซ้ำ 10 รอบ ทำให้มีจำนวนชุดข้อมูลที่ต้องวิเคราะห์ 320 ชุด ดังภาพ 2.23 (บน)

2) การศึกษาจากข้อมูลเชิงประจักษ์ เป็นการนำข้อมูลคะแนนสอบ O-NET วิชาคณิตศาสตร์ ระดับ ม.6 มาวิเคราะห์ด้วยโมเดล MMix3PL และ MMix2PLE เพื่อหาจำนวนกลุ่มแฝง พารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง โดยในการคัดเลือกโมเดลเพื่อหาโมเดลที่สอดคล้องกับข้อมูลมากที่สุด จะพิจารณาเฉพาะโมเดลแข่งขันที่มีจำนวนกลุ่มแฝงระดับนักเรียน (K) มากกว่าหรือเท่ากับจำนวนกลุ่มแฝงระดับโรงเรียน (G) เมื่อ $G \geq 2$ เนื่องจากในทางทฤษฎีจำนวนกลุ่มแฝงระดับโรงเรียนจะถูกจำแนกตามค่าสัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียนที่แตกต่างกัน ดังนั้นในทางปฏิบัติจึงเป็นไปได้ยากที่จำนวนกลุ่มแฝงระดับนักเรียนจะมีน้อยกว่าจำนวนกลุ่มแฝงระดับโรงเรียน (Cho, 2007) ดังภาพ 2.23 (ล่าง)



ภาพ 2.23 กรอบแนวคิดในการวิจัย

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยนี้มีวัตถุประสงค์ 4 ประการ ได้แก่ 1) เพื่อศึกษารูปแบบของโมเดล MMix2PLE และวิธีการประมาณค่าพารามิเตอร์ 2) เพื่อเปรียบเทียบความถูกต้องของการคัดเลือกโมเดลระหว่างเกณฑ์สารสนเทศ AIC, BIC, CAIC และ DBIC สำหรับโมเดล MMix3PL และ MMix2PLE ภายใต้สถานการณ์จำลอง 3) เพื่อศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง ระหว่างโมเดล MMix3PL กับ MMix2PLE ภายใต้สถานการณ์จำลอง และ 4) เพื่อวิเคราะห์หาจำนวนกลุ่มแฝง พารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง จากข้อมูลเชิงประจักษ์ด้วยโมเดล MMix3PL และ MMix2PLE การดำเนินการวิจัยแบ่งออกเป็น 2 ส่วนดังนี้

ส่วนที่ 1 เป็นการนำโมเดล MMix3PL และ MMix2PLE มาศึกษาภายใต้สถานการณ์จำลองเพื่อหาเกณฑ์สารสนเทศที่มีความเหมาะสมสำหรับการคัดเลือกโมเดลให้กับโมเดล MMix3PL และ MMix2PLE โดยจะเปรียบเทียบกันระหว่างเกณฑ์สารสนเทศ 4 ตัว ได้แก่ AIC, BIC, CAIC และ DBIC และเพื่อศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์ในโมเดล MMix3PL และ MMix2PLE ได้แก่ พารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบในกลุ่มแฝง และความสามารถเฉลี่ยในกลุ่มแฝง

ส่วนที่ 2 เป็นการสาธิตการใช้โมเดลกับข้อมูลเชิงประจักษ์ โดยจะนำข้อมูลผลการตอบข้อสอบ O-NET วิชาคณิตศาสตร์ ระดับ ม.6 มาวิเคราะห์ด้วยโมเดล MMix3PL และ MMix2PLE ที่มีจำนวนกลุ่มแฝงระดับโรงเรียนและระดับนักเรียนแตกต่างกัน เพื่อวิเคราะห์หาจำนวนกลุ่มแฝงของข้อมูลพารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบในกลุ่มแฝง และความสามารถเฉลี่ยในกลุ่มแฝง จากนั้นจะนำผลการวิเคราะห์ที่ได้จากทั้งสองโมเดลมาศึกษาเปรียบเทียบกัน

ส่วนที่ 1 การศึกษาภายใต้สถานการณ์จำลอง

ในส่วนนี้เป็นการตรวจสอบประสิทธิภาพของขั้นตอนวิธีที่ใช้ในการประมาณค่า และความถูกต้องของการคัดเลือกโมเดลของเกณฑ์สารสนเทศ ภายใต้สถานการณ์จำลองด้วยเทคนิคมอนติคาร์โล ผู้วิจัยขอแบ่งการพิจารณาเป็น 4 ด้าน ได้แก่ 1) การออกแบบสถานการณ์การจำลองข้อมูล 2) การประมาณค่าพารามิเตอร์ และการตรวจสอบความเหมาะสมของการประมาณค่า 3) เกณฑ์การพิจารณาความถูกต้องและความแม่นยำ และ 4) สถิติที่ใช้ในการวิเคราะห์ รายละเอียดดังนี้

1) การออกแบบสถานการณ์การจำลองข้อมูล

1.1) โมเดลจำลอง กำหนดโมเดลประชากร (population model) เพื่อจำลองข้อมูลเป็นโมเดล MMix3PL และ MMix2PLE (ดูสมการ (3.16) และ (3.19) ตามลำดับ)

1.2) จำนวนกลุ่มแฝง กำหนดข้อมูลเป็น 2 ระดับ ได้แก่ ระดับกลุ่ม (ระดับโรงเรียน) และระดับหน่วย (ระดับนักเรียน) โดยจำลองให้ระดับโรงเรียนมีจำนวนกลุ่มแฝง (G) 2 กลุ่ม และระดับนักเรียนมีจำนวนกลุ่มแฝง (K) 2 กลุ่ม ทั้งนี้เหตุผลในการเลือกจำนวนดังกล่าว เนื่องจากการศึกษาข้อมูลที่มีกลุ่มแฝงโดยทั่วไปสามารถแบ่งได้เป็น 3 กรณี ดังตาราง 3.1

ตาราง 3.1 สถานการณ์การศึกษาเมื่อข้อมูลมีจำนวนกลุ่มแฝงแตกต่างกัน

จำนวนกลุ่มแฝง ระดับนักเรียน	จำนวนกลุ่มแฝงระดับโรงเรียน		
	$G = 1$	$G = 2$	$G = 3$
$K = 1$	IRT หรือ multilevel IRT	-	-
$K = 2$	MixIRT	MMixIRT	MMixIRT
$K = 3$	MixIRT	MMixIRT	MMixIRT

จากตาราง 3.1 แต่ละสถานการณ์สามารถอธิบายได้ดังนี้

- เมื่อ $K = 1$ และ $G = 1$ (ไม่มีกลุ่มแฝง) → วิเคราะห์ด้วย IRT หรือ multilevel IRT
- เมื่อ $K = 2, 3$ และ $G = 1$ (มีกลุ่มแฝงระดับนักเรียน) → วิเคราะห์ด้วย MixIRT
- เมื่อ $K \geq 2$ และ $G \geq 2$ (มีกลุ่มแฝงระดับนักเรียนและโรงเรียน) → วิเคราะห์ด้วย MMixIRT

ส่วนกรณี $K = 1$ และ $G \geq 2$ จะไม่นำมาศึกษาในงานวิจัย เนื่องจากในทางทฤษฎีจำนวนกลุ่มแฝงระดับโรงเรียนจะถูกจำแนกตามสัดส่วนของสมาชิกในกลุ่มแฝงระดับนักเรียนที่แตกต่างกัน จึงเป็นไปได้ที่กลุ่มแฝงระดับโรงเรียนจะมีมากกว่าหนึ่งกลุ่ม ในขณะที่กลุ่มแฝงระดับนักเรียนมีเพียงกลุ่มเดียว นอกจากนี้กรณี $K = 2$ และ $G = 3$ ยังมีโอกาสเกิดขึ้นได้ยากในเชิงปฏิบัติ เนื่องจากมีโอกาสน้อยมากที่กลุ่มแฝงระดับนักเรียนจะแตกต่างกันเพียง 2 กลุ่ม ในขณะที่กลุ่มแฝงระดับโรงเรียนแตกต่างกันถึง 3 กลุ่ม (Cho, 2007) จากเหตุผลดังกล่าว งานวิจัยนี้จึงจำลองข้อมูลให้มีกลุ่มแฝงระดับโรงเรียน (G) 2 กลุ่ม และระดับนักเรียน (K) 2 กลุ่ม ซึ่งเป็นสถานการณ์ขั้นพื้นฐานสำหรับโมเดล MMixIRT

1.3) สัดส่วนของสมาชิกในกลุ่มแฝง กำหนดค่าสัดส่วนในระดับโรงเรียนทั้งสองกลุ่มเท่ากันคือ .50 โดยในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 ($g = 1$) สมมติให้กลุ่มแฝงระดับนักเรียนกลุ่มแรก ($k = 1$) เป็นกลุ่มเด่นที่มีค่าสัดส่วนเป็น .90 ส่วนในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 2 ($g = 2$) สมมติให้กลุ่มแฝงระดับนักเรียนกลุ่มที่สอง ($k = 2$) เป็นกลุ่มเด่นที่มีค่าสัดส่วนเป็น .70 ซึ่งค่าดังกล่าวเป็นการกำหนดในลักษณะใกล้เคียงกับงานวิจัยของ Cho และ Cohen (2010) ดังตาราง 3.2

ตาราง 3.2 ค่าจำลองสัดส่วนของสมาชิกในกลุ่มแฝง

สัดส่วนของสมาชิกใน กลุ่มแฝงระดับนักเรียน	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 ขนาด คือ 1,250 และ 2,500 คน คำนวณมาจากผลคูณระหว่างขนาดตัวอย่างของโรงเรียนคือ 50 โรงเรียน กับขนาดตัวอย่างของนักเรียนคือ 25 และ 50 คน โดยเหตุผลในการเลือกขนาดตัวอย่างของโรงเรียนเป็น 50 เนื่องจากจำนวนดังกล่าวเป็นที่นิยมสำหรับการวิจัยเชิงทดลองทั่วไป (Lee et al., 2018) ส่วนการเลือกขนาดตัวอย่างของนักเรียนเป็น 25 และ 50 เนื่องจากโดยทั่วไปพบว่าโรงเรียนแต่ละแห่งจะมีนักเรียนเข้าร่วมการทดสอบระดับชาติอย่างน้อย 25 คนโดยประมาณ (Cho, 2007) และ 50 คือจำนวนที่ใช้ศึกษาเมื่อนักเรียนมีจำนวนมากขึ้น

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

1.6) ความยาวแบบสอบ กำหนดเป็น 40 ข้อ เพื่อให้ใกล้เคียงกับแบบสอบมาตรฐานที่ใช้ในการสอบระดับชาติ โดยสมมติให้มีการวัดสาระการเรียนรู้ 4 ด้าน แต่ละด้านวัดด้วยข้อสอบ 10 ข้อ โดยสาระการเรียนรู้ที่ 1-4 วัดจากข้อ 1-10, 11-20, 21-30 และ 31-40 ตามลำดับ

1.7) พารามิเตอร์ของข้อสอบ กำหนดระดับสมรรถภาพของผู้สอบในกลุ่มแฝงสำหรับจำลองค่าพารามิเตอร์ของข้อสอบดังนี้ กลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 ($g = 1$) สมมติให้กลุ่มแฝงระดับนักเรียนกลุ่มแรก ($k = 1$) มีสมรรถภาพในสาระที่ 1-4 ระดับสูงมาก สูง ต่ำ และต่ำมากตามลำดับ และกลุ่มแฝงระดับนักเรียนกลุ่มที่สอง ($k = 2$) มีสมรรถภาพค่อนข้างตรงข้ามกับกลุ่มแรก ส่วนกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 2 ($g = 2$) สมมติให้กลุ่มแฝงระดับนักเรียนกลุ่มแรกมีสมรรถภาพต่ำกว่าสำหรับสาระที่ 1-2 และสูงกว่าสำหรับสาระที่ 3-4 เมื่อเทียบกับกลุ่มแฝงระดับนักเรียนกลุ่มแรกในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 และกลุ่มแฝงระดับนักเรียนกลุ่มที่สอง สมมติให้มีสมรรถภาพเหมือนกลุ่มแฝงระดับนักเรียนกลุ่มที่สองในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 ทั้งหมด ยกเว้นสาระที่ 4 ที่มีสมรรถภาพสูงและต่ำคละกัน ดังตาราง 3.3 และจำลองค่าพารามิเตอร์ของข้อสอบที่สมนัยกับระดับสมรรถภาพดังกล่าว ดังตาราง 3.4 ซึ่งการกำหนดค่าดังกล่าวมีลักษณะใกล้เคียงกับงานวิจัยก่อนหน้าคือ Cho (2007) และ Li และคณะ (2009)

1.8) โมเดลแข่งขัน กำหนดโมเดลเพื่อศึกษาความถูกต้องของการคัดเลือกโมเดลไว้ 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 (ดูตาราง 3.1)

จากสถานการณ์จำลองที่กล่าวมา มีเงื่อนไขการศึกษาดังนี้ 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

ตาราง 3.3 การจำลองรูปแบบสมรรถภาพของผู้สอบในกลุ่มแฝง

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 3.4 ค่าจำลองพารามิเตอร์ของข้อสอบในกลุ่มแฝง

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

2) การประมาณค่าพารามิเตอร์ และการตรวจสอบความเหมาะสมของการประมาณค่า

จากการทบทวนวรรณกรรมที่ผ่านมา พบว่าการประมาณค่าพารามิเตอร์สำหรับโมเดล MMixIRT มี 2 วิธี ได้แก่ 1) วิธีภาวะควรจะเป็นสูงสุด (ML) และ 2) วิธีการประมาณค่าแบบเบส์ อย่างไรก็ตามเนื่องจากเบส์ค่อนข้างมีจุดเด่นเหนือกว่าเมื่อเทียบกับวิธี ML อาทิ ขั้นตอนวิธีการคำนวณมีความซับซ้อนน้อยกว่า การนำสารสนเทศจากข้อมูลเชิงประจักษ์เข้ามาช่วยปรับปรุงการแจกแจงก่อนหน้าทำให้ผลการประมาณค่ามีความสมบูรณ์กว่า รวมถึงค่าประมาณที่ได้ยังไม่มีอคติ (unbiased) และมีความแกร่งในการประมาณค่าเมื่อข้อมูลมีการสูญหาย (Maier, 2001; Patz & Junker, 1999) หรือเมื่อขนาดตัวอย่างในกลุ่มแฝงมีขนาดเล็ก เนื่องจากวิธีการประมาณค่าแบบเบส์ไม่มีข้อตกลงเกี่ยวกับความเป็นปกติของข้อมูล (Frühwirth-Schnatter, 2006) จากเหตุผลดังกล่าวผู้วิจัยจึงขอเลือกใช้วิธีการประมาณค่าแบบเบส์ โดยจะดำเนินการวิเคราะห์ด้วยโปรแกรม MultiBUGS

ทั้งนี้ในการวิเคราะห์ข้อมูลโดยทั่วไป จำนวนกลุ่มแฝงจะถือว่าเป็นสิ่งที่ไม่สามารถทราบได้ล่วงหน้า ในทางปฏิบัติจึงต้องมีการวิเคราะห์เชิงสำรวจเพื่อตรวจสอบจำนวนกลุ่มแฝงของข้อมูลที่อาจเป็นไปได้ก่อน ตัวอย่างเช่น ถ้าข้อมูลไม่มีกลุ่มแฝงทั้งระดับนักเรียนและระดับโรงเรียน โมเดล IRT หรือ multilevel IRT ก็น่าจะมีแนวโน้มสอดคล้องกับข้อมูลมากที่สุด หรือหากข้อมูลมีเฉพาะกลุ่มแฝงระดับนักเรียน โมเดล MixIRT ก็น่าจะมีแนวโน้มสอดคล้องกับข้อมูลมากที่สุด เป็นต้น

จากแนวคิดดังกล่าว การวิเคราะห์ข้อมูลจึงจำเป็นต้องทดลองนำโมเดลที่มีจำนวนกลุ่มแฝงต่าง ๆ มาวิเคราะห์ข้อมูลแข่งขันกัน เพื่อคัดเลือกหาโมเดลที่มีจำนวนกลุ่มแฝงสอดคล้องกับข้อมูลมากที่สุด ในส่วนนี้ผู้วิจัยจึงได้ออกแบบขั้นตอนวิธีการประมาณค่าแบบเบส์สำหรับโมเดล IRT ไว้ 8 โมเดล ได้แก่ 1) 3PL 2) 2PLE 3) multilevel 3PL 4) multilevel 2PLE 5) Mix3PL 6) Mix2PLE 7) MMix3PL และ 8) MMix2PLE รายละเอียดดังนี้

2.1) การประมาณค่าพารามิเตอร์

ก) โมเดล 3PL

โมเดล จากสมการ (2.3) ทราบว่าโมเดล 3PL มีรูปแบบดังสมการ (3.1)

$$P(y_{ij}=1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]} \quad (3.1)$$

เมื่อ $\theta_j \sim N(0, \sigma^2)$ อย่างไรก็ตามเพื่อความสะดวกในการเขียนชุดคำสั่งในโปรแกรม MultiBUGS งานวิจัยนี้จะแปลงพารามิเตอร์ดังกล่าวให้มีความแปรปรวนเท่ากับ 1 เป็น $\eta_j = \theta_j / \sigma \sim N(0, 1)$ เมื่อ $\theta_j = \sigma \cdot \eta_j$ และกำหนดการแจกแจงก่อนหน้าและการแจกแจงภายหลังในลักษณะเดียวกับงานของ Cho (2007) และ Choi (2014) ดังนี้

การแจกแจงก่อนหน้าของพารามิเตอร์ในโมเดล

- นักเรียน: $\eta_j \sim N(0, 1)$; $\sigma \sim N(0, 1) T(0,)$

- ข้อสอบ: $a_i \sim N(0, 1) T(0,)$; $b_i \sim N(0, 1)$; $c_i \sim \text{Beta}(5, 17)$

ความน่าจะเป็นของการแจกแจงภายหลัง แสดงถึงความสัมพันธ์ (3.2)

$$P(S | y) \propto L(\theta_j) P(\eta_j) P(\sigma) P(a_i) P(b_i) P(c_i) \quad (3.2)$$

เมื่อ $S = \{\eta_j, \sigma, a_i, b_i, c_i\}$ คือเซตของพารามิเตอร์ที่สนใจ; $L(\theta_j)$ คือฟังก์ชันภาวะควรจะเป็นของโมเดล 3PL ซึ่งมีรูปแบบดังสมการ (3.3)

$$L(\theta_j) = \prod_{i=1}^I \prod_{j=1}^J \left[\left\{ P(y_{ij}=1 | \theta_j) \right\}^{y_{ij}} \cdot \left\{ 1 - P(y_{ij}=1 | \theta_j) \right\}^{1-y_{ij}} \right] \quad (3.3)$$

เมื่อ y_{ij} คือผลการตอบข้อสอบของนักเรียน ($y_{ij} = 1$ เมื่อตอบถูก; $y_{ij} = 0$ เมื่อตอบผิด)

การวิเคราะห์ใน MultiBUGS งานวิจัยนี้จะแปลงความสัมพันธ์ของตัวแปรในโมเดลให้เป็นรูปแบบเชิงเส้นด้วยการใช้ฟังก์ชันเชื่อมโยง (link function) คือ logit เนื่องจากจะทำให้การเขียนชุดคำสั่งในโปรแกรมทำได้สะดวกกว่า ทั้งนี้จากสมการ (3.1) ถ้ากำหนด $t_{ij} = 1 / (1 + \exp[-a_i(\theta_j - b_i)])$ อาจเขียนความสัมพันธ์ในรูปแบบเชิงเส้นได้เป็น $\text{logit}(t_{ij}) = a_i(\theta_j - b_i)$ ดังนั้นโมเดล 3PL ในรูปแบบเชิงเส้นจึงเขียนได้เป็น $p_{ij} = c_i + (1 - c_i)t_{ij}$ และแสดงกราฟการประมาณค่าได้ดังภาพ 3.1

ข) โมเดล 2PLE

โมเดล จากสมการ (2.8) ทราบว่าโมเดล 2PLE มีรูปแบบดังสมการ (3.4)

$$P(y_{ij}=1 | \theta_j) = g_{ij} + (1 - g_{ij}) \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]} \quad (3.4)$$

เมื่อ $g_{ij} = (1/m_i)^{1/(1+\exp[a_i(\theta_j - b_i)])}$ และ $\theta_j \sim N(0, \sigma^2)$ จากนั้นแปลงพารามิเตอร์ดังกล่าวให้มีความแปรปรวนเท่ากับ 1 เป็น $\eta_j \sim N(0, 1)$ เมื่อ $\theta_j = \sigma \cdot \eta_j$ และให้ $m_i = 5$ เพื่อให้สอดคล้องกับจำนวนตัวเลือกรวมของข้อสอบมาตรฐานทั่วไป และกำหนดการแจกแจงก่อนหน้าและการแจกแจงภายหลังในลักษณะเดียวกับโมเดล 3PL ดังนี้

การแจกแจงก่อนหน้าของพารามิเตอร์ในโมเดล

- นักเรียน: $\eta_j \sim N(0, 1)$; $\sigma \sim N(0, 1)T(0,)$

- ข้อสอบ: $a_i \sim N(0, 1)T(0,)$; $b_i \sim N(0, 1)$

ความน่าจะเป็นของการแจกแจงภายหลัง แสดงถึงความสัมพันธ์ (3.5)

$$P(S | y) \propto L(\theta_j) P(\eta_j) P(\sigma) P(a_i) P(b_i) \quad (3.5)$$

เมื่อ $S = \{\eta_j, \sigma, a_i, b_i\}$ คือเซตของพารามิเตอร์ที่สนใจ; $L(\theta_j)$ คือฟังก์ชันภาวะควรจะเป็นของโมเดล 2PLE ซึ่งมีรูปแบบเช่นเดียวกับโมเดล 3PL ดังสมการ (3.3)

การวิเคราะห์ใน MultiBUGS จากโมเดล 2PLE ในสมการ (3.4) ถ้ากำหนด $t_{ij} = 1/(1 + \exp[-a_i(\theta_j - b_i)])$ และ $s_{ij} = 1/(1 + \exp[a_i(\theta_j - b_i)])$ อาจเขียนความสัมพันธ์ในรูปแบบเชิงเส้นได้เป็น $\text{logit}(t_{ij}) = a_i(\theta_j - b_i)$; $\text{logit}(s_{ij}) = -a_i(\theta_j - b_i) = -\text{logit}(t_{ij})$ และ $g_{ij} = (1/5)^{s_{ij}} = .20^{s_{ij}}$ ดังนั้นโมเดล 2PLE ในรูปแบบเชิงเส้นจึงเขียนได้เป็น $p_{ij} = g_{ij} + (1 - g_{ij})t_{ij}$ และแสดงกราฟการประมาณค่าได้ดังภาพ 3.2

ค) โมเดล multilevel 3PL

โมเดล จากสมการ (2.23) ทราบว่าโมเดล multilevel 3PL มีรูปแบบดังสมการ (3.6)

$$P(y_{ijs} = 1 | r_{js}) = c_i + (1 - c_i) \frac{1}{1 + \exp[-a_i((r_{js} + u_{0s}) - b_i)]} \quad (3.6)$$

เมื่อ $r_{js} \sim N(0, \tau)$ จากนั้นแปลงพารามิเตอร์ดังกล่าวให้มีความแปรปรวนเท่ากับ 1 เป็น $\eta_{js} = r_{js}/\sqrt{\tau} \sim N(0, 1)$ เมื่อ $r_{js} = \sqrt{\tau} \cdot \eta_{js}$ และกำหนดการแจกแจงก่อนหน้าและการแจกแจงภายหลังในลักษณะเดียวกับงานของ Cho (2007) และ Choi (2014) ดังนี้

การแจกแจงก่อนหน้าของพารามิเตอร์ในโมเดล

- นักเรียน: $\eta_{js} \sim N(0, 1)$; $\sqrt{\tau} \sim N(0, 1)T(0,)$

- โรงเรียน: $u_{0s} \sim N(\gamma_{00}, \zeta)$; $\gamma_{00} \sim N(0, 1)$; $\zeta \sim \text{Gamma}(.1, .001)$

- ข้อสอบ: $a_i \sim N(0, 1)T(0,)$; $b_i \sim N(0, 1)$; $c_i \sim \text{Beta}(5, 17)$

ความน่าจะเป็นของการแจกแจงภายหลัง แสดงถึงความสัมพันธ์ (3.7)

$$P(S | y) \propto L(\theta_{js}) P(\eta_{js}) P(\sqrt{\tau}) P(u_{0s} | \gamma_{00}, \zeta) P(\gamma_{00}) P(\zeta) P(a_i) P(b_i) P(c_i) \quad (3.7)$$

เมื่อ $S = \{\eta_{js}, \sqrt{\tau}, u_{0s}, \gamma_{00}, \zeta, a_i, b_i, c_i\}$ คือเซตของพารามิเตอร์ที่สนใจ; $L(\theta_{js})$ คือฟังก์ชันภาวะควรจะเป็นของโมเดล multilevel 3PL ซึ่งมีรูปแบบดังสมการ (3.8)

$$L(\theta_{js}) = \prod_{i=1}^I \prod_{j=1}^J \left[\{P(y_{ijs} = 1 | \theta_{js})\}^{y_{ij}} \cdot \{1 - P(y_{ijs} = 1 | \theta_{js})\}^{1-y_{ij}} \right] \quad (3.8)$$

เมื่อ y_{ijs} คือผลการตอบข้อสอบของนักเรียนในโรงเรียน ($y_{ijs} = 1$ เมื่อตอบถูก; $y_{ijs} = 0$ เมื่อตอบผิด) การวิเคราะห์ใน MultiBUGS จากโมเดล multilevel 3PL ในสมการ (3.6) ถ้ากำหนด $t_{ij} = 1 / (1 + \exp[-a_i(r_{js} + u_{0s} - b_i)])$ อาจเขียนความสัมพันธ์ในรูปแบบเชิงเส้นได้เป็น $\text{logit}(t_{ij}) = a_i(r_{js} + u_{0s} - b_i)$ ทั้งนี้ดัชนีระดับโรงเรียน (s) จะไม่ถูกนำมาพิจารณาสำหรับ t_{ij} เนื่องจากโดยทั่วไปนักเรียนแต่ละคนจะสังกัดอยู่ในโรงเรียนได้เพียงแห่งเดียว การใช้ดัชนีระดับนักเรียน (j) จึงเพียงพอที่จะอธิบายแทน s ได้ ดังนั้นโมเดล multilevel 3PL ในรูปแบบเชิงเส้นจึงเขียนได้เป็น $p_{ij} = c_i + (1 - c_i)t_{ij}$ และแสดงกราฟการประมาณค่าได้ดังภาพ 3.3

ง) โมเดล multilevel 2PLE

โมเดล จากการบูรณาการระหว่างโมเดล 2PLE ในสมการ (3.4) กับ multilevel 3PL ในสมการ (3.6) งานวิจัยนี้จะกำหนดให้โมเดล multilevel 2PLE มีรูปแบบดังสมการ (3.9)

$$P(y_{ijs} = 1 | r_{js}) = g_{ijs} + (1 - g_{ijs}) \frac{1}{1 + \exp[-a_i((r_{js} + u_{0s}) - b_i)]} \quad (3.9)$$

เมื่อ $g_{ijs} = (1/m_i)^{1/(1 + \exp[a_i(r_{js} + u_{0s} - b_i)])}$ และ $r_{js} \sim N(0, \tau)$ จากนั้นแปลงพารามิเตอร์ดังกล่าวให้มีความแปรปรวนเท่ากับ 1 เป็น $\eta_{js} \sim N(0, 1)$ เมื่อ $r_{js} = \sqrt{\tau} \cdot \eta_{js}$ และให้ $m_i = 5$ และกำหนดการแจกแจงก่อนหน้าและการแจกแจงภายหลังในลักษณะเดียวกับโมเดล multilevel 3PL ดังนี้

การแจกแจงก่อนหน้าของพารามิเตอร์ในโมเดล

- นักเรียน: $\eta_{js} \sim N(0, 1)$; $\sqrt{\tau} \sim N(0, 1)T(0,)$
- โรงเรียน: $u_{0s} \sim N(\gamma_{00}, \zeta)$; $\gamma_{00} \sim N(0, 1)$; $\zeta \sim \text{Gamma}(1, .001)$
- ข้อสอบ: $a_i \sim N(0, 1)T(0,)$; $b_i \sim N(0, 1)$

ความน่าจะเป็นของการแจกแจงภายหลัง แสดงถึงความสัมพันธ์ (3.10)

$$P(S | y) \propto L(\theta_{js}) P(\eta_j) P(\sqrt{\tau}) P(u_{0s} | \gamma_{00}, \zeta) P(\gamma_{00}) P(\zeta) P(a_i) P(b_i) \quad (3.10)$$

เมื่อ $S = \{\eta_{js}, \sqrt{\tau}, u_{0s}, \gamma_{00}, \zeta, a_i, b_i\}$ คือเซตของพารามิเตอร์ที่สนใจ; $L(\theta_{js})$ คือฟังก์ชันภาวะควรจะเป็นของโมเดล multilevel 2PLE ซึ่งมีรูปแบบเช่นเดียวกับโมเดล multilevel 3PL ดังสมการ (3.8)

การวิเคราะห์ใน MultiBUGS จากโมเดล multilevel 2PLE ในสมการ (3.9) ถ้ากำหนด $t_{ij} = 1 / (1 + \exp[-a_i(r_{js} + u_{0s} - b_i)])$ และ $s_{ij} = 1 / (1 + \exp[a_i(r_{js} + u_{0s} - b_i)])$ อาจเขียนความสัมพันธ์ในรูปแบบเชิงเส้นได้เป็น $\text{logit}(t_{ij}) = a_i(r_{js} + u_{0s} - b_i)$; $\text{logit}(s_{ij}) = -\text{logit}(t_{ij})$ และ $g_{ij} = .20^{s_{ij}}$ ดังนั้นโมเดล multilevel 2PLE ในรูปแบบเชิงเส้นจึงเขียนได้เป็น $p_{ij} = g_{ij} + (1 - g_{ij})t_{ij}$ และแสดงกราฟการประมาณค่าได้ดังภาพ 3.4

จ) โมเดล Mix3PL

โมเดล จากสมการ (2.17) ทราบว่าโมเดล Mix3PL มีรูปแบบดังสมการ (3.11)

$$P(y_{ij} = 1 | \theta_{jk}) = \sum_{k=1}^K \pi_k \left(c_{ik} + (1 - c_{ik}) \frac{1}{1 + \exp[-a_{ik}(\theta_{jk} - b_{ik})]} \right) \quad (3.11)$$

เมื่อ $\theta_{jk} \sim N(\mu_k, \sigma_k^2)$ จากนั้นแปลงพารามิเตอร์ดังกล่าวให้มีความแปรปรวนเท่ากับ 1 เป็น $\eta_{jk} = \theta_{jk} / \sigma_k \sim N(\mu_k, 1)$ เมื่อ $\theta_{jk} = \sigma_k \cdot \eta_{jk}$ และกำหนดการแจกแจงก่อนหน้าและการแจกแจงภายหลังในลักษณะเดียวกับงานของ Cho (2007) และ Choi (2014) ดังนี้

การแจกแจงก่อนหน้าของพารามิเตอร์ในโมเดล

- นักเรียน: $\eta_{jk} \sim N(\mu_k, 1)$; $\mu_k \sim N(0, 1)$; $\mu_1 = 0$; $\sigma_k \sim N(0, 1) T(0, 1)$
- ข้อสอบ: $a_{ik} \sim N(0, 1) T(0, 1)$; $b_{ik} \sim N(0, 1)$; $c_{ik} \sim \text{Beta}(5, 17)$
- กลุ่มแฝง: $k \sim \text{Mult}(1, \pi_k)$; $(\pi_1, \dots, \pi_K) \sim \text{Dir}(1, \dots, 1)$

ความน่าจะเป็นของการแจกแจงภายหลัง แสดงถึงความสัมพันธ์ (3.12)

$$P(S | y) \propto L(k, \theta_{jk}) P(\eta_{jk} | \mu_k) P(\mu_k) P(\sigma_k) P(a_{ik}) P(b_{ik}) P(c_{ik}) P(k | \pi_k) P(\pi_k) \quad (3.12)$$

เมื่อ $S = \{\eta_{jk}, \mu_k, \sigma_k, a_{ik}, b_{ik}, c_{ik}, k, \pi_k\}$ คือเซตของพารามิเตอร์ที่สนใจ; $L(k, \theta_{jk})$ คือฟังก์ชันภาวะควรจะเป็นของโมเดล Mix3PL ซึ่งมีรูปแบบดังสมการ (3.13)

$$L(k, \theta_{jk}) = \prod_{i=1}^I \prod_{j=1}^J \left[\left\{ \sum_{k=1}^K \pi_k \cdot P(y_{ij} = 1 | k, \theta_{jk}) \right\}^{y_{ij}} \cdot \left\{ 1 - \sum_{k=1}^K \pi_k \cdot P(y_{ij} = 1 | k, \theta_{jk}) \right\}^{1 - y_{ij}} \right]^{c_{jk}} \quad (3.13)$$

เมื่อ y_{ij} คือผลการตอบข้อสอบของนักเรียน ($y_{ij} = 1$ เมื่อตอบถูก; $y_{ij} = 0$ เมื่อตอบผิด) และ $c_{jk} = 1$ เมื่อนักเรียนอยู่ในกลุ่มแฝง k ; $c_{jk} = 0$ สำหรับกรณีอื่น ๆ

การวิเคราะห์ใน MultiBUGS จากโมเดล Mix3PL ในสมการ (3.11) ถ้ากำหนด $t_{ij} = 1 / (1 + \exp[-a_{ik}(\theta_{jk} - b_{ik})])$ อาจเขียนความสัมพันธ์ในรูปแบบเชิงเส้นได้เป็น $\text{logit}(t_{ij}) = a_{ik}(\theta_{jk} - b_{ik})$ ทั้งนี้ดัชนีกลุ่มแฝง (k) จะไม่ถูกนำมาพิจารณาสำหรับ t_{ij} เนื่องจากนักเรียนแต่ละคนจะถูกจัดอยู่ในกลุ่มแฝงได้เพียงกลุ่มเดียว การใช้ดัชนีระดับนักเรียน (j) จึงเพียงพอที่จะอธิบายแทน k ได้ ดังนั้นโมเดล Mix3PL ในรูปแบบเชิงเส้นจึงเขียนได้เป็น $p_{ij} = \sum_{k=1}^K \pi_k (c_{ik} + (1 - c_{ik}) t_{ij})$ และแสดงกราฟการประมาณค่าได้ดังภาพ 3.5

ฉ) โมเดล Mix2PLE

โมเดล จากการบูรณาการระหว่างโมเดล 2PLE ในสมการ (3.4) กับ Mix3PL ในสมการ (3.11) งานวิจัยนี้จะกำหนดให้โมเดล Mix2PLE มีรูปแบบดังสมการ (3.14)

$$P(y_{ij} = 1 | \theta_{jk}) = \sum_{k=1}^K \pi_k \left(g_{ijk} + (1 - g_{ijk}) \frac{1}{1 + \exp[-a_{ik}(\theta_{jk} - b_{ik})]} \right) \quad (3.14)$$

เมื่อ $g_{ijk} = (1/m_i)^{1/(1+\exp[a_{ik}(\theta_{jk}-b_{ik})])}$ และ $\theta_{jk} \sim N(\mu_k, \sigma_k^2)$ จากนั้นแปลงพารามิเตอร์ดังกล่าวให้มีความแปรปรวนเท่ากับ 1 เป็น $\eta_{jk} \sim N(\mu_k, 1)$ เมื่อ $\theta_{jk} = \sigma_k \cdot \eta_{jk}$ และให้ $m_i = 5$ และกำหนดการแจกแจงก่อนหน้าและการแจกแจงภายหลังในลักษณะเดียวกับโมเดล Mix3PL ดังนี้

การแจกแจงก่อนหน้าของพารามิเตอร์ในโมเดล

- นักเรียน: $\eta_{jk} \sim N(\mu_k, 1)$; $\mu_k \sim N(0, 1)$; $\mu_1 = 0$; $\sigma_k \sim N(0, 1) T(0,)$

- ข้อสอบ: $a_{ik} \sim N(0, 1) T(0,)$; $b_{ik} \sim N(0, 1)$

- กลุ่มแฝง: $k \sim Mult(1, \pi_k)$; $(\pi_1, \dots, \pi_K) \sim Dir(1, \dots, 1)$

ความน่าจะเป็นของการแจกแจงภายหลัง แสดงถึงความสัมพันธ์ (3.15)

$$P(S | y) \propto L(k, \theta_{jk}) P(\eta_{jk} | \mu_k) P(\mu_k) P(\sigma_k) P(a_{ik}) P(b_{ik}) P(k | \pi_k) P(\pi_k) \quad (3.15)$$

เมื่อ $S = \{\eta_{jk}, \mu_k, \sigma_k, a_{ik}, b_{ik}, k, \pi_k\}$ คือเซตของพารามิเตอร์ที่สนใจ; $L(k, \theta_{jk})$ คือฟังก์ชันภาวะควรจะเป็นของโมเดล Mix2PLE ซึ่งมีรูปแบบเช่นเดียวกับโมเดล Mix3PL ดังสมการ (3.13)

การวิเคราะห์ใน MultiBUGS จากโมเดล Mix2PLE ในสมการ (3.14) ถ้ากำหนด $t_{ij} = 1/(1+\exp[-a_{ik}(\theta_{jk}-b_{ik})])$ และ $s_{ij} = 1/(1+\exp[a_{ik}(\theta_{jk}-b_{ik})])$ อาจเขียนความสัมพันธ์ในรูปแบบเชิงเส้นได้เป็น $\text{logit}(t_{ij}) = a_{ik}(\theta_{jk}-b_{ik})$; $\text{logit}(s_{ij}) = -\text{logit}(t_{ij})$ และ $g_{ij} = .20^{s_{ij}}$ ดังนั้นโมเดล Mix2PLE ในรูปแบบเชิงเส้นจึงเขียนได้เป็น $p_{ij} = \sum_{k=1}^K \pi_k (g_{ij} + (1-g_{ij}) t_{ij})$ และแสดงกราฟการประมาณค่าได้ดังภาพ 3.6

ข) โมเดล MMix3PL

โมเดล จากสมการ (2.32) ทราบว่าโมเดล MMix3PL มีรูปแบบดังสมการ (3.16)

$$P(y_{ijs} = 1 | \theta_{jskg}) = \sum_{g=1}^G \sum_{k=1}^K \pi_g \cdot \pi_{k|g} \left(c_{ikg} + (1 - c_{ikg}) \frac{1}{1 + \exp[-a_{ikg}(\theta_{jskg} - b_{ikg})]} \right) \quad (3.16)$$

เมื่อ $\theta_{jskg} \sim N(\mu_{kg}, \sigma_{kg}^2)$ จากนั้นแปลงพารามิเตอร์ดังกล่าวให้มีความแปรปรวนเท่ากับ 1 เป็น $\eta_{jskg} = \theta_{jskg} / \sigma_{kg} \sim N(\mu_{kg}, 1)$ เมื่อ $\theta_{jskg} = \sigma_{kg} \cdot \eta_{jskg}$ และกำหนดการแจกแจงก่อนหน้าและการแจกแจงภายหลังในลักษณะเดียวกับงานของ Cho (2007) และ Choi (2014) ดังนี้

การแจกแจงก่อนหน้าของพารามิเตอร์ในโมเดล

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

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

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

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

ความน่าจะเป็นของการแจกแจงภายหลัง แสดงถึงความสัมพันธ์ (3.17)

$$P(S | y) \propto L(k, g, \theta_{jskg}) P(\eta_{jskg} | \mu_{kg}) P(\mu_{kg}) P(\sigma_{kg}) P(a_{ikg}) P(b_{ikg}) P(c_{ikg}) P(g | \pi_g) P(\pi_g) P(k | \pi_{k|g}) P(\pi_{k|g}) \quad (3.17)$$

เมื่อ $S = \{\eta_{jsg}, \mu_{kg}, \sigma_{kg}, a_{ikg}, b_{ikg}, c_{ikg}, g, k, \pi_g, \pi_{k|g}\}$ คือเซตของพารามิเตอร์ที่สนใจ ;
 $L(k, g, \theta_{jsg})$ คือฟังก์ชันภาวะควรจะเป็นของโมเดล MMix3PL ซึ่งมีรูปแบบดังสมการ (3.18)

$$L(k, g, \theta_{jsg}) = \prod_{i=1}^I \prod_{j=1}^J \left[\left\{ \sum_{g=1}^G \sum_{k=1}^K \pi_g \pi_{k|g} P(y_{ijs} = 1 | k, g, \theta_{jsg}) \right\}^{y_{ij}} \cdot \left\{ 1 - \sum_{g=1}^G \sum_{k=1}^K \pi_g \pi_{k|g} P(y_{ijs} = 1 | k, g, \theta_{jsg}) \right\}^{1-y_{ij}} \right]^{c_{jsg}} \quad (3.18)$$

เมื่อ y_{ijs} คือผลการตอบข้อสอบของนักเรียนในโรงเรียน ($y_{ijs} = 1$ เมื่อตอบถูก; $y_{ijs} = 0$ เมื่อตอบผิด) และ $c_{jsg} = 1$ เมื่อนักเรียนอยู่ในกลุ่มแฟงระดับนักเรียน k และกลุ่มแฟงระดับโรงเรียน g และ $c_{jsg} = 0$ สำหรับกรณีอื่น ๆ

การวิเคราะห์ใน MultiBUGS จากโมเดล MMix3PL ในสมการ (3.16) ถ้ากำหนด $t_{ij} = 1/(1 + \exp[-a_{ikg}(\theta_{jsg} - b_{ikg})])$ อาจเขียนความสัมพันธ์ในรูปแบบเชิงเส้นได้เป็น $\text{logit}(t_{ij}) = a_{ikg}(\theta_{jsg} - b_{ikg})$ ทั้งนี้ดัชนีระดับโรงเรียน (s) กลุ่มแฟงระดับนักเรียน (k) และกลุ่มแฟงระดับโรงเรียน (g) จะไม่ถูกนำมาพิจารณาสำหรับ t_{ij} เนื่องจากโดยทั่วไปนักเรียนแต่ละคนจะสังกัดอยู่ในโรงเรียนได้เพียงแห่งเดียว รวมถึงจะอยู่ในกลุ่มแฟงระดับนักเรียนหรือระดับโรงเรียนได้เพียงกลุ่มเดียวเช่นกัน การใช้ดัชนีระดับนักเรียน (j) จึงเพียงพอที่จะอธิบายแทนดัชนีดังกล่าวได้ ดังนั้นโมเดล MMix3PL รูปแบบเชิงเส้นจึงเขียนได้เป็น $p_{ij} = \sum_{g=1}^G \sum_{k=1}^K \pi_g \pi_{k|g} (c_{ikg} + (1 - c_{ikg}) t_{ij})$ และแสดงกราฟการประมาณค่าได้ดังภาพ 3.7

ข) โมเดล MMix2PLE

โมเดล จากการบูรณาการระหว่างโมเดล 2PLE ในสมการ (3.4) กับ MMix3PL ในสมการ (3.16) งานวิจัยนี้จะกำหนดให้โมเดล MMix2PLE มีรูปแบบดังสมการ (3.19)

$$P(y_{ijs} = 1 | \theta_{jsg}) = \sum_{g=1}^G \sum_{k=1}^K \pi_g \cdot \pi_{k|g} \left(g_{ijskg} + (1 - g_{ijskg}) \frac{1}{1 + \exp[-a_{ikg}(\theta_{jsg} - b_{ikg})]} \right) \quad (3.19)$$

เมื่อ $g_{ijskg} = (1/m_i)^{1/(1 + \exp[a_{ikg}(\theta_{jsg} - b_{ikg})])}$ และ $\theta_{jsg} \sim N(\mu_{kg}, \sigma_{kg}^2)$ จากนั้นแปลงพารามิเตอร์ดังกล่าวให้มีความแปรปรวนเท่ากับ 1 เป็น $\eta_{jsg} \sim N(\mu_{kg}, 1)$ เมื่อ $\theta_{jsg} = \sigma_{kg} \cdot \eta_{jsg}$ และให้ $m_i = 5$ และกำหนดการแจกแจงก่อนหน้าและการแจกแจงภายหลังในลักษณะเดียวกับโมเดล MMix3PL ดังนี้

การแจกแจงก่อนหน้าของพารามิเตอร์ในโมเดล

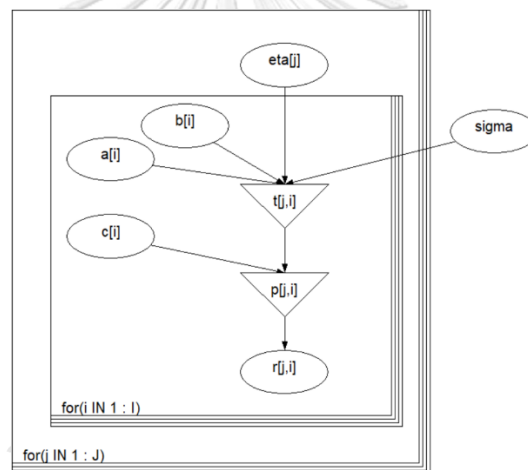
- นักเรียน: $\eta_{jsg} \sim N(\mu_{kg}, 1)$; $\mu_{kg} \sim N(0, 1)$; $\mu_{11} = 0$; $\sigma_{kg} \sim N(0, 1) T(0)$
- ข้อสอบ: $a_{ikg} \sim N(0, 1) T(0)$; $b_{ikg} \sim N(0, 1)$
- กลุ่มแฟง: $g \sim \text{Mult}(1, \pi_g)$; $(\pi_1, \dots, \pi_G) \sim \text{Dir}(1, \dots, 1)$;
 $k \sim \text{Mult}(1, \pi_{k|g})$; $(\pi_{1|1}, \dots, \pi_{K|G}) \sim \text{Dir}(1, \dots, 1)$

ความน่าจะเป็นของการแจกแจงภายหลัง แสดงดังความสัมพันธ์ (3.20)

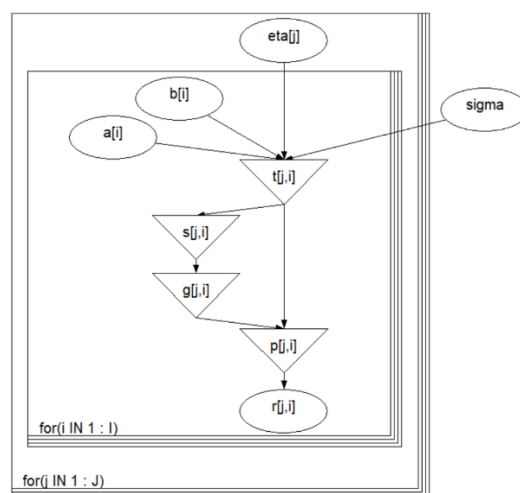
$$P(S | y) \propto L(k, g, \theta_{j skg}) P(\eta_{j skg} | \mu_{kg}) P(\mu_{kg}) P(\sigma_{kg}) P(a_{ikg}) P(b_{ikg}) \\ P(g | \pi_g) P(\pi_g) P(k | \pi_{k|g}) P(\pi_{k|g}) \quad (3.20)$$

เมื่อ $S = \{\eta_{j skg}, \mu_{kg}, \sigma_{kg}, a_{ikg}, b_{ikg}, g, k, \pi_g, \pi_{k|g}\}$ คือเซตของพารามิเตอร์ที่สนใจ; $L(k, g, \theta_{j skg})$ คือฟังก์ชันภาวะควรจะเป็นของโมเดล MMix2PLE ซึ่งมีรูปแบบเช่นเดียวกับโมเดล MMix3PL ดังสมการ (3.18)

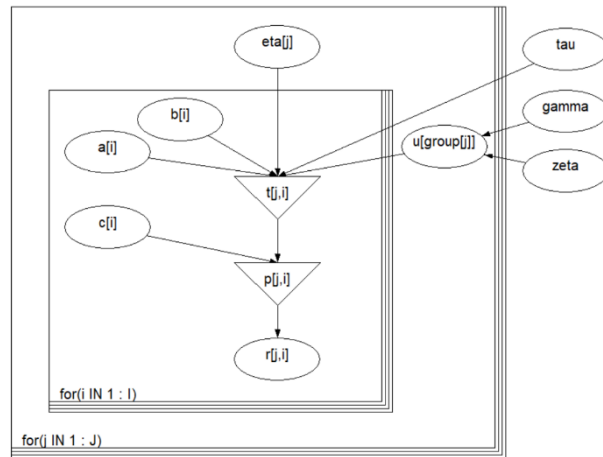
การวิเคราะห์ใน MultiBUGS จากโมเดล MMix2PLE ในสมการ (3.19) ถ้ากำหนด $t_{ij} = 1 / (1 + \exp[-a_{ikg}(\theta_{j skg} - b_{ikg})])$ และ $s_{ij} = 1 / (1 + \exp[a_{ikg}(\theta_{j skg} - b_{ikg})])$ อาจเขียนความสัมพันธ์ในรูปแบบเชิงเส้นได้เป็น $\text{logit}(t_{ij}) = a_{ikg}(\theta_{j skg} - b_{ikg})$; $\text{logit}(s_{ij}) = -\text{logit}(t_{ij})$ และ $g_{ij} = .20^{s_{ij}}$ ดังนั้นโมเดล MMix2PLE ในรูปแบบเชิงเส้นจึงเขียนได้เป็น $p_{ij} = \sum_{g=1}^G \sum_{k=1}^K \pi_g \pi_{k|g} (g_{ij} + (1 - g_{ij}) t_{ij})$ และแสดงกราฟิกการประมาณค่าได้ดังภาพ 3.8



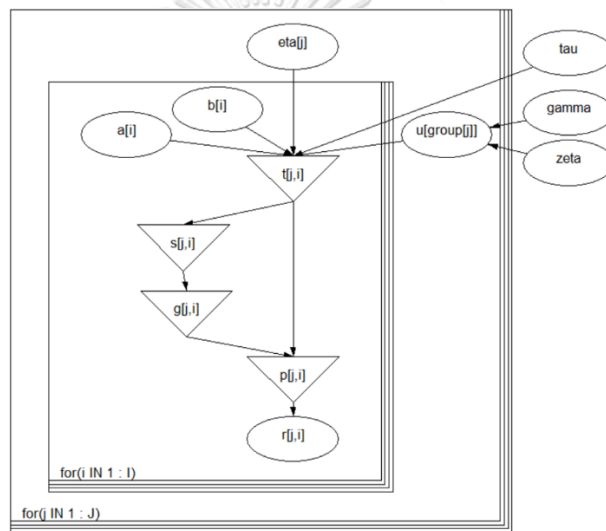
ภาพ 3.1 กราฟิกการประมาณแบบเบส์ค่าสำหรับโมเดล 3PL



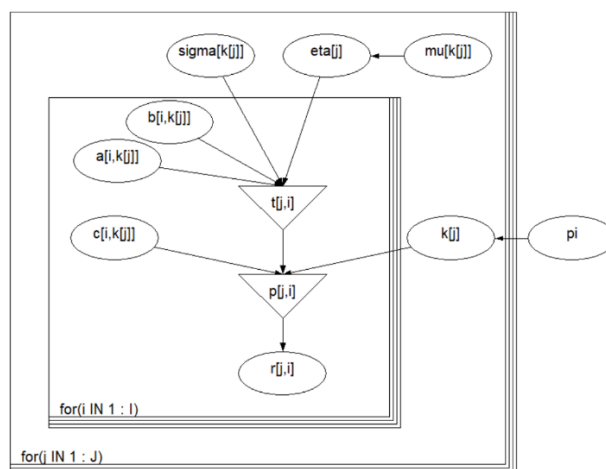
ภาพ 3.2 กราฟิกการประมาณค่าแบบเบส์สำหรับโมเดล 2PLE



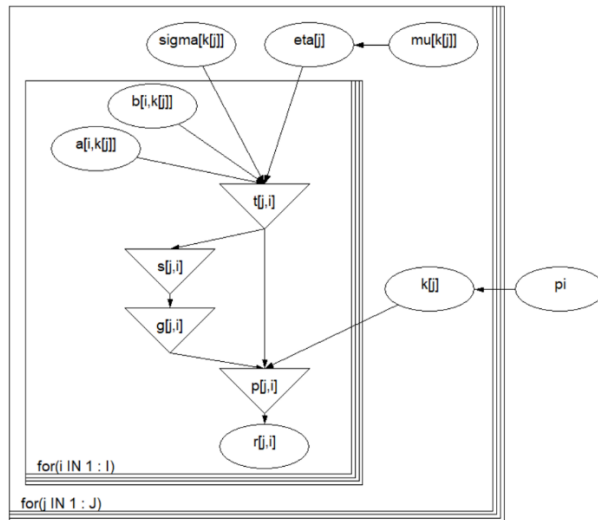
ภาพ 3.3 กราฟิกการประมาณค่าแบบเบส์สำหรับโมเดล multilevel 3PL



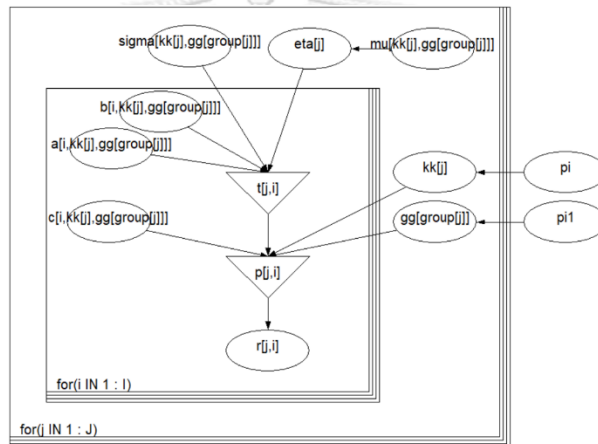
ภาพ 3.4 กราฟิกการประมาณค่าแบบเบส์สำหรับโมเดล multilevel 2PLE



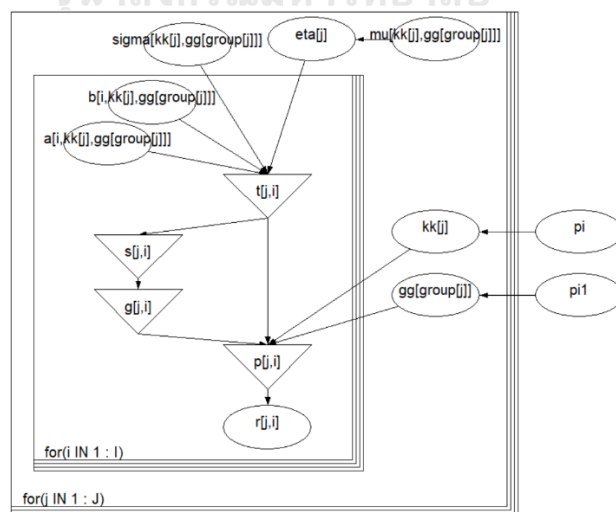
ภาพ 3.5 กราฟิกการประมาณค่าแบบเบส์สำหรับโมเดล Mix3PL



ภาพ 3.6 กราฟการประมาณค่าแบบเบย์สำหรับโมเดล Mix2PLE



ภาพ 3.7 กราฟการประมาณค่าแบบเบย์สำหรับโมเดล MMix3PL



ภาพ 3.8 กราฟการประมาณค่าแบบเบย์สำหรับโมเดล MMix2PLE

หมายเหตุ: \circ คือ stochastic node หมายความว่าตัวแปรถูกสุ่มมาจากการแจกแจงก่อนหน้าที่ทราบค่า; ∇ คือ logical node หมายความว่าตัวแปรมีความสัมพันธ์กับตัวแปรก่อนหน้าแบบระบุรูปแบบได้; \square คือ การดำเนินการทวนซ้ำ

2.2) การตรวจสอบความเหมาะสมของการประมาณค่า

ก) การสลับสัญลักษณ์ (label switching)

ในการประมาณค่าแบบเบส์กับโมเดลกลุ่มแฝง มีความเป็นไปได้ว่าพารามิเตอร์กลุ่มแฝงที่แตกต่างกันอาจมีการแจกแจงภายหลังเหมือนกัน ค่าพารามิเตอร์กลุ่มแฝงที่ประมาณได้จึงอาจมีได้มีเพียงหนึ่งเดียว (unique) เรียกปัญหาดังกล่าวว่าการสลับสัญลักษณ์ (Stephens, 2000) ทั้งนี้การสลับสัญลักษณ์สำหรับโมเดล MMixIRT มี 2 ประเภทหลัก (Cho & Cohen, 2010) ดังนี้

ประเภทที่ 1 เป็นการสลับสัญลักษณ์จากการทวนซ้ำภายในลูกโซ่ (within a chain) ส่งผลให้การประมาณค่าพารามิเตอร์ไม่สามารถระบุค่าได้ (non-identifiable) อาจมีสาเหตุมาจากความไม่เพียงพอของสารสนเทศที่ใช้ในขั้นตอนวิธีเพื่อจำแนกกลุ่มแฝง การสลับสัญลักษณ์ประเภทนี้สามารถตรวจสอบได้จากกราฟความหนาแน่น (density plot) ของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์กลุ่มแฝง หากมีการสลับสัญลักษณ์เกิดขึ้น กราฟจะมีลักษณะเป็นพหุฐานนิยม (multimodal) ซึ่งแสดงถึงการที่ค่าประมาณพารามิเตอร์มีได้มีค่าเพียงหนึ่งเดียว หรือไม่สามารระบุค่าที่แน่นอนได้ การแก้ปัญหาเมื่อมีการสลับสัญลักษณ์ประเภทนี้เกิดขึ้นมี 3 วิธี ได้แก่ 1) กำหนดข้อจำกัดให้กับค่าพารามิเตอร์ 2) ใช้ขั้นตอนวิธีการสลับสัญลักษณ์ใหม่ (relabelling algorithm) และ 3) กำหนดฟังก์ชันสูญเสียที่มีความไม่แปรเปลี่ยน (invariant loss function) (Choi, 2014)

ประเภทที่ 2 เป็นการสลับสัญลักษณ์จากการทวนซ้ำระหว่างลูกโซ่ (across chains) ส่งผลให้ค่าประมาณพารามิเตอร์กลุ่มแฝงระหว่างลูกโซ่มีการสลับสัญลักษณ์กัน ตัวอย่างเช่น เมื่อประมาณค่าด้วยลูกโซ่จำนวน 2 ลูก อาจพบว่าลูกโซ่แรกประมาณค่าพารามิเตอร์กลุ่มแฝงที่ 1 (π_1) และกลุ่มแฝงที่ 2 (π_2) ได้ .3 และ .7 ในขณะที่ลูกโซ่ที่สองประมาณค่า π_1 และ π_2 ได้ .7 และ .3 ซึ่งปัญหานี้อาจมีสาเหตุมาจากการใช้ค่าเริ่มต้นของพารามิเตอร์ในลูกโซ่ที่แตกต่างกัน อย่างไรก็ตามหากเป็นการศึกษาเชิงจำลอง การสลับสัญลักษณ์ประเภทนี้สามารถจัดการได้ด้วยการนำค่าที่ประมาณได้มาเทียบเคียงกับค่าจริงที่จำลองขึ้น แต่ถ้าเป็นการศึกษาเชิงประจักษ์ซึ่งไม่ทราบค่าจริง อาจต้องใช้การสังเกตเปรียบเทียบจากพารามิเตอร์กลุ่มแฝงและความสามารถเฉลี่ยในกลุ่มแฝง

ข) การวินิจฉัยการลู่เข้าของลูกโซ่

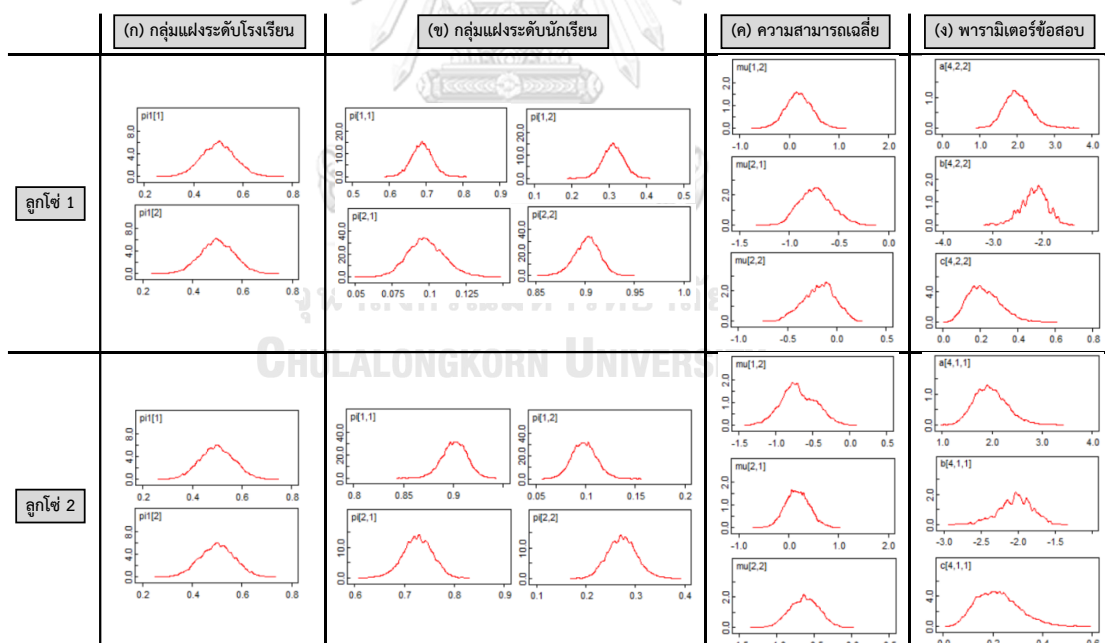
การวินิจฉัยการลู่เข้าของลูกโซ่ในงานวิจัยนี้จะพิจารณาจากวิธีของ Gelman และ Rubin (1992) และอัตสหสัมพันธ์ (autocorrelation) ซึ่งดำเนินการวิเคราะห์ด้วยโปรแกรม MultiBUGS โดยในการประมาณค่าแบบเบส์สำหรับโมเดล IRT จะกำหนดให้มีการตัดค่าสังเกต (burn-in) 1,000 ค่าแรกทิ้ง และนำค่าสังเกตภายหลังการตัด (post-burn-in) 8,000 ค่า มาประมาณหาค่าเฉลี่ย ส่วนโมเดล multilevel IRT และ MixIRT กำหนดให้ burn-in เท่ากับ 4,000 และ post-burn-in เท่ากับ 8,000 และโมเดล MMixIRT กำหนดให้ burn-in เท่ากับ 7,000 และ post-burn-in เท่ากับ 8,000 ซึ่งจำนวน burn-in ดังกล่าวเป็นการกำหนดในลักษณะใกล้เคียงกับงานวิจัยของ Cho (2007)

ค) ผลการตรวจสอบความเหมาะสมของการประมาณค่า

ค.1) การตรวจสอบการสลับสัญลักษณ์

จากการพิจารณากราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์กลุ่มแฝง พบว่ากราฟมีแนวโน้มเป็นเอกฐานนิยม (unimodal) (ดูภาพ 3.9ก-3.9ข) โดยมีลักษณะเป็นเช่นนี้กับทุกเงื่อนไขการศึกษาสำหรับโมเดล MMix3PL และ MMix2PLE ซึ่งผลดังกล่าวแสดงถึงการประมาณค่าพารามิเตอร์ที่สามารถระบุค่าที่แน่นอนได้เพียงหนึ่งเดียว จึงอนุมานได้ว่าไม่มีการสลับสัญลักษณ์ประเภทที่ 1 หรือสารสนเทศที่ใช้ในขั้นตอนวิธีเพื่อประมาณค่าสำหรับโมเดล MMix3PL (ดูภาคผนวก ข.7) และ MMix2PLE (ดูภาคผนวก ข.8) มีความเพียงพอแล้ว

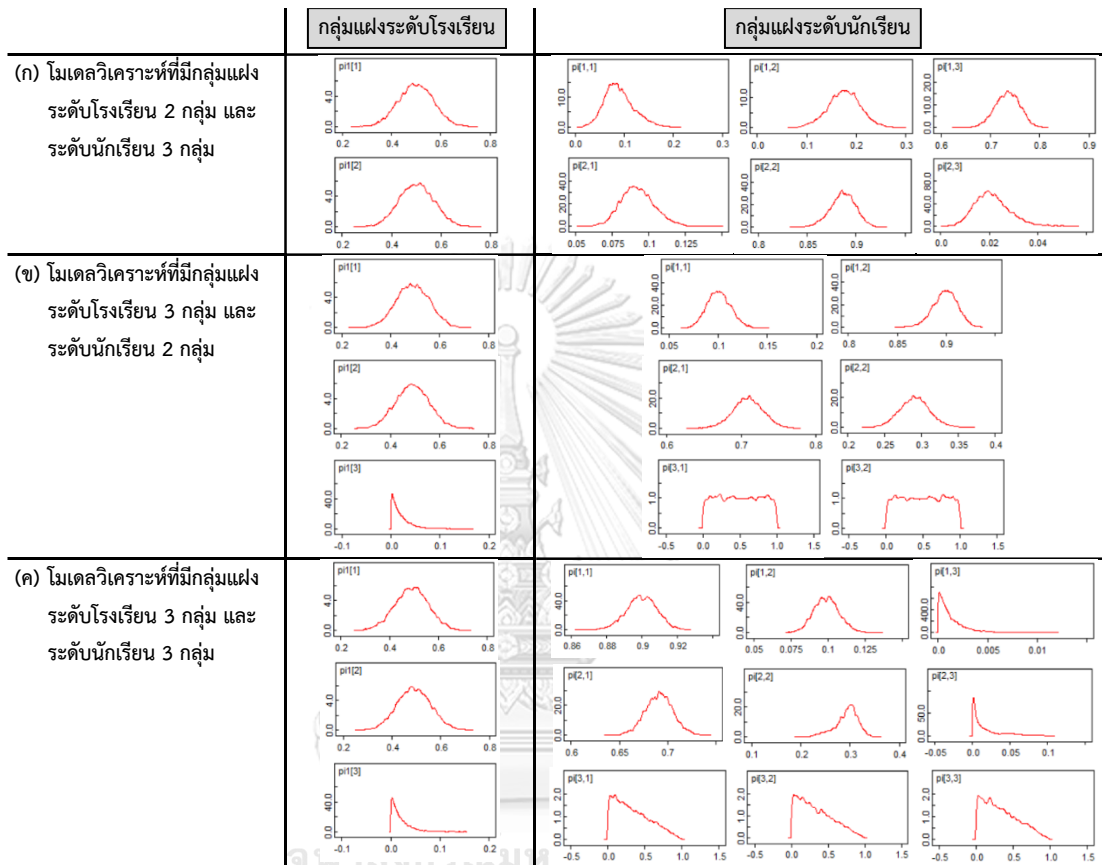
อย่างไรก็ตามบางเงื่อนไขการศึกษาอาจพบว่าค่าพารามิเตอร์กลุ่มแฝงมีการสลับสัญลักษณ์ประเภทที่ 2 เกิดขึ้น สังเกตได้จากค่าประมาณระหว่างลูกโซ่ 1 กับ 2 ที่มีการสลับค่ากัน เช่น ลูกโซ่ 1: $\pi_{22} \approx .90$; ลูกโซ่ 2: $\pi_{11} \approx .90$ (ดูภาพ 3.9ข) และผลดังกล่าวยังทำให้ค่าประมาณพารามิเตอร์ความสามารถเฉลี่ยและพารามิเตอร์ของข้อสอบในกลุ่มแฝงมีการสลับสัญลักษณ์ด้วยเช่นกัน (ดูภาพ 3.9ค-3.9ง) ทั้งนี้ในงานวิจัยจะแก้ปัญหาดังกล่าวด้วยการนำค่าพารามิเตอร์ที่ประมาณได้ไปเทียบเคียงกับค่าจริงที่จำลองขึ้น เพื่อศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์ต่าง ๆ ต่อไป



ภาพ 3.9 กราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์

นอกจากนี้อาจสังเกตได้ว่าเมื่อนำข้อมูลดังกล่าวซึ่งมีกลุ่มแฝงระดับโรงเรียน 2 กลุ่ม และระดับนักเรียน 2 กลุ่ม มาวิเคราะห์ด้วยโมเดลแข่งขันที่มีจำนวนกลุ่มแฝงมากกว่าจำนวนกลุ่มแฝงจริง ค่าประมาณพารามิเตอร์กลุ่มแฝงของกลุ่มที่เกินมาจะมีแนวโน้มเข้าสู่ 0 หรืออนุมานได้ว่ากลุ่มแฝงดังกล่าวอาจไม่มีอยู่จริง (ดูภาพ 3.10) ซึ่งหากกรณีดังกล่าวเกิดขึ้นกับกลุ่มแฝงระดับโรงเรียนก็

อาจส่งผลทำให้พารามิเตอร์กลุ่มแฝงระดับนักเรียนซึ่งซ่อนอยู่ในกลุ่มแฝงระดับโรงเรียนเกิดปัญหาในการประมาณค่า ดังสังเกตได้จากกราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์กลุ่มแฝงระดับนักเรียนในกลุ่มแฝงระดับโรงเรียนที่ 3 ซึ่งมีโซ่เอกรฐานนิยม (ดูภาพ 3.10) หรือกราฟความหนาแน่นฯ ที่จำแนกความต่างระหว่างกลุ่มแฝงระดับนักเรียนไม่ได้ (ดูภาพ 3.10ค)

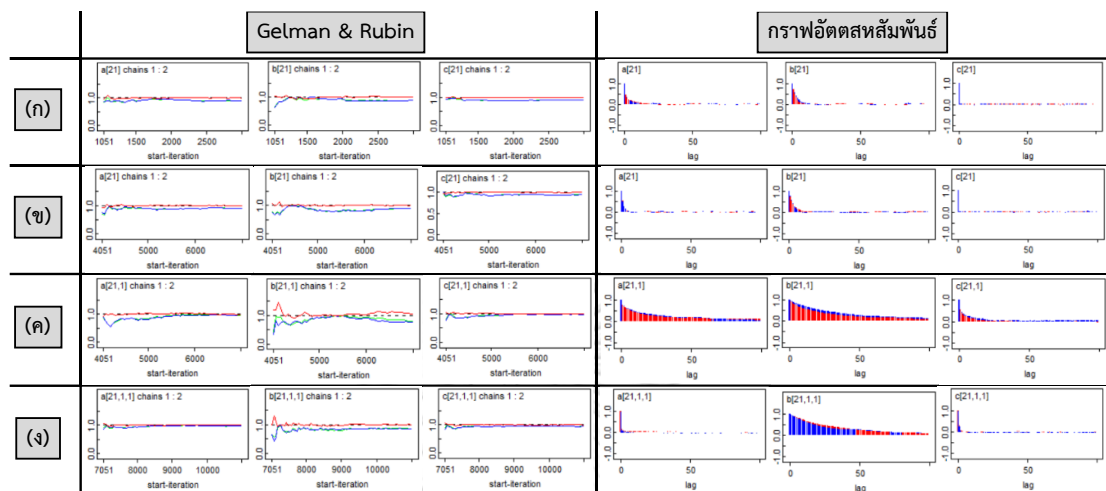


ภาพ 3.10 กราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์กลุ่มแฝงเมื่อโมเดลวิเคราะห์มีจำนวนกลุ่มแฝงมากกว่าข้อมูล

ค.2) การวินิจฉัยการลู่เข้าของลูกโซ่

จากการพิจารณากราฟแสดงค่าสถิติ Gelman & Rubin และอัตราสหสัมพันธ์ พบว่าโมเดล IRT (3PL, 2PLE), multilevel IRT (multilevel 3PL, multilevel 2PLE), MixIRT (Mix3PL, Mix2PLE) และ MMixIRT (MMix3PL, MMix2PLE) มีแนวโน้มการลู่เข้าของลูกโซ่ในการประมาณค่าพารามิเตอร์ค่อนข้างดี กล่าวคือกรณีโมเดล IRT และ multilevel IRT ลูกโซ่มีแนวโน้มลู่เข้าหลังจากตัดค่าสังเกต (burn-in) 1,000 และ 4,000 ค่าแรกทิ้ง ตามลำดับ สังเกตได้จากค่าสถิติ Gelman & Rubin ที่มีแนวโน้มลู่เข้า 1 และระดับอัตราสหสัมพันธ์ที่ลดลงค่อนข้างเร็วเมื่อ lag เพิ่มขึ้น (ดูภาพ 3.11ก-3.11ข) ในทำนองเดียวกันสำหรับกรณีโมเดล MixIRT และ MMixIRT พบว่าลูกโซ่มีแนวโน้มลู่เข้าค่อนข้างดีสำหรับการประมาณค่าพารามิเตอร์อำนาจจำแนกและการเดาของข้อสอบ

หลังจากตัดค่าสังเกต 4,000 และ 7,000 ค่าแรกทิ้ง ตามลำดับ อย่างไรก็ตามสำหรับพารามิเตอร์ความยากของข้อสอบ พบว่าลูกโซ่มีแนวโน้มลู่ออกเข้ายังไม่ดีเท่าที่ควร สังเกตได้จากค่าสถิติ Gelman & Rubin ที่มีค่าผันแปรใกล้ 1 และระดับอัตโนมัติสหสัมพันธ์ลดลงค่อนข้างช้าเมื่อ lag เพิ่มขึ้น (ดูภาพ 3.11ค-3.11ง) โดยมีลักษณะเป็นเช่นนี้กับทุกเงื่อนไขการศึกษา



ภาพ 3.11 การวินิจฉัยการลู่ออกของลูกโซ่: (ก) โมเดล IRT, (ข) โมเดล multilevel IRT, (ค) โมเดล MixIRT และ (ง) โมเดล MMixIRT

ค.3) ผลการประมาณค่าพารามิเตอร์

ผลการประมาณค่าพารามิเตอร์กลุ่มแฝงของข้อมูลจำลอง 10 ชุด และพารามิเตอร์ของข้อสอบจากข้อมูลชุดหนึ่งสำหรับโมเดล MMix3PL และ MMix2PLE กรณี $N = 2,500$ สามารถแสดงได้ดังตาราง 3.5-3.8 และการเปรียบเทียบค่าโอกาสในการเดาถูกจากข้อมูลชุดหนึ่งระหว่างโมเดล MMix3PL กับ MMix2PLE สามารถแสดงได้ดังตาราง 3.9

ตาราง 3.5 ค่าประมาณพารามิเตอร์กลุ่มแฝงสำหรับโมเดล MMix3PL กรณี $N = 2,500$

ชุดข้อมูล	ระดับโรงเรียน				ระดับนักเรียน							
	π_1	π_1	π_2	π_2	π_{111}	π_{111}	π_{211}	π_{211}	π_{112}	π_{112}	π_{212}	π_{212}
1	.50	.50	.50	.50	.90	.90	.10	.10	.30	.24	.70	.76
2	.50	.50	.50	.50	.90	.90	.10	.10	.30	.26	.70	.74
3	.50	.50	.50	.50	.90	.90	.10	.10	.30	.29	.70	.71
4	.50	.50	.50	.50	.90	.90	.10	.10	.30	.28	.70	.72
5	.50	.50	.50	.50	.90	.90	.10	.10	.30	.29	.70	.71
6	.50	.50	.50	.50	.90	.90	.10	.10	.30	.29	.70	.71
7	.50	.50	.50	.50	.90	.90	.10	.10	.30	.29	.70	.71
8	.50	.50	.50	.50	.90	.90	.10	.10	.30	.31	.70	.69
9	.50	.50	.50	.50	.90	.90	.10	.10	.30	.33	.70	.67
10	.50	.50	.50	.50	.90	.90	.10	.10	.30	.30	.70	.70

ตาราง 3.8 ค่าประมาณพารามิเตอร์ของข้อสอบสำหรับโมเดล MMix2PLE กรณี $N = 2,500$

ข้อ	$k = 1, g = 1$				$k = 2, g = 1$				$k = 1, g = 2$				$k = 2, g = 2$			
	b_{11}	\hat{b}_{11}	a_{11}	\hat{a}_{11}	b_{21}	\hat{b}_{21}	a_{21}	\hat{a}_{21}	b_{12}	\hat{b}_{12}	a_{12}	\hat{a}_{12}	b_{22}	\hat{b}_{22}	a_{22}	\hat{a}_{22}
1	-2.50	-2.37	2.00	2.06	0.60	0.86	1.00	0.92	-0.50	-0.52	1.00	1.10	0.60	0.65	1.00	0.91
2	-2.40	-2.28	2.00	2.18	0.70	0.65	1.00	1.12	-0.40	-0.17	1.00	1.15	0.70	0.50	1.00	1.22
3	-2.30	-2.58	2.00	1.57	0.80	0.96	1.00	1.35	-0.30	-0.17	1.00	1.10	0.80	1.02	1.00	0.79
4	-2.20	-2.18	2.00	1.79	0.90	1.25	1.00	0.93	-0.20	-0.25	1.00	1.10	0.90	0.77	1.00	1.17
5	-2.10	-2.02	2.00	2.17	1.00	1.50	1.00	0.84	-0.10	-0.03	1.00	1.16	1.00	0.89	1.00	1.16
6	-2.00	-2.08	2.00	1.87	1.10	1.39	1.00	0.98	0.00	-0.06	1.00	0.97	1.10	0.95	1.00	1.19
7	-1.90	-1.92	2.00	2.09	1.20	1.50	1.00	0.86	0.10	0.17	1.00	0.59	1.20	1.20	1.00	0.94
8	-1.80	-1.71	2.00	2.23	1.30	1.55	1.00	1.18	0.20	0.17	1.00	1.11	1.30	1.33	1.00	0.86
9	-1.70	-1.69	2.00	1.87	1.40	1.45	1.00	0.99	0.30	0.59	1.00	0.89	1.40	1.28	1.00	1.16
10	-1.60	-1.70	2.00	1.94	1.50	1.18	1.00	1.29	0.40	0.48	1.00	1.19	1.50	1.37	1.00	0.92
11	-1.50	-1.45	2.00	1.99	1.60	1.78	1.00	0.86	0.50	0.30	1.00	1.34	1.60	1.44	1.00	1.06
12	-1.40	-1.40	2.00	2.06	1.70	1.89	1.00	1.14	0.60	0.86	1.00	0.79	1.70	1.66	1.00	1.24
13	-1.30	-1.29	2.00	2.25	1.80	1.42	1.00	1.78	0.70	0.71	1.00	1.07	1.80	2.02	1.00	0.81
14	-1.20	-1.22	2.00	2.02	1.90	1.68	1.00	1.59	0.80	0.63	1.00	1.14	1.90	1.78	1.00	1.14
15	-1.10	-1.24	2.00	1.61	2.00	2.21	1.00	1.49	0.90	0.89	1.00	1.18	2.00	2.31	1.00	0.98
16	-1.00	-0.93	2.00	2.08	2.10	2.21	1.00	1.05	1.00	1.10	1.00	0.99	2.10	2.61	1.00	0.83
17	-0.90	-0.94	1.00	0.94	2.20	1.80	1.00	1.47	1.10	1.19	1.00	1.15	2.20	2.67	1.00	0.83
18	-0.80	-0.74	1.00	1.06	2.30	2.19	1.00	1.07	1.20	1.32	1.00	1.01	2.30	2.09	1.00	1.12
19	-0.70	-0.62	1.00	1.00	2.40	1.88	1.00	1.50	1.30	1.68	1.00	0.74	2.40	2.40	1.00	1.26
20	-0.60	-0.77	1.00	0.85	2.50	1.91	1.00	1.45	1.40	1.44	1.00	1.06	2.50	2.06	1.00	1.46
21	0.60	0.67	1.00	0.87	-1.50	-1.95	2.00	1.43	-0.40	-0.53	1.00	1.26	-1.50	-1.61	2.00	1.88
22	0.70	0.78	1.00	0.92	-1.40	-1.22	2.00	1.73	-0.30	-0.07	1.00	1.20	-1.40	-1.44	2.00	2.04
23	0.80	0.95	1.00	0.83	-1.30	-1.24	2.00	1.64	-0.20	-0.07	1.00	1.51	-1.30	-1.27	2.00	1.94
24	0.90	0.90	1.00	1.01	-1.20	-0.95	2.00	1.43	-0.10	-0.27	1.00	1.07	-1.20	-1.20	2.00	1.78
25	1.00	0.97	1.00	1.12	-1.10	-0.93	2.00	1.64	0.00	-0.01	1.00	1.02	-1.10	-1.02	2.00	2.49
26	1.10	1.11	1.00	0.93	-1.00	-0.95	2.00	1.29	0.10	-0.03	1.00	1.66	-1.00	-1.01	2.00	1.67
27	1.20	1.10	1.00	1.21	-0.90	-1.33	1.00	0.59	0.20	0.11	1.00	1.24	-0.90	-0.76	1.00	1.07
28	1.30	1.52	1.00	1.00	-0.80	-0.62	1.00	1.21	0.30	0.42	1.00	0.88	-0.80	-0.78	1.00	1.26
29	1.40	1.51	1.00	1.09	-0.70	-0.81	1.00	1.26	0.40	0.50	1.00	0.87	-0.70	-0.49	1.00	1.14
30	1.50	1.40	1.00	1.07	-0.60	-0.35	1.00	1.25	0.50	0.22	1.00	1.03	-0.60	-0.58	1.00	1.11
31	1.60	1.54	1.00	1.11	-2.50	-2.17	2.00	1.45	0.60	0.46	1.00	0.97	1.50	1.44	1.00	0.98
32	1.70	1.52	1.00	1.20	-2.40	-2.16	2.00	1.59	0.70	0.61	1.00	1.22	1.60	1.36	1.00	1.33
33	1.80	1.97	1.00	0.98	-2.30	-2.54	2.00	1.59	0.80	0.79	1.00	0.83	1.70	1.49	1.00	1.25
34	1.90	1.75	1.00	1.48	-2.20	-2.50	2.00	1.60	0.90	1.02	1.00	0.94	1.80	1.88	1.00	0.95
35	2.00	2.28	1.00	1.00	-2.10	-1.75	2.00	1.44	1.00	0.77	1.00	1.01	1.90	1.92	1.00	1.15
36	2.10	2.14	1.00	1.01	-2.00	-1.83	2.00	1.60	1.10	0.77	1.00	1.49	-2.00	-2.16	2.00	1.84
37	2.20	2.49	1.00	0.84	-1.90	-2.15	2.00	1.60	1.20	0.98	1.00	1.47	-1.90	-1.82	2.00	1.84
38	2.30	2.20	1.00	0.98	-1.80	-2.03	2.00	1.08	1.30	1.23	1.00	1.16	-1.80	-1.72	2.00	1.60
39	2.40	2.19	1.00	1.04	-1.70	-1.86	2.00	1.28	1.40	1.28	1.00	1.35	-1.70	-1.74	2.00	1.76
40	2.50	2.13	1.00	1.41	-1.60	-1.93	2.00	1.50	1.50	1.48	1.00	1.34	-1.60	-1.48	2.00	2.11

ตาราง 3.9 การเปรียบเทียบค่าโอกาสในการเดาถูกระหว่างโมเดล MMix3PL กับ MMix2PLE

กลุ่มแฝง	c ในโมเดล MMix3PL				g ในโมเดล MMix2PLE			
	ค่าเฉลี่ย	ค่าต่ำสุด	ค่าสูงสุด	cor (c, c)	ค่าเฉลี่ย	ค่าต่ำสุด	ค่าสูงสุด	cor (g, g)
$k = 1, g = 1$.213	.162	.275	.159	.553	.200	.999	.799
$k = 2, g = 1$.214	.162	.277	-.603	.502	.200	.998	.711
$k = 1, g = 2$.235	.192	.313	-.284	.395	.201	.973	.241
$k = 2, g = 2$.218	.174	.280	-.089	.468	.200	.999	.773
ภาพรวม	.220	.162	.313	-.098	.497	.200	.999	.767

ง) การปรับเทียบค่าพารามิเตอร์

ค่าพารามิเตอร์ของข้อสอบที่ประมาณได้จำเป็นจะต้องถูกปรับเทียบให้อยู่บนสเกลเดียวกับค่าพารามิเตอร์จริงเสียก่อน เพื่อให้ค่าดังกล่าวสามารถนำมาเปรียบเทียบกันได้ โดยงานวิจัยนี้จะเลือกใช้การปรับเทียบด้วยวิธีค่าเฉลี่ยและซิกมา (mean and sigma equating) (Choi, 2014) ซึ่งมีสูตรคำนวณดังสมการ (3.21) - (3.23)

$$b_i^* = \frac{S_{b_B}}{S_{b_T}} b_i + \left(\bar{b}_B - \frac{S_{b_B}}{S_{b_T}} \bar{b}_T \right) \quad (3.21)$$

$$a_i^* = \frac{S_{b_T}}{S_{b_B}} a_i \quad (3.22)$$

$$c_i^* = c_i \quad (3.23)$$

เมื่อ a_i^* , b_i^* และ c_i^* คือค่าประมาณพารามิเตอร์ของข้อสอบที่ถูกปรับเทียบ; a_i , b_i และ c_i คือค่าประมาณพารามิเตอร์ของข้อสอบ; \bar{b}_B กับ S_{b_B} คือค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของข้อสอบจริง เรียกว่าสเกลฐาน (base scale); \bar{b}_T กับ S_{b_T} คือค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าประมาณพารามิเตอร์ความยากของข้อสอบ เรียกว่าสเกลเป้าหมาย (target scale) ทั้งนี้หลังจากปรับเทียบแล้วจะนำค่าดังกล่าวไปคำนวณ bias RMSE และ corr ต่อไป

3) เกณฑ์การพิจารณาความถูกต้องและความแม่นยำ

3.1) ความถูกต้องของการคัดเลือกโมเดล พิจารณาจากค่าร้อยละของจำนวนครั้งที่เกณฑ์สารสนเทศ AIC, BIC, CAIC และ DBIC เลือกโมเดลได้ตรงกับโมเดลประชากรจากข้อมูลจำลอง 10 ชุด โดยเกณฑ์สารสนเทศดังกล่าวมีสูตรคำนวณดังสมการ (3.24) - (3.27)

$$AIC = -2 \log L + 2p \quad (3.24)$$

$$BIC = -2 \log L + p \cdot \ln(N) \quad (3.25)$$

$$CAIC = -2 \log L + p \cdot [\ln(N) + 1] \quad (3.26)$$

$$DBIC = -2 \log L + p \cdot [\ln(N) - \ln(2\pi)] \quad (3.27)$$

เมื่อ p คือจำนวนพารามิเตอร์; N คือขนาดตัวอย่างของข้อมูล; $\log L$ คือค่า log-likelihood ซึ่งคำนวณได้จากค่าเฉลี่ยภายหลังของความเบี่ยงเบน (deviance) จากขั้นตอนวิธี MCMC ในการประมาณค่าแบบเบย์หรือ $\log L = \bar{D}$ (Congdon, 2003)

ทั้งนี้หากโมเดลใดมีค่าเกณฑ์สารสนเทศต่ำที่สุด จะหมายความว่าโมเดลนั้นถูกเลือกโดยเกณฑ์สารสนเทศให้มีความสอดคล้องกับข้อมูลมากที่สุด และเกณฑ์สารสนเทศใดมีร้อยละของจำนวนครั้งที่เลือกโมเดลได้ตรงกับโมเดลประชากรมากกว่า จะหมายความว่าเกณฑ์สารสนเทศนั้นมีความถูกต้องของการคัดเลือกโมเดลมากกว่า

3.2) ความแม่นยำของการประมาณค่าพารามิเตอร์

- พารามิเตอร์กลุ่มแฝง คำนวณจากค่าร้อยละของความใกล้เคียง (p) ระหว่างค่าประมาณกับค่าจริงของสัดส่วนสมาชิกในกลุ่มแฝงจากข้อมูลแต่ละชุด มีสูตรการคำนวณสำหรับกลุ่มแฝงระดับโรงเรียนและระดับนักเรียนดังสมการ (3.28) และ (3.29) ตามลำดับ ทั้งนี้หาก p เข้าใกล้ 100 จะหมายความว่า การประมาณค่ามีความแม่นยำสูง

$$P_{school} = 100 - \left(\frac{\sum_{r=1}^R \sum_{g=1}^G |\hat{\pi}_{g,r} - \pi_{g,r}|}{\sum_{r=1}^R \sum_{g=1}^G \pi_{g,r}} \times 100 \right) \quad (3.28)$$

$$P_{student} = 100 - \left(\frac{\sum_{r=1}^R \sum_{g=1}^G \sum_{k=1}^K |\hat{\pi}_{k|g,r} - \pi_{k|g,r}|}{\sum_{r=1}^R \sum_{g=1}^G \sum_{k=1}^K \pi_{k|g,r}} \times 100 \right) \quad (3.29)$$

เมื่อ $\pi_{g,r}$ และ $\hat{\pi}_{g,r}$ คือค่าจริงและค่าประมาณของพารามิเตอร์กลุ่มแฝงระดับโรงเรียน g จากข้อมูลชุดที่ r ; $\pi_{k|g,r}$ และ $\hat{\pi}_{k|g,r}$ คือค่าจริงและค่าประมาณของพารามิเตอร์กลุ่มแฝงระดับนักเรียน k ในกลุ่มแฝงระดับโรงเรียน g จากข้อมูลจำลองชุดที่ r

- พารามิเตอร์ของข้อสอบและความสามารถเฉลี่ย คำนวณจาก bias และ RMSE ซึ่งบอกถึงความคลาดเคลื่อนและรากที่สองของความคลาดเคลื่อนกำลังสองระหว่างค่าประมาณกับค่าจริงของพารามิเตอร์ในกลุ่มแฝงจากข้อมูลแต่ละชุด และค่า corr ซึ่งบอกถึงความสัมพันธ์ระหว่างค่าประมาณกับค่าจริง มีสูตรคำนวณดังสมการ (3.30) - (3.32) ทั้งนี้หาก bias และ RMSE เข้าใกล้ 0 และ corr เข้าใกล้ 1 จะหมายความว่า การประมาณค่ามีความแม่นยำสูง

$$\text{bias} = \frac{\sum_{r=1}^R \sum_{g=1}^G \sum_{k=1}^K \sum_{i=1}^I (\hat{\Theta}_{ikgr} - \Theta_{ikg})}{RGKI} \quad (3.30)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{r=1}^R \sum_{g=1}^G \sum_{k=1}^K \sum_{i=1}^I (\hat{\Theta}_{ikgr} - \Theta_{ikg})^2}{RGKI}} \quad (3.31)$$

$$\text{corr} = \frac{1}{R} \sum_{r=1}^R \frac{\text{Cov}(\hat{\Theta}_{ikgr}, \Theta_{ikg})}{\sigma_{\hat{\Theta}_{ikgr}} \sigma_{\Theta_{ikg}}} \quad (3.32)$$

เมื่อ Θ_{ikg} และ $\hat{\Theta}_{ikgr}$ คือค่าจริงและค่าประมาณพารามิเตอร์ของข้อสอบ i ในกลุ่มแฝงระดับนักเรียน k และระดับโรงเรียน g จากข้อมูลจำลองชุดที่ r ส่วน bias และ RMSE สำหรับการประมาณค่าความสามารถเฉลี่ยสามารถคำนวณได้จากสูตรเดียวกัน เพียงแต่จะไม่พิจารณาดัชนีของข้อสอบ (i)

4) สถิติที่ใช้ในการวิเคราะห์

สถิติที่ใช้มี 2 ประเภท ดังนี้ 1) สถิติเชิงบรรยาย ได้แก่ ค่าร้อยละ, bias, RMSE และ corr และ 2) สถิติเชิงอนุมาน ได้แก่ การประมาณค่าพารามิเตอร์แบบเบสด้วยโมเดล MMix3PL และ MMix2PLE

ส่วนที่ 2 การศึกษาจากข้อมูลเชิงประจักษ์

ในส่วนนี้เป็นการสาธิตการใช้โมเดลกับข้อมูลเชิงประจักษ์ โดยจะนำข้อมูลคะแนนสอบ O-NET วิชาคณิตศาสตร์ ม.6 มาวิเคราะห์ด้วยโมเดล MMix3PL และ MMix2PLE เพื่อหาจำนวนกลุ่มแฝงของข้อมูล พารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง โดยผู้วิจัยขอแบ่งการพิจารณาเป็น 5 ด้าน ได้แก่ 1) ประชากรและตัวอย่าง 2) ตัวแปรที่ศึกษา 3) เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล 4) ขั้นตอนการศึกษา และ 5) สถิติที่ใช้ในการวิเคราะห์ รายละเอียดดังนี้

1) ประชากรและตัวอย่าง

ประชากรคือนักเรียนระดับ ม.6 ที่เข้าร่วมการสอบ O-NET วิชาคณิตศาสตร์ ปีการศึกษา 2562 จำนวน 363,752 คน จากสังกัดต่าง ๆ ได้แก่ 1) สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) จำนวน 280,358 คน 2) สำนักงานคณะกรรมการส่งเสริมการศึกษาเอกชน (สช.) จำนวน 49,032 คน 3) กระทรวงการอุดมศึกษา (อว.) จำนวน 4,670 คน 4) กรมส่งเสริมการปกครองท้องถิ่น (กสท.) จำนวน 21,176 คน และ 5) อื่น ๆ จำนวน 8,516 คน ในจำนวนนี้เป็นนักเรียนที่มาจากรโรงเรียนขนาดเล็ก (น้อยกว่า 300 คน) 22,765 คน โรงเรียนขนาดกลาง (300-999 คน) 81,294 คน โรงเรียนขนาดใหญ่ (1,000-1,999 คน) 85,582 คน และโรงเรียนขนาดใหญ่พิเศษ (ตั้งแต่ 2,000 คนขึ้นไป) 174,111 คน (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2562ข)

ตัวอย่างที่นำมาศึกษาคือนักเรียนระดับ ม.6 ที่เข้าร่วมการสอบ O-NET วิชาคณิตศาสตร์ ปีการศึกษา 2562 จากโรงเรียนใน 4 สังกัด (ไม่รวมสังกัดอื่น ๆ) จำนวน 2,500 คน โดยดำเนินการสุ่มนักเรียนจากโรงเรียนแต่ละแห่ง โรงเรียนละ 50 คนมาเป็นกลุ่มตัวอย่าง รายละเอียดดังตาราง 3.10

ตาราง 3.10 จำนวนตัวอย่างของโรงเรียนและนักเรียน

สังกัด	จำนวนโรงเรียนตามขนาด				จำนวนนักเรียน
	เล็ก	กลาง	ใหญ่	ใหญ่พิเศษ	
1. สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.)	4	6	6	6	1,100
2. สำนักงานคณะกรรมการส่งเสริมการศึกษาเอกชน (สช.)	1	2	2	2	350
3. กระทรวงการอุดมศึกษา (อว.)	1	2	2	2	350
4. กรมส่งเสริมการปกครองท้องถิ่น (กสท.)	2	4	4	4	700
รวม	8	14	14	14	2,500

2) ตัวแปรที่ศึกษา

ตัวแปรที่ศึกษาจากข้อมูลเชิงประจักษ์แบ่งเป็น 2 ประเภท ดังนี้ 1) ตัวแปรสังเกตได้ ได้แก่ จำนวนข้อสอบ (32 ข้อ) จำนวนโรงเรียน (50 โรงเรียน) จำนวนนักเรียนในแต่ละโรงเรียน (50 คน) และผลการตอบข้อสอบของนักเรียน (0, 1) และ 2) ตัวแปรแฝง ได้แก่ จำนวนกลุ่มแฝง พารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบในกลุ่มแฝง และความสามารถเฉลี่ยในกลุ่มแฝง

3) เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลคือ แบบสอบ O-NET วิชาคณิตศาสตร์ ระดับ ม.6 ปีการศึกษา 2562 ประกอบด้วยข้อสอบปรนัยแบบ 5 ตัวเลือกจำนวน 32 ข้อ และข้อสอบอัตนัยแบบเติมตัวเลขจำนวน 8 ข้อ อย่างไรก็ตามงานวิจัยนี้จะสนใจเฉพาะข้อมูลผลการตอบข้อสอบปรนัย ซึ่งมีเนื้อหาครอบคลุม 4 สาระ ได้แก่ 1) จำนวนและการดำเนินการ (5 ข้อ) 2) การวัด (2 ข้อ) 3) พีชคณิต (16 ข้อ) และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น (9 ข้อ)

4) ขั้นตอนการศึกษา

4.1) ดำเนินการขอความอนุเคราะห์ข้อมูลคะแนนสอบ O-NET วิชาคณิตศาสตร์ ระดับ ม.6 ปีการศึกษา 2562 จากสถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.)

4.2) ดำเนินการสุ่มตัวอย่างโรงเรียนจำแนกตามขนาดและสังกัดมา 50 โรงเรียน และสุ่มตัวอย่างนักเรียนจากโรงเรียนแต่ละแห่งมา 50 คน รวมมีขนาดตัวอย่าง 2,500 คน ทั้งนี้กระบวนการสุ่มดังกล่าวจะดำเนินการด้วยโปรแกรม R แพคเกจ dplyr (Wickham et al., 2020)

4.3) ตรวจสอบจำนวนมิติของความสามารถที่มุ่งวัดจากข้อมูลผลการตอบข้อสอบที่สุ่มมา โดยจะตรวจสอบด้วยวิธีไร้พารามิเตอร์ (nonparametric method) ด้วยดัชนี *DETECT*, *ASSI* และ *RATIO* ซึ่งเป็นตัวชี้วัดขนาดอิทธิพล (effect size) ความเป็นพหุมิติของความสามารถที่มุ่งวัด (Jang & Roussos, 2007; Zhang, 2007) ทั้งนี้การตรวจสอบดังกล่าวจะดำเนินการวิเคราะห์ด้วยโปรแกรม R แพคเกจ sirt โดยถ้า $DETECT < .20$, $ASSI < .25$ และ $RATIO < .36$ จะถือว่าความสามารถที่มุ่งวัดเป็นแบบเอกมิติ ส่วนกรณีอื่น ๆ จะเป็นแบบพหุมิติ (Robitzsch, 2020)

4.4) ตรวจสอบความเป็นพหุระดับของข้อมูลจากค่าสหสัมพันธ์ภายในชั้น (*ICC*) ด้วยโมเดลพหุระดับและโมเดล multilevel IRT หาก $ICC > .05$ จะถือว่าข้อมูลมีความเหมาะสมสำหรับการวิเคราะห์แบบพหุระดับ (Hox & Maas, 2001)

4.5) ตรวจสอบจำนวนกลุ่มแฝงของข้อมูล จากการคัดเลือกโมเดลระหว่างโมเดลแข่งขัน MMix3PL กับ MMix2PLE ด้วยเกณฑ์สารสนเทศ AIC, BIC, CAIC และ DBIC หากโมเดลใดมีเกณฑ์สารสนเทศต่ำที่สุด จะหมายความว่าโมเดลนั้นถูกเลือกให้มีความสอดคล้องกับข้อมูลมากที่สุด

4.6) ตรวจสอบการสลับสัญลักษณ์ของพารามิเตอร์กลุ่มแฝงและการลู่ออกของลูกโซ่

4.7) วิเคราะห์ข้อมูลเพื่อประมาณค่าพารามิเตอร์ต่าง ๆ และสรุปผลการวิเคราะห์

5) สถิติที่ใช้ในการวิเคราะห์

สถิติที่ใช้มี 2 ประเภท ดังนี้ 1) สถิติเชิงบรรยาย ได้แก่ ความถี่ สัดส่วน ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน และ 2) สถิติเชิงอนุมาน ได้แก่ การประมาณค่าพารามิเตอร์แบบเบสด้วยโมเดล MMix3PL และ MMix2PLE และการวิเคราะห์ด้วยโมเดลถดถอยโลจิสติก (logistic regression)

บทที่ 4

ผลการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อต้องการหาเกณฑ์สารสนเทศที่มีความเหมาะสมสำหรับการคัดเลือกโมเดลให้กับโมเดล MMix3PL และ MMix2PLE รวมถึงต้องการศึกษารูปแบบของโมเดล MMix2PLE ที่มีการนำคุณลักษณะของผู้สอบและข้อสอบเข้ามาช่วยอธิบายโอกาสในการเดาถูก เพื่อหลีกเลี่ยงปัญหาการประมาณค่าพารามิเตอร์การเดาโดยตรงจากโมเดล MMix3PL ที่มีแนวโน้มไม่สัมพันธ์กับค่าจริง และมีการแปลผลการเดาที่ยังไม่ค่อยสอดคล้องกับธรรมชาติของการตอบข้อสอบ เนื่องจากโมเดลเชื่อว่าโอกาสในการเดาถูกเป็นคุณลักษณะของข้อสอบที่มีได้ผันแปรตามความสามารถของผู้สอบ ทั้งนี้การศึกษาดังกล่าวจะดำเนินการด้วยวิธีการเชิงจำลองมอนติคาร์โล รวมถึงจะมีการสาธิตการใช้โมเดลกับข้อมูลเชิงประจักษ์ เพื่อเปรียบเทียบผลการวิเคราะห์ที่ได้ระหว่างโมเดล MMix3PL กับ MMix2PLE ทั้งนี้การนำเสนอผลการวิจัยจะแบ่งเป็น 2 ตอน ได้แก่ 1) ผลการศึกษาภายใต้สถานการณ์จำลอง และ 2) ผลการศึกษาจากข้อมูลเชิงประจักษ์ รายละเอียดดังนี้

ตอนที่ 1 ผลการศึกษาภายใต้สถานการณ์จำลอง

ผลการศึกษาในส่วนนี้เป็นการศึกษาเพื่อหาเกณฑ์สารสนเทศที่มีความเหมาะสมสำหรับการคัดเลือกโมเดลให้กับโมเดล MMix3PL และ MMix2PLE และเพื่อตรวจสอบประสิทธิภาพของขั้นตอนวิธีการประมาณค่าแบบเบย์สำหรับโมเดล MMix3PL และ MMix2PLE เกี่ยวกับความแม่นยำในการประมาณค่าพารามิเตอร์ การนำเสนอแบ่งเป็น 2 ส่วน ได้แก่ ความถูกต้องของการคัดเลือกโมเดล และความแม่นยำของการประมาณค่าพารามิเตอร์ รายละเอียดดังนี้

1) ความถูกต้องของการคัดเลือกโมเดล

จากการพิจารณาผลการคัดเลือกโมเดลด้วยเกณฑ์สารสนเทศ AIC, BIC, CAIC และ DBIC พบว่า AIC สามารถเลือกโมเดลได้ตรงกับโมเดลประชากรที่มีกลุ่มแฝงระดับนักเรียน (K) 2 กลุ่ม และกลุ่มแฝงระดับโรงเรียน (G) 2 กลุ่ม ทั้ง 10 ชุดข้อมูล (ร้อยละ 100) ทั้งในกรณี $N = 1,250$ และ $2,500$ สำหรับโมเดล MMix3PL และ MMix2PLE ส่วน BIC และ CAIC พบว่าไม่สามารถเลือกโมเดลได้ตรงกับโมเดลประชากรแม้แต่ชุดเดียว (ร้อยละ 0) ยกเว้นกรณี $N = 2,500$ สำหรับโมเดล MMix2PLE ที่สามารถเลือกโมเดลได้ตรงกับโมเดลประชากรทั้ง 10 ชุดข้อมูล (ร้อยละ 100) และ DBIC พบว่าไม่สามารถเลือกโมเดลได้ตรงกับโมเดลประชากรแม้แต่ชุดเดียว (ร้อยละ 0) ในกรณี $N = 1,250$ สำหรับโมเดล MMix3PL ตรงกันข้ามกับโมเดล MMix2PLE ที่สามารถเลือกโมเดลได้ตรงถึง 6 ชุดข้อมูล (ร้อยละ 60) ส่วนกรณี $N = 2,500$ พบว่า DBIC สามารถเลือกโมเดลได้ตรง 8 ชุดข้อมูล (ร้อยละ 80) และ 10 ชุดข้อมูล (ร้อยละ 100) สำหรับโมเดล MMix3PL และ MMix2PLE ตามลำดับ รายละเอียดดังตาราง 4.1

ตาราง 4.1 ค่าเฉลี่ยเกณฑ์สารสนเทศกับความถูกต้องของการคัดเลือกโมเดล

โมเดล	จำนวนพารามิเตอร์	N = 1,250				N = 2,500			
		AIC	BIC	CAIC	DBIC	AIC	BIC	CAIC	DBIC
3PL	121	57722 (0.0)	58342 (0.0)	58462 (0.0)	58122 (0.0)	115092 (0.0)	115796 (0.0)	115922 (0.0)	115572 (0.0)
multilevel 3PL	123	58509 (0.0)	59140 (0.0)	59259 (0.0)	58919 (0.0)	116875 (0.0)	117592 (0.0)	117715 (0.0)	117365 (0.0)
Mix3PL (K=2)	244	52438 (0.0)	53401 (100.0)	53938 (100.0)	53238 (30.0)	104295 (0.0)	105715 (0.0)	105965 (0.0)	105265 (0.0)
Mix3PL (K=3)	367	51964 (0.0)	53846 (0.0)	54214 (0.0)	53174 (70.0)	102986 (0.0)	105122 (100.0)	105486 (100.0)	104446 (20.0)
*MMix3PL (K=2, G=2)	490	51797 (100.0)	54312 (0.0)	54797 (0.0)	53407 (0.0)	102431 (100.0)	105283 (0.0)	105771 (0.0)	104381 (80.0)
MMix3PL (K=3, G=2)	736	52090 (0.0)	55866 (0.0)	56600 (0.0)	54510 (0.0)	102663 (0.0)	107051 (0.0)	107683 (0.0)	105593 (0.0)
MMix3PL (K=2, G=3)	736	52295 (0.0)	56071 (0.0)	56805 (0.0)	54715 (0.0)	102929 (0.0)	107215 (0.0)	107949 (0.0)	105859 (0.0)
MMix3PL (K=3, G=3)	1105	52828 (0.0)	58498 (0.0)	59598 (0.0)	56468 (0.0)	103401 (0.0)	109834 (0.0)	110941 (0.0)	107801 (0.0)
2PLE	81	53501 (0.0)	53917 (0.0)	54001 (0.0)	53771 (0.0)	106067 (0.0)	106537 (0.0)	106617 (0.0)	106387 (0.0)
multilevel 2PLE	83	54123 (0.0)	54547 (0.0)	54633 (0.0)	54393 (0.0)	108001 (0.0)	108486 (0.0)	108571 (0.0)	108331 (0.0)
Mix2PLE (K=2)	164	46467 (0.0)	47308 (0.0)	47477 (0.0)	47007 (0.0)	92477 (0.0)	93431 (0.0)	93597 (0.0)	93127 (0.0)
Mix2PLE (K=3)	247	45279 (0.0)	46548 (100.0)	46789 (100.0)	46089 (40.0)	89966 (0.0)	91406 (0.0)	91656 (0.0)	90946 (0.0)
*MMix2PLE (K=2, G=2)	330	45006 (100.0)	46700 (0.0)	47026 (0.0)	46096 (60.0)	89212 (100.0)	91133 (100.0)	91462 (100.0)	90532 (100.0)
MMix2PLE (K=3, G=2)	496	45203 (0.0)	47751 (0.0)	48243 (0.0)	46833 (0.0)	89473 (0.0)	92361 (0.0)	92853 (0.0)	91453 (0.0)
MMix2PLE (K=2, G=3)	496	45354 (0.0)	47900 (0.0)	48394 (0.0)	46984 (0.0)	89579 (0.0)	92467 (0.0)	92959 (0.0)	91559 (0.0)
MMix2PLE (K=3, G=3)	745	45694 (0.0)	49515 (0.0)	50264 (0.0)	48144 (0.0)	89956 (0.0)	94295 (0.0)	95036 (0.0)	92926 (0.0)

หมายเหตุ: ตัวหนาและขีดเส้นใต้คือ เกณฑ์สารสนเทศเลือกโมเดลได้ตรงกับโมเดลประชากร; เลขในวงเล็บคือ ร้อยละที่โมเดลถูกเลือกจากข้อมูล 10 ชุด

2) ความแม่นยำของการประมาณค่าพารามิเตอร์

ในส่วนนี้เป็นการพิจารณาผลการประมาณค่าพารามิเตอร์ระหว่างโมเดล MMix3PL ($K=2, G=2$) กับ MMix2PLE ($K=2, G=2$) ซึ่งเป็นโมเดลแข่งขันที่มีกลุ่มแฝงตรงกับโมเดลประชากร เพื่อศึกษาว่าโมเดลสามารถประมาณค่าพารามิเตอร์ย้อนกลับได้สอดคล้องกับค่าจริงเพียงใด รายละเอียดดังนี้

- **กลุ่มแฝง** พบว่าโมเดล MMix3PL และ MMix2PLE มีความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝงระดับโรงเรียน (π_g) สูงร้อยละ 100.0 เท่ากัน ส่วนกลุ่มแฝงระดับนักเรียน (π_{lg}) มีค่าสูงร้อยละ 96.02 ถึง 97.71 และ 98.92 ถึง 99.25 ตามลำดับ (ดูตาราง 4.2ก)

- **ความสามารถเฉลี่ย** พบว่าโมเดล MMix2PLE (bias=-.117 ถึง -.084, RMSE=.082 ถึง .097) และโมเดล MMix3PL (bias=-.193 ถึง -.156, RMSE=.135) มีแนวโน้มให้ความแม่นยำของการประมาณค่าพารามิเตอร์ความสามารถเฉลี่ย (μ) ค่อนข้างสูงใกล้เคียงกัน (ดูตาราง 4.2ข)

- **ข้อสอบ** ในการประมาณค่าพารามิเตอร์อำนาจจำแนก (a) และความยาก (b) พบว่าโมเดล MMix2PLE (bias=-.039 ถึง .000, RMSE=.037 ถึง .114, corr=.667 ถึง .991) และโมเดล MMix3PL (bias=-.154 ถึง .000, RMSE=.053 ถึง .115, corr=.772 ถึง .987) มีแนวโน้มให้ความแม่นยำของการประมาณค่าค่อนข้างสูงใกล้เคียงกัน อย่างไรก็ตามค่าพารามิเตอร์การเดา (c) ของโมเดล MMix3PL (bias=.025 ถึง .027, RMSE=.006 ถึง .007, corr=-.290 ถึง -.186) พบว่ายังคงมีข้อผิดพลาดเล็กน้อย (bias และ RMSE ต่ำเข้าใกล้ 0) แต่กลับไม่มีความสัมพันธ์กับค่าจริง (corr น้อยกว่า 0) ทั้งกรณี $N=1,250$ และ $2,500$ (ดูตาราง 4.2ข)

ตาราง 4.2 ความแม่นยำของการประมาณค่าพารามิเตอร์

(ก) พารามิเตอร์กลุ่มแฝง							
N	โมเดล MMix3PL			โมเดล MMix2PLE			
	p (ระดับโรงเรียน)	p (ระดับนักเรียน)		p (ระดับโรงเรียน)	p (ระดับนักเรียน)		
π	1,250	100	96.02	100	98.92		
	2,500	100	97.71	100	99.25		
(ข) ความสามารถเฉลี่ยและพารามิเตอร์ของข้อสอบ							
N	โมเดล MMix3PL			โมเดล MMix2PLE			
	bias	RMSE	corr	bias	RMSE	corr	
μ	1,250	-.156	.135	-	-.117	.097	-
	2,500	-.193	.135	-	-.084	.082	-
a	1,250	-.154	.115	.772	-.039	.114	.667
	2,500	-.101	.077	.838	-.010	.078	.782
b	1,250	.000	.082	.980	.000	.055	.987
	2,500	.000	.053	.987	.000	.037	.991
c	1,250	.027	.007	-.290	-	-	-
	2,500	.025	.006	-.186	-	-	-

ตอนที่ 2 ผลการศึกษาจากข้อมูลเชิงประจักษ์

ในส่วนนี้เป็นการสาธิตการใช้โมเดล MMix3PL และ MMix2PLE วิเคราะห์ข้อมูลคะแนนสอบ O-NET ที่สุ่มมาจากนักเรียนจำนวน 2,500 คน การนำเสนอแบ่งเป็น 4 ส่วน ได้แก่ 1) การตรวจสอบความเหมาะสมของข้อมูล 2) การคัดเลือกโมเดล 3) การตรวจสอบความเหมาะสมของการประมาณค่า และ 4) ผลการวิเคราะห์ข้อมูล รายละเอียดดังนี้

1) การตรวจสอบความเหมาะสมของข้อมูล

- **ข้อมูลคะแนนสอบ** จากข้อมูลตัวอย่างที่สุ่มมาพบว่านักเรียนจากโรงเรียนสังกัด อว. มีคะแนนเฉลี่ยรวมสูงที่สุด (37.88 คะแนน) รองลงมาคือ สช. (25.68 คะแนน) สพฐ. (17.84 คะแนน) และต่ำที่สุดคือ กสท. (16.41 คะแนน) นอกจากนี้เมื่อพิจารณาคะแนนจำแนกตามกลุ่มสาระของข้อสอบ ก็ยังพบว่านักเรียนจากโรงเรียนสังกัด อว. มีคะแนนเฉลี่ยสูงกว่านักเรียนจากโรงเรียนสังกัดอื่นในทุกกลุ่มสาระด้วยเช่นกัน รายละเอียดดังตาราง 4.3 จากผลดังกล่าวจึงอาจอนุมานได้ว่าข้อมูลน่าจะมีภาวะวิวิธพันธุ์เกิดขึ้นในกลุ่มผู้สอบ และน่าจะจำแนกผู้สอบออกเป็นกลุ่มแฝงตามแบบแผนการตอบข้อสอบที่แตกต่างกันได้

ตาราง 4.3 ข้อมูลคะแนนสอบจำแนกตามสังกัดและขนาดของโรงเรียน

สังกัด	ขนาด	1) จำนวนและการดำเนินการ (12.5)	2) การวัด (5)	3) พีชคณิต (40)	4) การวิเคราะห์ข้อมูล (22.5)	คะแนนรวม (80)
		ค่าเฉลี่ย (SD)	ค่าเฉลี่ย (SD)	ค่าเฉลี่ย (SD)	ค่าเฉลี่ย (SD)	ค่าเฉลี่ย (SD)
สพฐ.	เล็ก	2.53 (2.28)	1.05 (1.43)	7.75 (4.21)	4.59 (3.23)	15.91 (6.59)
	กลาง	2.75 (2.22)	1.07 (1.57)	7.82 (4.82)	4.43 (3.28)	16.06 (7.84)
	ใหญ่	2.80 (2.39)	1.08 (1.54)	8.41 (5.11)	5.39 (3.65)	17.68 (8.37)
	ใหญ่พิเศษ	3.18 (2.75)	1.57 (1.84)	10.7 (6.59)	5.62 (3.99)	21.07 (11.71)
	ภาพรวม	2.84 (2.44)	1.20 (1.63)	8.75 (5.47)	5.04 (3.61)	17.84 (9.24)
สช.	เล็ก	11.6 (1.49)	4.85 (0.60)	38.15 (3.19)	19.6 (2.44)	74.20 (5.68)
	กลาง	2.65 (2.24)	1.18 (1.53)	8.03 (4.25)	5.25 (3.45)	17.10 (7.37)
	ใหญ่	2.28 (2.19)	0.75 (1.31)	7.68 (4.22)	4.00 (3.22)	14.70 (7.00)
	ใหญ่พิเศษ	3.05 (2.90)	1.68 (1.92)	10.08 (6.64)	6.18 (4.24)	20.98 (12.48)
	ภาพรวม	3.94 (3.92)	1.72 (2.00)	12.81 (11.51)	7.21 (6.21)	25.68 (21.85)
อว.	เล็ก	9.70 (3.45)	4.25 (1.45)	32.20 (8.90)	15.95 (5.48)	62.10 (16.60)
	กลาง	4.83 (3.68)	2.05 (2.11)	15.23 (9.29)	8.75 (5.79)	30.85 (18.36)
	ใหญ่	5.13 (3.87)	2.73 (2.01)	18.60 (10.45)	10.23 (5.80)	36.68 (19.46)
	ใหญ่พิเศษ	5.00 (4.29)	2.08 (2.10)	17.55 (11.17)	9.38 (5.82)	34.00 (21.30)
	ภาพรวม	5.66 (4.21)	2.56 (2.13)	19.28 (11.47)	10.38 (6.19)	37.88 (21.76)
กสท.	เล็ก	3.18 (2.10)	0.68 (1.22)	7.58 (4.37)	4.43 (3.05)	15.85 (6.23)
	กลาง	2.66 (2.28)	0.85 (1.43)	7.31 (4.79)	4.38 (3.36)	15.20 (7.95)
	ใหญ่	2.89 (2.26)	1.00 (1.42)	8.33 (4.91)	4.50 (3.23)	16.71 (8.41)
	ใหญ่พิเศษ	3.08 (2.32)	1.19 (1.56)	8.46 (5.37)	4.89 (3.31)	17.61 (8.10)
	ภาพรวม	2.92 (2.26)	0.96 (1.45)	7.97 (4.96)	4.56 (3.27)	16.41 (7.95)

- มิติของความสามารถแฝงที่มุ่งวัด พบว่าความสามารถแฝงที่มุ่งวัดมีแนวโน้มเป็นแบบเอกมิติ เนื่องจากการวัดขนาดอิทธิพลความเป็นพหุมิติของความสามารถแฝงภายใต้การวิเคราะห์เชิงยืนยัน ตามกลุ่มสาระของข้อสอบ (ดูภาพ 4.1 ซ้าย) ด้วยโมเดล 3PL จากโปรแกรม R แพคเกจ sirt พบว่าดัชนี *DETECT*, *ASSI* และ *RATIO* มีค่าเป็น .14, .21 และ .31 ตามลำดับ (ดูภาพ 4.1 ขวา) ซึ่งค่าดังกล่าวอยู่ในเงื่อนไข $DETECT < .20$, $ASSI < .25$ และ $RATIO < .36$ (Robitzsch, 2020) จากผลดังกล่าวจึงสรุปได้ว่าความสามารถที่มุ่งวัดมีเพียงมิติเดียว

```

> iteminfo <- read.csv("item.csv", header=T)
> iteminfo
  item Format      CONTENT DOMAIN
1  C1      MC      Number and Operation
2  C2      MC      Number and Operation
3  C3      MC      Number and Operation
4  C4      MC      Number and Operation
5  C5      MC      Number and Operation
6  C6      MC      Measurement
7  C7      MC      Measurement
8  C8      MC      Algebra
9  C9      MC      Algebra
10 C10     MC      Algebra
11 C11     MC      Algebra
12 C12     MC      Algebra
13 C13     MC      Algebra
14 C14     MC      Algebra
15 C15     MC      Algebra
16 C16     MC      Algebra
17 C17     MC      Algebra
18 C18     MC      Algebra
19 C19     MC      Algebra
20 C20     MC      Algebra
21 C21     MC      Algebra
22 C22     MC      Algebra
23 C23     MC      Algebra
24 C24     MC      Data Analysis and Probability
25 C25     MC      Data Analysis and Probability
26 C26     MC      Data Analysis and Probability
27 C27     MC      Data Analysis and Probability
28 C28     MC      Data Analysis and Probability
29 C29     MC      Data Analysis and Probability
30 C30     MC      Data Analysis and Probability
31 C31     MC      Data Analysis and Probability
32 C32     MC      Data Analysis and Probability

> # O-NEI math62 data set (dichotomous data)
> head(dat)
  C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C12 C13 C14 C15 C16 C17 C18 C19 C20
1 0 0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0
2 0 0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0
3 0 0 0 0 0 0 0 1 1 1 1 0 0 1 0 0 1 0 0 1 0 0 0
4 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
5 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6 0 0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 1 0 0 1 0 0 0
  C21 C22 C23 C24 C25 C26 C27 C28 C29 C30 C31 C32
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0
5 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0
6 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

> I <- ncol(dat)
> quiet <- function(x) {
+   sink(tempfile())
+   on.exit(sink())
+   invisible(force(x))
+ }
> # Estimate 3PL model
> mod1 <- quiet(rasch.mm12(dat, est.a=I, est.c=I, mm1iter=400))
> wle1 <- quiet(wle.rasch(dat, b=mod1$itemb)$theta)
>
> # DETECT for content domains
> detect1 <- quiet(conf.detect(data=dat, score=wle1,
+ itemcluster=iteminfo$CONTENT_DOMAIN))
> detect1$detect
  unweighted  weighted
DETECT  0.1394619  0.1394619
ASSI    0.2056774  0.2056774
RATIO   0.3052710  0.3052710
MADCOV100 0.4509376  0.4509376
MCCOV100  -0.3902018  -0.3902018

```

ภาพ 4.1 กลุ่มสาระข้อสอบ (ซ้าย) และผลการตรวจสอบมิติของความสามารถที่มุ่งวัด (ขวา)

- **โครงสร้างพหุระดับของข้อมูล** พบว่าข้อมูลมีความเหมาะสมสำหรับการวิเคราะห์แบบพหุระดับ เนื่องจากการวิเคราะห์ด้วยโมเดลพหุระดับและโมเดล multilevel IRT พบว่า $ICC > .05$ ซึ่งหมายความว่าความแปรปรวนของข้อมูลมีความเพียงพอที่จะอธิบายได้ด้วยระดับโรงเรียน รายละเอียดดังนี้

ผลการวิเคราะห์ด้วยโมเดลพหุระดับจากโปรแกรม R แพคเกจ lme4 ที่วัดสมรรถภาพทางคณิตศาสตร์จากคะแนนรวมซึ่งเป็นตัวแปรสังเกตได้ พบว่าความแปรปรวนของความคลาดเคลื่อนในระดับนักเรียนและระดับโรงเรียนมีค่าเท่ากับ 17.78 และ 21.21 ตามลำดับ และมี $ICC = .54$ (ดูภาพ 4.2 ซ้าย) ส่วนผลการวิเคราะห์ด้วยโมเดล multilevel 3PL และ multilevel 2PLE จากโปรแกรม MultiBUGS (ดูภาคผนวก ข.3-ข.4) ที่วัดสมรรถภาพฯ จากค่าความสามารถซึ่งเป็นตัวแปรแฝง พบว่าความแปรปรวนของความคลาดเคลื่อนในระดับนักเรียนมีค่าเท่ากับ 3.35 และ 4.49 และในระดับโรงเรียนมีค่าเท่ากับ 3.39 และ 5.47 และมี $ICC = .51, .55$ ตามลำดับ (ดูภาพ 4.2 ขวา)

<pre>> # Demographic information > head(dat) STUDENT_SEQ PARENT_INCOME INCOME_RANGE NAME SCHOOL_CODE BUREA_CODE 1 670174 1 < 50,000 1048022002 1 2 670152 1 < 50,000 1048022002 1 3 670154 1 < 50,000 1048022002 1 4 670148 1 < 50,000 1048022002 1 5 670134 3 100,001-200,000 1048022002 1 6 670172 1 < 50,000 1048022002 1 SCHOOL_SIZE SCHOOL_SIZE_CODE STD_KEY GENDER MATH_SCORE 1 S 1 12553529 F 4 2 S 1 12553561 F 7 3 S 1 12553530 F 7 4 S 1 12553551 F 5 5 S 1 12553538 M 5 6 S 1 12553540 F 5 > > # Compute ICC for multilevel Model > null.model <- lmer(MATH_SCORE~1+(1 SCHOOL_CODE),data=dat); null.model Linear mixed model fit by REML ['EigenModLmerTest'] Formula: MATH_SCORE ~ 1 + (1 SCHOOL_CODE) Data: dat REML criterion at convergence: 14493.96 Random effects: Groups Name Std.Dev. SCHOOL_CODE (Intercept) 4.605 Residual 4.217 Number of obs: 2500, groups: SCHOOL_CODE, 50 Fixed Effects: (Intercept) 8.538 > var.stu <- 4.217^2; var.stu [1] 17.78309 > var.sch <- 4.605^2; var.sch [1] 21.20603 > ICC <- var.sch/(var.sch+var.stu); ICC [1] 0.543896</pre>	<pre>> # Results of Multilevel 3PL from MultiBUGS > # mean median sd MC_error val2.5pc val97.5pc > # gamma -3.34 -3.34 0.303 0.0158 -3.93 -2.75 > # tau 1.83 1.84 0.125 0.0130 1.56 2.04 > # zeta 0.29 0.29 0.077 0.0048 0.17 0.47 > # Note: tau = sqrt(tau.model); zeta.model = 1/zeta > > # ICC for multilevel 3PL > tau.model <- 1.83^2; tau.model [1] 3.3489 > zeta.model <- 1/.29; zeta.model [1] 3.448276 > ICC.M3PL <- zeta.model/(zeta.model+tau.model) > ICC.M3PL [1] 0.5073101 > > # Results of Multilevel 2PLE from MultiBUGS > # mean median sd MC_error val2.5pc val97.5pc > # gamma -4.67 -4.67 0.417 0.0259 -5.49 -3.87 > # tau 2.12 2.10 0.142 0.0148 1.90 2.41 > # zeta 0.18 0.18 0.046 0.0025 0.10 0.29 > # Note: tau = sqrt(tau.model); zeta.model = 1/zeta > > # ICC for multilevel 2PLE > tau.model <- 2.12^2; tau.model [1] 4.4944 > zeta.model <- 1/.18; zeta.model [1] 5.555556 > ICC.M2PLE <- zeta.model/(zeta.model+tau.model) > ICC.M2PLE [1] 0.552794</pre>
--	--

ภาพ 4.2 ผลการวิเคราะห์ด้วยโมเดลพหุระดับ (ซ้าย) และโมเดล multilevel IRT (ขวา)

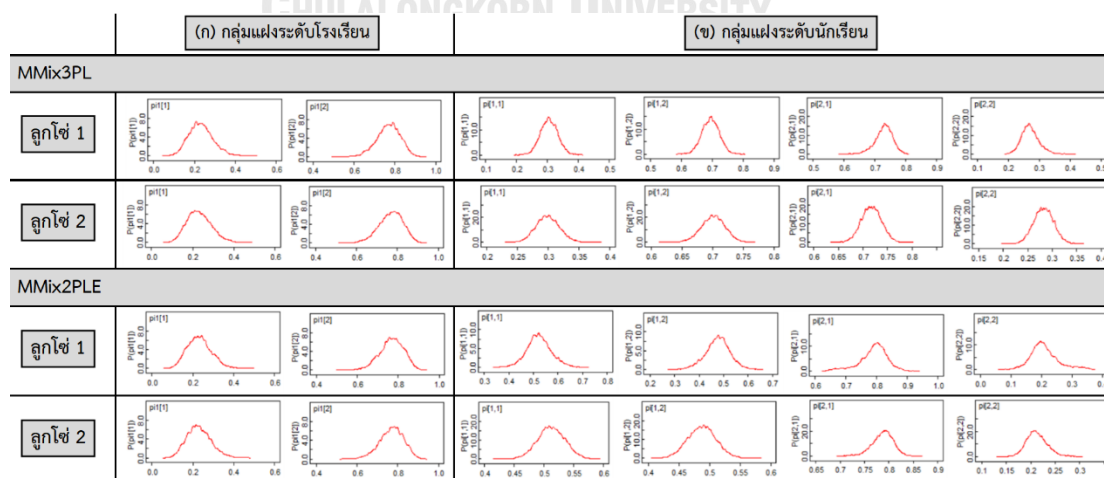
2) **การคัดเลือกโมเดล** พบว่าเกณฑ์สารสนเทศ AIC, BIC, CAIC และ DBIC เลือกให้โมเดล MMix3PL ($K=2, G=2$) มีความสอดคล้องกับข้อมูลมากที่สุด เนื่องจากเกณฑ์สารสนเทศของโมเดลดังกล่าวมีค่าต่ำที่สุดเมื่อเทียบกับโมเดลแข่งขันอื่น ๆ (ดูตาราง 4.4ก) จึงอาจสรุปได้ว่าข้อมูลนี้น่าจะมีกลุ่มแฝงระดับนักเรียน (K) 2 กลุ่ม และกลุ่มแฝงระดับโรงเรียน (G) 2 กลุ่ม อย่างไรก็ตามเมื่อพิจารณาผลการคัดเลือกโมเดลตามปรัชญา 2PLE แม้จะพบว่าโมเดล MMix2PLE มีความสอดคล้องกับข้อมูลน้อยกว่าโมเดล MMix3PL แต่ก็สามารถให้ผลสรุปเกี่ยวกับจำนวนกลุ่มแฝงของข้อมูลได้ไม่ต่างจาก MMix3PL ดังสังเกตได้จากเกณฑ์สารสนเทศของโมเดล MMix2PLE ($K=2, G=2$) ที่มีค่าต่ำที่สุดเมื่อเทียบกับโมเดลแข่งขันตามปรัชญา 2PLE (ดูตาราง 4.4ข) ดังนั้นผู้วิจัยจึงขอแนะนำผลการวิเคราะห์ที่ได้จากโมเดล MMix3PL ($K=2, G=2$) และ MMix2PLE ($K=2, G=2$) มาศึกษาเปรียบเทียบเพื่ออธิบายคุณลักษณะของกลุ่มแฝงต่อไป

ตาราง 4.4 การคัดเลือกโมเดลกรณีการศึกษาข้อมูลเชิงประจักษ์

โมเดล	จำนวนพารามิเตอร์	AIC	BIC	CAIC	DBIC
(ก) โมเดลตามปรัชญา 3PL					
MMix3PL ($K = 2, G = 2$)	394	77570	79870	80260	79140
MMix3PL ($K = 3, G = 2$)	592	77790	81240	81830	80150
MMix3PL ($K = 4, G = 2$)	790	78030	82630	83420	81180
MMix3PL ($K = 3, G = 3$)	889	78590	83770	84660	82130
MMix3PL ($K = 4, G = 3$)	1186	79110	86020	87200	83840
(ข) โมเดลตามปรัชญา 2PLE					
MMix2PLE ($K = 2, G = 2$)	266	78640	80190	80460	79700
MMix2PLE ($K = 3, G = 2$)	400	78880	81210	81610	80470
MMix2PLE ($K = 4, G = 2$)	534	79190	82300	82830	81320
MMix2PLE ($K = 3, G = 3$)	601	79220	82720	83320	81620
MMix2PLE ($K = 4, G = 3$)	802	79750	84420	85220	82950

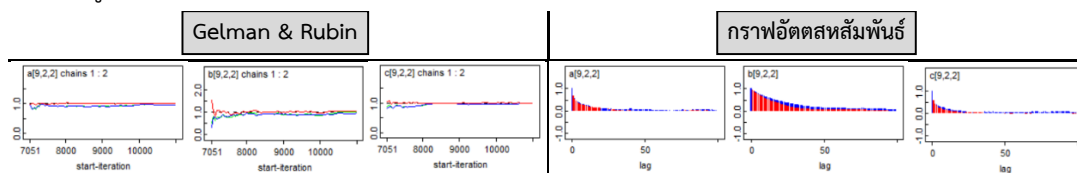
3) การตรวจสอบความเหมาะสมของการประมาณค่า

- การตรวจสอบการสลับสัญลักษณ์ พบว่ากราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์กลุ่มแฝงมีลักษณะเป็นเอกฐานนิยม ซึ่งแสดงถึงการที่ค่าประมาณพารามิเตอร์สามารถระบุค่าที่แน่นอนได้เพียงหนึ่งเดียว จึงอาจสรุปได้ว่าไม่มีการสลับสัญลักษณ์ประเภทที่ 1 นอกจากนี้เมื่อเทียบเคียงค่าประมาณพารามิเตอร์ที่ได้ระหว่างลูกโซ่ 1 และ 2 ของโมเดล MMix3PL และ MMix2PLE ยังพบว่าไม่มีการสลับสัญลักษณ์ประเภทที่ 2 เกิดขึ้นด้วย สังเกตได้จากค่าประมาณพารามิเตอร์ระหว่างลูกโซ่ทั้งสองที่มีค่าเฉลี่ยภายหลังใกล้เคียงกันมาก ดังภาพ 4.3



ภาพ 4.3 กราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์กลุ่มแฝงกรณีการศึกษาข้อมูลเชิงประจักษ์

- การวินิจฉัยการลู่เข้าของลูกโซ่ พบว่าโมเดล MMix3PL และ MMix2PLE มีแนวโน้มการลู่เข้าของลูกโซ่ในการประมาณค่าพารามิเตอร์ของข้อสอบค่อนข้างดีหลังจากตัดค่าสังเกต 7,000 ค่าแรกทิ้ง สังเกตได้จากค่าสถิติ Gelman & Rubin ที่มีแนวโน้มลู่เข้า 1 และระดับอัตโนมัติสหสัมพันธ์ลดค่อนข้างเร็วเมื่อ lag เพิ่มขึ้น ดังภาพ 4.4 อย่างไรก็ตามสำหรับค่าความยากของข้อสอบ พบว่าลูกโซ่มีแนวโน้มลู่เข้ายังไม่ดีเท่าที่ควร ซึ่งผลดังกล่าวค่อนข้างสอดคล้องกับผลที่ได้จากการศึกษาเชิงจำลอง



ภาพ 4.4 การวินิจฉัยการลู่เข้าของลูกโซ่กรณีการศึกษาข้อมูลเชิงประจักษ์

4) ผลการวิเคราะห์ข้อมูล ในส่วนนี้เป็นการนำเสนอผลการประมาณค่าพารามิเตอร์จากโมเดล MMix3PL ($K=2, G=2$) และ MMix2PLE ($K=2, G=2$) ซึ่งให้ผลสรุปเกี่ยวกับจำนวนกลุ่มแฝงของข้อมูลเหมือนกัน เพื่อศึกษาเปรียบเทียบว่าค่าประมาณพารามิเตอร์ในกลุ่มแฝงที่ได้จากโมเดลทั้งสองมีความคล้ายคลึงหรือแตกต่างกันเพียงใด การนำเสนอผลการประมาณค่าแบ่งเป็น 3 ส่วน ได้แก่ พารามิเตอร์กลุ่มแฝง พารามิเตอร์ความสามารถ และพารามิเตอร์ของข้อสอบ รายละเอียดดังนี้

- **กลุ่มแฝง** ผลการวิเคราะห์จากโมเดล MMix3PL พบว่าค่าประมาณสัดส่วนของสมาชิกในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 และ 2 มีค่าเป็น .231 และ .769 ตามลำดับ โดยในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 ($g=1$) มีกลุ่มแฝงระดับนักเรียนกลุ่มที่ 2 ($k=2$) เป็นกลุ่มเด่นด้วยค่าสัดส่วน .695 และในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 2 ($g=2$) มีกลุ่มแฝงระดับนักเรียนกลุ่มที่ 1 ($k=1$) เป็นกลุ่มเด่นด้วยค่าสัดส่วน .729 ส่วนผลการวิเคราะห์จากโมเดล MMix2PLE พบว่าค่าประมาณสัดส่วนของสมาชิกในกลุ่มแฝงระดับโรงเรียนมีความใกล้เคียงกับโมเดล MMix3PL เป็นอย่างมาก แต่มีข้อแตกต่างตรงที่กลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 ($g=1$) มีกลุ่มแฝงระดับนักเรียนกลุ่มที่ 1 ($k=1$) เป็นกลุ่มเด่นด้วยค่าสัดส่วน .524 และกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 2 ($g=2$) มีกลุ่มแฝงระดับนักเรียนกลุ่มที่ 1 ($k=1$) เป็นกลุ่มเด่นด้วยค่าสัดส่วน .790 ดังตาราง 4.5

ตาราง 4.5 ค่าสัดส่วนของสมาชิกในกลุ่มแฝงระดับโรงเรียน (π_g) และระดับนักเรียน ($\pi_{k|g}$)

โมเดล	กลุ่มแฝงระดับโรงเรียนที่ g	$\hat{\pi}_g$	กลุ่มแฝงระดับนักเรียนที่ k	
			$k=1$	$k=2$
MMix3PL	$g=1$.231	$\hat{\pi}_{1 1} = .305$	$\hat{\pi}_{2 1} = .695$
	$g=2$.769	$\hat{\pi}_{1 2} = .729$	$\hat{\pi}_{2 2} = .271$
MMix2PLE	$g=1$.232	$\hat{\pi}_{1 1} = .524$	$\hat{\pi}_{2 1} = .476$
	$g=2$.768	$\hat{\pi}_{1 2} = .790$	$\hat{\pi}_{2 2} = .210$

เมื่อพิจารณาผลการจำแนกโรงเรียนจำนวน 40 แห่ง ตามกลุ่มแฝงและตัวแปรสังเกตได้ระดับโรงเรียน (สังกัด, ขนาดโรงเรียน, ภูมิภาค) พบว่าโมเดล MMix3PL มีแนวโน้มไม่สามารถจำแนกลักษณะเด่นของสมาชิกในกลุ่มแฝงที่ 1 และ 2 ออกจากกันได้ สังเกตได้จากผลการจำแนกที่โรงเรียนส่วนใหญ่ถูกจัดให้อยู่ในกลุ่มแฝงที่ 2 สำหรับทุกกลุ่มตัวแปรสังเกตได้ ในขณะที่โมเดล MMix2PLE ให้ผลการจำแนกว่ากลุ่มแฝงกลุ่มที่ 1 มีลักษณะเด่นคือสมาชิกส่วนใหญ่เป็นโรงเรียนในสังกัดสำนักงานคณะกรรมการการอุดมศึกษา (อว.) และอยู่ในกรุงเทพฯ ส่วนกลุ่มแฝงกลุ่มที่ 2 มีลักษณะเด่นคือสมาชิกส่วนใหญ่เป็นโรงเรียนในสังกัดอื่นซึ่งมิใช่สังกัด อว. (สังกัด สพฐ., สช. และ กสท.) และอยู่ในจังหวัดอื่นที่มิใช่กรุงเทพฯ ดังตาราง 4.6

ตาราง 4.6 จำนวนและค่าสัดส่วนของโรงเรียนจำแนกตามตัวแปรและกลุ่มแฝงระดับโรงเรียน

ตัวแปรระดับโรงเรียน		MMix3PL		MMix2PLE	
		$g = 1$	$g = 2$	$g = 1$	$g = 2$
สังกัด	สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.)	4 (.18)	18 (.82)	2 (.09)	20 (.91)
	สำนักงานคณะกรรมการส่งเสริมการศึกษาเอกชน (สช.)	2 (.29)	5 (.71)	2 (.29)	5 (.71)
	สำนักงานคณะกรรมการการอุดมศึกษา (อว.)	1 (.14)	6 (.86)	7 (1.00)	0 (.00)
	กรมส่งเสริมการปกครองท้องถิ่น (กสท.)	4 (.29)	10 (.71)	0 (.00)	14 (1.00)
ขนาดโรงเรียน	เล็ก	3 (.38)	5 (.63)	2 (.25)	6 (.75)
	กลาง	2 (.14)	12 (.86)	2 (.14)	12 (.86)
	ใหญ่	1 (.07)	13 (.93)	2 (.14)	12 (.86)
	ใหญ่พิเศษ	5 (.36)	9 (.64)	5 (.36)	9 (.64)
ภูมิภาค	ใต้	3 (.27)	8 (.73)	1 (.09)	10 (.91)
	กรุงเทพฯ	1 (.14)	6 (.86)	6 (.86)	1 (.14)
	กลาง	2 (.25)	6 (.75)	3 (.38)	5 (.63)
	ตะวันออก	1 (.25)	3 (.75)	1 (.25)	3 (.75)
	เหนือ	1 (.25)	3 (.75)	0 (.00)	4 (1.00)
	ตะวันออกเฉียงเหนือ	3 (.21)	11 (.79)	0 (.00)	14 (1.00)
	ตะวันตก	0 (.00)	2 (1.00)	0 (.00)	2 (1.00)

จากผลการวิเคราะห์ข้างต้น เนื่องจากการจำแนกกลุ่มแฝงระดับโรงเรียนด้วยโมเดล MMix3PL ยังไม่มีความชัดเจนในการระบุลักษณะเด่นของสมาชิกในกลุ่มแฝง ดังนั้นผู้วิจัยจึงขอนำผลการจำแนกกลุ่มแฝงระดับโรงเรียนจากโมเดล MMix2PLE มาอธิบายการจำแนกกลุ่มแฝงระดับนักเรียนต่อไป

ผลการจำแนกกลุ่มแฝงระดับนักเรียนของโมเดล MMix2PLE พบว่าในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 ($g = 1$) มีกลุ่มแฝงระดับนักเรียนกลุ่มที่ 1 ($k = 1$) เป็นกลุ่มเด่น โดยสมาชิกส่วนใหญ่เป็นนักเรียนที่มีผู้ปกครองรายได้สูง และในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 2 ($g = 2$) มีกลุ่มแฝงระดับนักเรียนกลุ่มที่ 1 ($k = 1$) เป็นกลุ่มเด่น โดยสมาชิกส่วนใหญ่เป็นนักเรียนที่มีผู้ปกครองรายได้ต่ำถึงปานกลาง นอกจากนี้หากพิจารณาคะแนนสอบตามกลุ่มแฝง จะพบว่าสมาชิกในกลุ่มแฝง $k = 1$ และ $g = 1$ เป็นกลุ่มที่มีคะแนนเฉลี่ยรวมสูงที่สุด (56.34) รองลงมาคือกลุ่มแฝง $k = 2$ และ $g = 2$ (30.78) กลุ่มแฝง $k = 2$ และ $g = 1$ (18.88) และต่ำที่สุดคือกลุ่มแฝง $k = 1$ และ $g = 2$ (14.68) ดังตาราง 4.7

ตาราง 4.7 จำนวนและค่าสัดส่วนของนักเรียนจำแนกตามตัวแปรและกลุ่มแฝงระดับนักเรียนซึ่งซ้อนอยู่ในกลุ่มแฝงระดับโรงเรียน และข้อมูลทั่วไปของคะแนนสอบจำแนกตามกลุ่มแฝง

ตัวแปรระดับนักเรียน	MMix3PL				MMix2PLE					
	g = 1		g = 2		g = 1		g = 2			
	k = 1	k = 2	k = 1	k = 2	k = 1	k = 2	k = 1	k = 2		
เพศ	ชาย	75 (0.08)	139 (0.15)	568 (0.60)	166 (0.18)	134 (0.14)	106 (0.11)	630 (0.66)	78 (0.08)	
	หญิง	69 (0.04)	267 (0.17)	934 (0.60)	282 (0.18)	147 (0.09)	163 (0.11)	1082 (0.70)	160 (0.10)	
รายได้ผู้ปกครอง	1-50,000 บาท	27 (0.03)	191 (0.20)	601 (0.64)	121 (0.13)	23 (0.02)	29 (0.03)	779 (0.83)	109 (0.12)	
	50,001-100,000 บาท	11 (0.02)	91 (0.14)	472 (0.73)	69 (0.11)	7 (0.01)	21 (0.03)	544 (0.85)	71 (0.11)	
	100,001-200,000 บาท	12 (0.03)	57 (0.15)	250 (0.66)	61 (0.16)	23 (0.06)	39 (0.10)	279 (0.73)	39 (0.10)	
	200,001-300,000 บาท	7 (0.05)	19 (0.14)	61 (0.46)	46 (0.35)	38 (0.29)	38 (0.29)	52 (0.39)	5 (0.04)	
	มากกว่า 300,000 บาท	87 (0.22)	48 (0.12)	118 (0.29)	151 (0.37)	190 (0.47)	142 (0.35)	58 (0.14)	14 (0.03)	
คะแนน	จำนวนฯ (12.5)	เฉลี่ย (SD)	8.42 (3.71)	2.50 (2.17)	2.48 (2.16)	5.76 (3.50)	8.66 (3.49)	2.61 (2.42)	2.51 (2.16)	4.56 (2.53)
	การวัด (5)	เฉลี่ย (SD)	3.98 (1.52)	0.88 (1.41)	0.92 (1.39)	2.66 (2.04)	4.03 (1.54)	1.12 (1.55)	0.91 (1.39)	2.14 (1.94)
	พิชคณิต (40)	เฉลี่ย (SD)	27.26 (10.73)	7.32 (3.78)	7.15 (3.76)	19.67 (9.01)	28.52 (8.98)	9.75 (4.88)	7.03 (3.67)	15.85 (5.78)
	การวิเคราะห์ฯ (22.5)	เฉลี่ย (SD)	13.84 (6.12)	4.41 (3.00)	4.15 (2.88)	10.91 (4.97)	15.13 (4.75)	5.40 (3.56)	4.23 (2.93)	8.22 (3.98)
	รวม (80)	เฉลี่ย (SD)	53.49 (20.05)	15.10 (5.39)	14.68 (5.52)	39.00 (16.16)	56.34 (15.89)	18.88 (7.99)	14.68 (5.44)	30.78 (9.41)

ทั้งนี้เพื่อให้การระบุลักษณะเด่นของสมาชิกในกลุ่มแฝงระดับนักเรียนของโมเดล MMix2PLE มีความชัดเจนขึ้น ผู้วิจัยจะนำผลการจัดกลุ่มแฝงที่ได้มาวิเคราะห์ด้วยโมเดลถดถอยโลจิสติก โดยกำหนดตัวแปรตามคือ การเป็นสมาชิกในกลุ่มแฝงระดับนักเรียน และตัวแปรต้นคือ เพศ และรายได้ผู้ปกครอง ผลการวิเคราะห์พบว่าในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 ($g = 1$) นักเรียนที่มีผู้ปกครองรายได้ 50,001-100,000 บาท และ 100,001-200,000 บาท มีแนวโน้มที่จะอยู่ในกลุ่มแฝงระดับนักเรียนกลุ่มที่ 1 ($k = 1$) น้อยกว่า เมื่อเทียบกับนักเรียนที่มีผู้ปกครองรายได้มากกว่า 300,000 บาท ส่วนในกลุ่มแฝงระดับโรงเรียนกลุ่มที่ 2 ($g = 2$) นักเรียนที่มีผู้ปกครองรายได้ 1-50,000 บาท 50,001-100,000 บาท และ 100,001-200,000 บาท มีแนวโน้มที่จะอยู่ในกลุ่มแฝงระดับนักเรียนกลุ่มที่ 1 ($k = 1$) มากกว่า เมื่อเทียบกับนักเรียนที่มีผู้ปกครองรายได้มากกว่า 300,000 บาท ดังตาราง 4.8

ตาราง 4.8 ค่าประมาณสัมประสิทธิ์ของโมเดลถดถอยโลจิสติก

ตัวแปรระดับนักเรียน	g = 1		g = 2		
	k = 1	k = 2	k = 1	k = 2	
เพศ	หญิง	-0.389 (-0.779, -0.001)	0 ^a	-0.192 (-0.481, 0.097)	0 ^a
	ชาย ^b	-	-	-	-
รายได้ผู้ปกครอง	1-50,000 บาท	-0.538 (-1.130, 0.054)	0 ^a	0.569 (0.009, 1.129)*	0 ^a
	50,001-100,000 บาท	-1.440 (-2.333, -0.553)*	0 ^a	0.642 (0.006, 1.278)*	0 ^a
	100,001-200,000 บาท	-0.846 (-1.406, -0.282)*	0 ^a	0.573 (0.003, 1.143)*	0 ^a
	200,001-300,000 บาท	-0.257 (-0.759, 0.245)	0 ^a	0.937 (-0.151, 2.026)	0 ^a
	มากกว่า 300,000 บาท ^b	-	-	-	-

หมายเหตุ: * หมายถึง มีนัยสำคัญทางสถิติที่ระดับ .05

a, b หมายถึง กำหนดเป็นกลุ่มอ้างอิงเพื่อใช้ในการเปรียบเทียบ

- **ความสามารถ** พบว่าการประมาณค่าพารามิเตอร์ความสามารถระหว่างโมเดล MMix3PL กับ MMix2PLE มีแนวโน้มให้ผลสรุปค่อนข้างสอดคล้องกัน กล่าวคือเมื่อกำหนดให้กลุ่มแฝง $k = 1$ และ $g = 1$ เป็นกลุ่มอ้างอิงซึ่งมีการตรึง (fixed) ค่าความสามารถเฉลี่ยเป็น 0 จะพบว่ากลุ่มแฝงที่มีความสามารถเฉลี่ยรองลงมาคือ กลุ่มแฝง $k = 2$ และ $g = 2$ ($\mu = -0.797$ และ -1.197) กลุ่มแฝง $k = 2$ และ $g = 1$ ($\mu = -2.559$ และ -2.125) และต่ำที่สุดคือกลุ่มแฝง $k = 1$ และ $g = 2$ ($\mu = -2.985$ และ -3.128) ดังตาราง 4.9 ซึ่งผลดังกล่าวค่อนข้างสอดคล้องกับคะแนนรวมเฉลี่ยในกลุ่มแฝง

ทั้งนี้เนื่องจากกลุ่มแฝง $k = 1$ และ $g = 1$ เป็นกลุ่มอ้างอิงที่มีความสามารถเฉลี่ยสูงกว่าเมื่อเทียบกับกลุ่มแฝงอื่นที่เหลือ ค่าความสามารถเฉลี่ยในกลุ่มแฝงที่นำมาเทียบจึงมีค่าน้อยกว่า 0 สำหรับทุกกลุ่ม การแปลผลคุณลักษณะของกลุ่มแฝงที่เหลือจึงมีเพียงกลุ่มความสามารถปานกลางและต่ำเท่านั้น ค่าความสามารถที่ได้จึงไม่ครอบคลุมตามสเกลของคะแนนมาตรฐานซึ่งมีค่าเป็นได้ทั้งบวกและลบ ดังนั้นโดยไม่สูญเสียความหมายทั่วไป ในที่นี้อาจกำหนดให้กลุ่มแฝง $k = 2$ และ $g = 2$ เป็นกลุ่มอ้างอิงใหม่ที่มีความสามารถเฉลี่ยเป็น 0 และพิจารณาใหม่ให้กลุ่มแฝง $k = 1$ และ $g = 1$ คือ “กลุ่มความสามารถสูง” กลุ่มแฝง $k = 2$ และ $g = 2$ คือ “กลุ่มความสามารถปานกลาง” กลุ่มแฝง $k = 2$ และ $g = 1$ คือ “กลุ่มความสามารถต่ำ” และกลุ่มแฝง $k = 1$ และ $g = 2$ คือ “กลุ่มความสามารถต่ำมาก”

ตาราง 4.9 การแจกแจงของพารามิเตอร์ความสามารถในกลุ่มแฝง

โมเดล	กลุ่มแฝงระดับโรงเรียนที่ g	กลุ่มแฝงระดับนักเรียนที่ k	
		$k = 1$	$k = 2$
MMix3PL	$g = 1$	$N(0^*, 3.348^2)$	$N(-2.559, 2.921^2)$
	$g = 2$	$N(-2.985, 3.278^2)$	$N(-0.797, 2.363^2)$
MMix2PLE	$g = 1$	$N(0^*, 1.75^2)$	$N(-2.125, 3.034^2)$
	$g = 2$	$N(-3.128, 3.223^2)$	$N(-1.197, 1.948^2)$

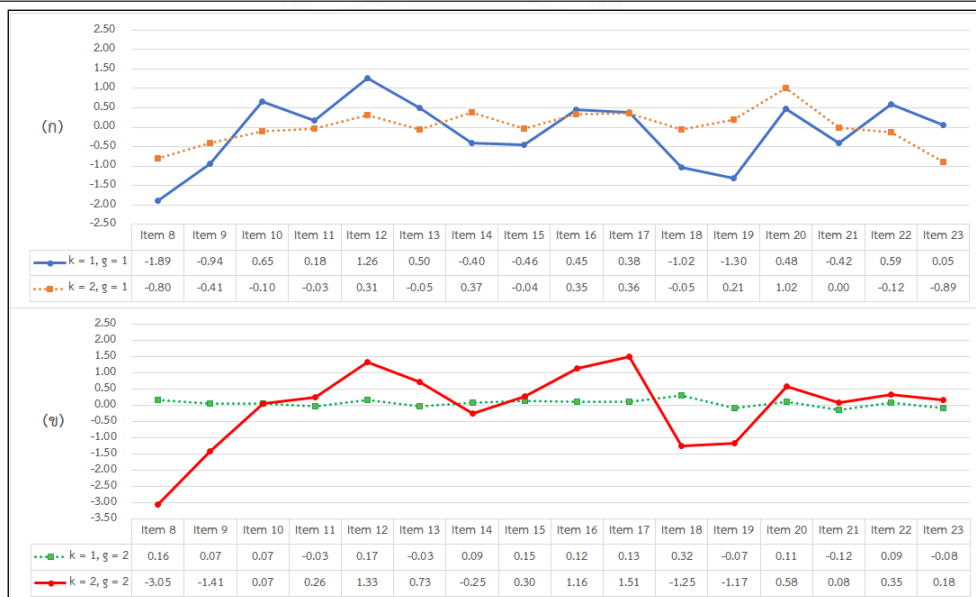
หมายเหตุ: * หมายถึง ความสามารถเฉลี่ยในกลุ่มแฝงถูกตรึงค่าเป็น 0 เพื่อใช้เป็นกลุ่มอ้างอิง

- **ข้อสอบ** พบว่าการประมาณค่าพารามิเตอร์ของข้อสอบจากโมเดล MMix3PL และ MMix2PLE มีแนวโน้มให้ค่าประมาณความยาก (b) ไปในทิศทางเดียวกัน โดยเฉพาะในกลุ่มแฝง $k = 1$ และ $g = 1$ และกลุ่มแฝง $k = 2$ และ $g = 2$ ส่วนในกลุ่มแฝงที่เหลือพบว่ายังให้ผลไม่สอดคล้องกันเท่าใดนัก ดังตาราง 4.10 ซึ่งผลดังกล่าวอาจมาจากการจัดสมาชิกในกลุ่มแฝงของทั้งสองโมเดลที่มีความแตกต่างกัน ผลการตอบข้อสอบของสมาชิกในกลุ่มแฝงบางกลุ่มจึงมีความต่างกันด้วย

ทั้งนี้ผู้วิจัยขอยกตัวอย่างโปรไฟล์ (profile) ความยากของข้อสอบในกลุ่มสาระที่คณิตศาสตร์จากการวิเคราะห์ด้วยโมเดล MMix2PLE ดังแสดงในภาพ 4.5 เนื่องจากสาระนั้นกล่าวมีการวัดด้วยจำนวนข้อสอบมากที่สุด (16 ข้อ) และมีตำแหน่งอยู่ในช่วงกลางแบบสอบ (ข้อ 8-23) ซึ่งน่าจะแสดงถึงความยากของข้อสอบที่มีความผันแปรตามสมาชิกในกลุ่มแฝงที่มีความสามารถสูงและต่ำได้ค่อนข้างชัดเจน

ตาราง 4.10 ค่าประมาณพารามิเตอร์ของข้อสอบจากข้อมูล O-NET วิชาคณิตศาสตร์

สาระ	ข้อ	MMix3PL												MMix2PLE							
		k = 1, g = 1			k = 2, g = 1			k = 1, g = 2			k = 2, g = 2			k = 1, g = 1		k = 2, g = 1		k = 1, g = 2		k = 2, g = 2	
		\hat{b}_{11}	\hat{a}_{11}	\hat{c}_{11}	\hat{b}_{21}	\hat{a}_{21}	\hat{c}_{21}	\hat{b}_{12}	\hat{a}_{12}	\hat{c}_{12}	\hat{b}_{22}	\hat{a}_{22}	\hat{c}_{22}	\hat{b}_{11}	\hat{a}_{11}	\hat{b}_{21}	\hat{a}_{21}	\hat{b}_{12}	\hat{a}_{12}	\hat{b}_{22}	\hat{a}_{22}
1	1	1.47	1.03	0.12	0.13	0.88	0.24	-0.15	0.97	0.26	1.78	1.04	0.12	1.51	1.17	0.99	1.51	0.07	0.20	1.73	1.77
1	2	-0.38	1.32	0.27	-0.02	0.94	0.17	-0.47	1.00	0.20	-0.11	1.07	0.22	-0.29	1.15	0.32	1.02	-0.02	0.87	-0.02	0.73
1	3	-0.82	1.61	0.23	0.02	1.01	0.11	0.46	1.31	0.11	-0.69	0.91	0.21	-0.66	1.16	0.41	0.96	0.11	1.26	-0.59	0.99
1	4	-0.08	1.51	0.20	0.28	1.11	0.10	-0.74	1.13	0.09	0.21	1.59	0.17	-0.21	1.33	0.77	1.55	0.15	1.26	0.59	1.55
1	5	-0.23	0.44	0.30	0.04	0.67	0.29	-0.74	0.73	0.26	-0.58	0.35	0.31	-0.10	0.36	-0.04	0.43	-0.09	0.19	-0.19	0.14
2	6	-0.51	0.93	0.27	0.04	0.97	0.16	-0.83	0.80	0.13	-0.84	0.88	0.18	-0.82	0.58	0.27	1.23	0.05	1.17	-1.25	1.29
2	7	-1.24	0.80	0.30	-0.12	0.91	0.17	-0.54	0.76	0.18	-0.55	1.35	0.22	-0.70	0.87	-0.38	0.95	-0.09	0.99	-0.68	0.75
3	8	-1.69	0.77	0.33	-0.10	0.80	0.19	0.17	0.83	0.19	-2.19	0.92	0.33	-1.89	0.62	-0.80	0.39	0.16	0.97	-3.05	0.94
3	9	-0.39	1.19	0.29	-0.03	0.93	0.21	-0.01	0.94	0.21	-1.13	1.38	0.21	-0.94	1.00	-0.41	0.79	0.07	0.72	-1.41	0.95
3	10	0.91	0.90	0.20	0.09	1.04	0.11	0.47	0.91	0.10	0.46	0.80	0.23	0.65	0.74	-0.10	0.90	0.07	1.26	0.07	1.21
3	11	-0.13	0.75	0.28	0.06	0.97	0.21	-0.41	0.79	0.21	0.41	0.92	0.35	0.18	0.86	-0.03	0.36	-0.03	0.97	0.26	0.27
3	12	1.37	1.14	0.10	0.19	1.10	0.11	0.31	1.26	0.13	1.06	0.87	0.09	1.26	1.06	0.31	1.60	0.17	1.26	1.33	1.69
3	13	0.25	0.99	0.22	-0.04	0.91	0.22	-0.21	0.66	0.18	0.58	0.93	0.20	0.50	1.02	-0.05	0.86	-0.03	0.73	0.73	0.73
3	14	-0.37	1.56	0.21	-0.17	0.92	0.16	-0.57	0.99	0.15	-0.20	1.14	0.14	-0.40	0.86	0.37	1.29	0.09	1.18	-0.25	1.55
3	15	-0.54	0.81	0.30	0.04	0.94	0.19	0.28	1.07	0.20	-0.26	1.02	0.21	-0.46	0.77	-0.04	0.69	0.15	1.07	0.30	0.64
3	16	0.34	1.41	0.20	0.10	0.98	0.19	-0.33	1.07	0.15	0.93	1.59	0.21	0.45	1.61	0.35	1.33	0.12	1.14	1.16	1.25
3	17	0.26	2.23	0.12	0.25	1.18	0.11	0.73	1.07	0.11	0.73	1.59	0.11	0.38	2.10	0.36	1.70	0.13	1.26	1.51	1.74
3	18	-0.67	1.44	0.27	0.10	1.14	0.07	0.28	1.29	0.07	-1.23	1.37	0.16	-1.02	1.30	-0.05	1.64	0.32	1.28	-1.25	1.52
3	19	-0.85	1.42	0.27	0.01	0.98	0.20	-0.34	1.26	0.20	-1.29	1.09	0.22	-1.30	1.01	0.21	1.07	-0.07	1.03	-1.17	0.66
3	20	0.42	1.57	0.11	0.08	1.13	0.07	0.26	1.11	0.09	0.73	1.67	0.14	0.48	1.81	1.02	1.47	0.11	1.27	0.58	1.82
3	21	-0.46	0.57	0.29	-0.04	0.76	0.23	-0.28	0.76	0.25	-0.37	0.54	0.30	-0.42	0.41	0.00	0.44	-0.12	0.25	0.08	0.25
3	22	0.33	0.48	0.23	0.01	0.83	0.23	0.50	1.10	0.21	0.37	0.81	0.27	0.59	0.60	-0.12	0.46	0.09	0.72	0.35	0.49
3	23	-0.30	1.02	0.30	-0.10	0.82	0.23	0.53	1.11	0.20	-0.16	0.42	0.31	0.05	0.78	-0.89	0.46	-0.08	0.95	0.18	0.25
4	24	2.47	0.34	0.13	0.24	1.06	0.20	0.28	0.87	0.18	0.63	0.69	0.13	2.46	0.34	0.53	1.27	0.08	1.10	0.30	1.47
4	25	-0.99	0.95	0.28	-0.07	0.90	0.22	0.12	1.06	0.23	-1.00	1.12	0.35	-1.44	0.75	-1.18	0.74	-0.03	0.25	-0.51	0.52
4	26	0.98	1.22	0.19	0.18	1.09	0.17	0.58	1.25	0.18	1.42	1.01	0.19	1.11	1.14	0.61	1.30	0.12	1.14	0.49	1.64
4	27	0.57	0.46	0.19	0.04	1.01	0.15	0.65	1.31	0.15	0.98	0.76	0.24	1.16	0.64	-0.78	0.79	0.11	1.20	0.39	0.83
4	28	1.75	1.23	0.10	0.30	1.11	0.17	0.14	1.02	0.15	1.30	1.07	0.09	1.44	1.25	1.22	1.62	0.07	1.22	1.15	1.66
4	29	-1.09	1.07	0.21	-0.03	1.00	0.18	0.37	1.00	0.17	-1.45	0.76	0.28	-1.14	0.72	-0.63	0.74	0.08	1.13	-1.31	0.66
4	30	-0.40	1.38	0.27	-0.32	0.80	0.19	0.20	1.10	0.20	-0.68	0.79	0.34	-0.82	0.74	-1.18	0.63	0.07	0.95	0.28	0.38
4	31	-0.39	0.51	0.25	0.00	0.83	0.21	-0.33	0.71	0.17	-0.38	0.45	0.31	-0.47	0.34	-0.02	0.52	0.03	1.07	0.12	0.31
4	32	0.79	0.67	0.17	0.10	1.07	0.15	0.71	1.29	0.15	0.97	1.14	0.20	0.91	0.83	0.75	1.32	0.11	1.21	0.62	1.42



ภาพ 4.5 โปรไฟล์ความยากของข้อสอบกลุ่มสาระพีชคณิตในกลุ่มแฝงระดับนักเรียนซึ่งซ่อนอยู่ใน:
 (ก) กลุ่มแฝงระดับโรงเรียนกลุ่มที่ 1 (g = 1) และ (ข) กลุ่มแฝงระดับโรงเรียนกลุ่มที่ 2 (g = 2)

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยนี้มีวัตถุประสงค์ 4 ประการ ได้แก่ 1) เพื่อศึกษารูปแบบของโมเดล MMix2PLE และวิธีการประมาณค่าพารามิเตอร์ 2) เพื่อเปรียบเทียบความถูกต้องของการคัดเลือกโมเดลระหว่างเกณฑ์สารสนเทศ AIC, BIC, CAIC และ DBIC สำหรับโมเดล MMix3PL และ MMix2PLE ภายใต้สถานการณ์จำลอง 3) เพื่อศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง ระหว่างโมเดล MMix3PL กับ MMix2PLE ภายใต้สถานการณ์จำลอง และ 4) เพื่อวิเคราะห์หาจำนวนกลุ่มแฝง พารามิเตอร์กลุ่มแฝง พารามิเตอร์ของข้อสอบ และความสามารถเฉลี่ยในกลุ่มแฝง จากข้อมูลเชิงประจักษ์ด้วยโมเดล MMix3PL และ MMix2PLE ทั้งนี้ผู้วิจัยขอสรุปผลการวิจัย อภิปรายผล และเสนอข้อเสนอแนะดังนี้

5.1 สรุปผลการวิจัย

1) สรุปผลการวิจัยตามวัตถุประสงค์ข้อ 1

รูปแบบของโมเดล MMix2PLE ที่มีการนำคุณลักษณะของผู้สอบและข้อสอบเข้ามาช่วยอธิบายโอกาสในการเดาถูก (g) พบว่ามีแนวโน้มสามารถช่วยหลีกเลี่ยงปัญหาที่อาจเกิดขึ้นจากการประมาณค่าพารามิเตอร์การเดา (c) ของโมเดล MMix3PL ได้ เนื่องจากโอกาสในการเดาถูก (g) ของโมเดล MMix2PLE ถูกนิยามให้มีความแปรผันตามพารามิเตอร์ของผู้สอบและข้อสอบ g จึงไม่ใช่พารามิเตอร์ที่ต้องประมาณค่าในโมเดล แตกต่างจากโอกาสในการเดาถูก (c) ของโมเดล MMix3PL ซึ่งถูกนิยามให้เป็นพารามิเตอร์ของข้อสอบที่ต้องประมาณค่า ส่วนวิธีการประมาณค่าแบบเบส์ที่นำมาใช้ พบว่ามีแนวโน้มให้ค่าประมาณพารามิเตอร์ย้อนกลับได้ค่อนข้างแม่นยำ สัมพันธ์กับค่าจริง สามารถระบุค่าที่แน่นอนได้เพียงหนึ่งเดียว ดังสังเกตได้จากกราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังของค่าประมาณพารามิเตอร์ซึ่งมีแนวโน้มเป็นเอกฐานนิยม รวมถึงยังมีแนวโน้มมีความแกร่งในการประมาณค่าแม้ว่าขนาดตัวอย่างในกลุ่มแฝงจะมีขนาดเล็ก

2) สรุปผลการวิจัยตามวัตถุประสงค์ข้อ 2

การเปรียบเทียบความถูกต้องของการคัดเลือกโมเดลภายใต้สถานการณ์จำลอง พบว่า AIC มีแนวโน้มให้ความถูกต้องของการคัดเลือกโมเดลสูงที่สุด เนื่องจาก AIC มีความคงเส้นคงวาในการเลือกโมเดลได้ตรงกับโมเดลประชากรมากกว่าเกณฑ์สารสนเทศตัวอื่น ๆ ทั้งกรณี $N = 1,250$ และ $2,500$ สำหรับโมเดล MMix3PL และ MMix2PLE รองลงมาคือ DBIC ซึ่งพบว่ามีแนวโน้มให้ความถูกต้องของการคัดเลือกโมเดลสูงขึ้น เมื่อ N เพิ่มขึ้นจาก $1,250$ เป็น $2,500$ ทั้งกรณีโมเดล MMix3PL และ

MMix2PLE ส่วน BIC และ CAIC พบว่ามีแนวโน้มให้ความถูกต้องของการคัดเลือกโมเดลใกล้เคียงกัน โดยสามารถเลือกโมเดลได้ถูกต้องเฉพาะกรณี $N = 2,500$ สำหรับโมเดล MMix2PLE

จากผลดังกล่าวจึงอาจสรุปได้ว่าเกณฑ์สารสนเทศ AIC มีความถูกต้องของการคัดเลือกโมเดลสูงที่สุด รองลงมาคือ DBIC และต่ำที่สุดคือ BIC และ CAIC สำหรับโมเดล MMix3PL และ MMix2PLE ภายใต้เงื่อนไขการศึกษาข้อมูลทุกระดับที่มี 2 ระดับ ขนาดตัวอย่างเป็น 1,250 และ 2,500 และมีจำนวนกลุ่มแฝงระดับโรงเรียน 2 กลุ่ม และกลุ่มแฝงระดับนักเรียน 2 กลุ่ม

3) สรุปผลการวิจัยตามวัตถุประสงค์ข้อ 3

การศึกษาความแม่นยำของการประมาณค่าย้อนกลับ พบว่าโมเดล MMix3PL และ MMix2PLE มีความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝงระดับโรงเรียน (π_g) สูงร้อยละ 100.0 ในขณะที่กลุ่มแฝงระดับนักเรียน (π_{kg}) มีค่าสูงร้อยละ 96.02 ถึง 97.71 และ 98.92 ถึง 99.25 ตามลำดับ ส่วนพารามิเตอร์ความสามารถเฉลี่ย พบว่าโมเดล MMix2PLE (bias=-.117 ถึง -.084, RMSE=.082 ถึง .097) และโมเดล MMix3PL (bias=-.193 ถึง -.156, RMSE=.135) มีแนวโน้มให้ความแม่นยำของการประมาณค่าพารามิเตอร์ความสามารถเฉลี่ย (μ) ค่อนข้างสูงใกล้เคียงกัน ส่วนพารามิเตอร์ของข้อสอบคือ พารามิเตอร์อำนาจจำแนก (a) และความยาก (b) พบว่าโมเดล MMix2PLE (bias=-.039 ถึง .000, RMSE=.037 ถึง .114, corr=.667 ถึง .991) และโมเดล MMix3PL (bias=-.154 ถึง .000, RMSE=.053 ถึง .115, corr=.772 ถึง .987) ก็ยังมีแนวโน้มให้ความแม่นยำของการประมาณค่าค่อนข้างสูงใกล้เคียงกันด้วย อย่างไรก็ตามพารามิเตอร์การเดา (c) ของโมเดล MMix3PL (bias=.025 ถึง .027, RMSE=.006 ถึง .007, corr=-.290 ถึง -.186) พบว่ายังค่อนข้างมีปัญหาในการประมาณค่าพารามิเตอร์ย้อนกลับ เนื่องจากค่าประมาณที่ได้แทบจะไม่มีความสัมพันธ์กับค่าจริงแม้แต่น้อย (corr น้อยกว่า 0)

จากผลดังกล่าวจึงอาจสรุปได้ว่าโมเดล MMix2PLE และ MMix3PL มีแนวโน้มให้ความแม่นยำของการประมาณค่าพารามิเตอร์ส่วนใหญ่ค่อนข้างสูงใกล้เคียงกัน ยกเว้นการประมาณค่าพารามิเตอร์การเดาของโมเดล MMix3PL ที่มีแนวโน้มพบปัญหาได้ค่าประมาณไม่สัมพันธ์กับค่าจริง ภายใต้สถานการณ์จำลองที่โมเดลวิเคราะห์เป็นโมเดลเดียวกับโมเดลประชากร ในเงื่อนไขการศึกษาที่ข้อมูลมี 2 ระดับ มีขนาดตัวอย่างเป็น 1,250 และ 2,500 และมีจำนวนกลุ่มแฝงระดับโรงเรียน 2 กลุ่ม และกลุ่มแฝงระดับนักเรียน 2 กลุ่ม

4) สรุปผลการวิจัยตามวัตถุประสงค์ข้อ 4

การคัดเลือกโมเดลเพื่อหาจำนวนกลุ่มแฝงของข้อมูลเชิงประจักษ์สำหรับโมเดล MMix3PL และ MMix2PLE พบว่าเกณฑ์สารสนเทศ AIC, BIC, CAIC และ DBIC เลือกโมเดลที่มีจำนวนกลุ่มแฝงระดับนักเรียน (K) 2 กลุ่ม และระดับโรงเรียน (G) 2 กลุ่ม เป็นโมเดลที่มีความสอดคล้องกับข้อมูลเชิงประจักษ์ โดยพบแนวโน้มว่าโมเดล MMix3PL มีความสอดคล้องกับข้อมูลมากกว่าโมเดล MMix2PLE

อย่างไรก็ตามเมื่อพิจารณาจากผลการจำแนกกลุ่มแฝงแล้ว โมเดล MMix2PLE สามารถระบุลักษณะเด่นของสมาชิกในกลุ่มแฝงได้ชัดเจนกว่าโมเดล MMix3PL โดยพบว่ากลุ่มแฝงของข้อมูลอาจแบ่งได้เป็น 4 กลุ่ม ได้แก่ 1) กลุ่ม $k = 1$ และ $g = 1$ หรือ “กลุ่มความสามารถสูง” 2) กลุ่ม $k = 2$ และ $g = 1$ หรือ “กลุ่มความสามารถต่ำ” 3) กลุ่ม $k = 2$ และ $g = 2$ หรือ “กลุ่มความสามารถปานกลาง” และ 4) กลุ่ม $k = 1$ และ $g = 2$ หรือ “กลุ่มความสามารถต่ำมาก” ทั้งนี้กลุ่มแฝงสองกลุ่มแรกที่กล่าวมา มีลักษณะเด่นคือสมาชิกส่วนใหญ่มีแนวโน้มเป็นนักเรียนที่มีผู้ปกครองรายได้มากกว่า 300,000 บาท และศึกษาอยู่ในโรงเรียนสังกัดสำนักงานคณะกรรมการการอุดมศึกษา (อว.) ในกรุงเทพฯ ส่วนกลุ่มแฝงสองกลุ่มหลังมีลักษณะเด่นคือสมาชิกส่วนใหญ่มีแนวโน้มเป็นนักเรียนที่มีผู้ปกครองรายได้ไม่เกิน 300,000 บาท และศึกษาอยู่ในโรงเรียนซึ่งมิใช่สังกัด อว. และไม่ได้อยู่ในกรุงเทพฯ

ส่วนการประมาณค่าพารามิเตอร์ของข้อสอบในกลุ่มแฝง ในภาพรวมพบว่าโมเดล MMix3PL และ MMix2PLE มีแนวโน้มให้ค่าประมาณพารามิเตอร์ส่วนใหญ่ไปในทิศทางเดียวกัน อย่างไรก็ตามค่าประมาณพารามิเตอร์ของข้อสอบในบางกลุ่มพบว่ายังไม่มีความสอดคล้องกันเท่าที่ควร ซึ่งอาจเป็นผลมาจากการจัดสมาชิกในกลุ่มแฝงของทั้งสองโมเดลที่มีความแตกต่างกัน

5.2 อภิปรายผลการวิจัย

จากผลสรุปการวิจัยข้างต้น ผู้วิจัยขอแบ่งการอภิปรายผลเป็น 4 ประเด็นดังนี้

1) ขั้นตอนวิธีการประมาณค่าพารามิเตอร์แบบเบย์สำหรับโมเดล MMix2PLE ในงานวิจัยนี้ ผู้วิจัยได้ประยุกต์แนวคิดมาจากโมเดล MMixRasch ของ Cho และ Cohen (2010) โดยใช้วิธีการประมาณค่าแบบเบย์ด้วยขั้นตอนวิธี MCMC และกำหนดการแจกแจงก่อนหน้าของพารามิเตอร์กลุ่มแฝงเป็นการแจกแจงตรีเคิลเหมือนกัน เนื่องจากแนวทางแบบเบย์ไม่มีข้อตกลงเกี่ยวกับความเป็นปกติของข้อมูล จึงสามารถหลีกเลี่ยงปัญหาที่อาจเกิดขึ้นจากการฝ่าฝืนข้อตกลงดังกล่าวในการประมาณค่าพารามิเตอร์กลุ่มแฝง และอาจช่วยบรรเทาปัญหาความไม่แน่นอน (uncertainty) ในการหาจำนวนกลุ่มแฝงที่เหมาะสมของข้อมูลลงได้ (Cho, 2007) อย่างไรก็ตามงานวิจัยนี้ได้ใช้โมเดลการวัดเป็นโมเดล IRT ทางเลือกที่คำนึงถึงการเดา (โมเดล 2PLE) เพื่อต้องการนำโมเดลไปประยุกต์ใช้กับข้อมูลทางการศึกษาที่มักมีภาวะวิวิธพันธุ์และมีการเดาเข้ามาเกี่ยวข้อง รวมถึงต้องการหลีกเลี่ยงปัญหาที่อาจเกิดขึ้นจากการประมาณค่าพารามิเตอร์การเดาและการแปลผลการเดาจากโมเดล 3PL ในขณะที่ Cho และ Cohen (2010) ใช้โมเดลการวัดเป็นโมเดล IRT แบบดั้งเดิม (โมเดล Rasch) ทั้งนี้แม้โมเดลการวัดที่นำมาประยุกต์ใช้จะมีรูปแบบที่แตกต่างกัน แต่การประมาณค่าพารามิเตอร์ก็ยังคงให้ผลใกล้เคียงกัน สังเกตได้จากความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝงในงานวิจัยนี้ที่มีค่าร้อยละ 96.0 ถึง 100.0 ซึ่งใกล้เคียงกับ Cho และ Cohen (2010) 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

นอกจากนี้ขั้นตอนวิธีการประมาณค่าพารามิเตอร์ที่ใช้ยังพบว่าสามารถประมาณค่าพารามิเตอร์ย้อนกลับได้ค่อนข้างดี ระบุค่าที่แน่นอนได้เพียงหนึ่งเดียว ดังสังเกตได้จากกราฟความหนาแน่นของการแจกแจงความน่าจะเป็นภายหลังที่มีแนวโน้มเป็นเอกฐานนิยม รวมถึงยังให้ความแม่นยำของการประมาณค่าพารามิเตอร์ค่อนข้างดี แม้ว่าค่าสัดส่วนของสมาชิกในกลุ่มแฝงจะมีเพียงร้อยละ 10 ก็ตาม ซึ่งผลดังกล่าวอาจเป็นเพราะวิธีการประมาณค่าแบบเบย์มีความแกร่งในการประมาณค่าเมื่อข้อมูลมีการสูญหายหรือขนาดตัวอย่างในกลุ่มแฝงมีขนาดเล็ก (Frühwirth-Schnatter, 2006) อย่างไรก็ตามขั้นตอนวิธีการประมาณค่าที่ใช้ในงานวิจัยนี้ยังคงมีข้อจำกัดเกี่ยวกับประเภทของข้อมูลที่ต้องเป็นแบบทวิภาค และจำนวนมิติของความสามารถแฝงที่มุ่งวัดที่ต้องเป็นแบบเอกมิติเท่านั้น

2) การเปรียบเทียบความถูกต้องของการคัดเลือกโมเดลภายใต้สถานการณ์จำลอง พบว่า AIC มีแนวโน้มมีความถูกต้องของการคัดเลือกโมเดลสูงที่สุด ทั้งกรณี $N = 1,250$ และ $2,500$ สำหรับโมเดล MMix3PL และ MMix2PLE รองลงมาคือ DBIC ที่มีความถูกต้องของการคัดเลือกโมเดลสูงขึ้น เมื่อ N เพิ่มขึ้นสำหรับทั้งสองโมเดล และต่ำที่สุดคือ BIC และ CAIC ที่มีความถูกต้องของการคัดเลือกโมเดล เฉพาะกรณี $N = 2,500$ สำหรับโมเดล MMix2PLE ซึ่งผลดังกล่าวขัดแย้งกับสมมติฐานที่ตั้งไว้ว่า “BIC, CAIC และ DBIC มีความถูกต้องของการคัดเลือกโมเดลมากกว่า AIC สำหรับทั้งสองโมเดล” ทั้งนี้อาจเนื่องมาจากโมเดล MMix3PL และ MMix2PLE ที่นำมาศึกษาเป็นโมเดลที่มีโครงสร้างซับซ้อนมากกว่าโมเดลการวัดทั่วไป ซึ่งโมเดลลักษณะดังกล่าว AIC จะมีแนวโน้มให้ความถูกต้องของการคัดเลือกโมเดลสูงกว่าเกณฑ์สารสนเทศตัวอื่น แต่หากเป็นโมเดลที่มีโครงสร้างไม่ซับซ้อน BIC และ CAIC จะมีแนวโน้มให้ความถูกต้องของการคัดเลือกโมเดลได้สูงกว่า (Sen et al., 2019; Shao, 1997) ส่วน DBIC เป็นเกณฑ์สารสนเทศที่มีการปรับปรุงมาจาก BIC ให้มีความถูกต้องของการคัดเลือกโมเดลได้ดีขึ้นเมื่อตัวอย่างมีขนาดเล็กถึงปานกลาง (Yang & Yang, 2007) DBIC จึงมีแนวโน้มมีความถูกต้องของการคัดเลือกโมเดลสูงกว่า BIC ในหลายสถานการณ์การศึกษา นอกจากนี้ผลการวิเคราะห์ยังสอดคล้องกับ Choi (2014) ที่ศึกษาการคัดเลือกโมเดลสำหรับโมเดล MixIRT (MixRasch, Mix2PL, Mix3PL) โดยพบว่า BIC มีแนวโน้มให้ความถูกต้องของการคัดเลือกโมเดลได้สูงกว่า หากเป็นโมเดล MixRasch หรือ Mix2PL ในขณะที่ AIC มีแนวโน้มให้ความถูกต้องของการคัดเลือกโมเดลได้สูงกว่าหากเป็นโมเดล Mix3PL ซึ่งข้อสรุปดังกล่าวน่าจะอนุมานถึงโมเดล MMix3PL และ MMix2PLE ได้เช่นกัน

3) การศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์ย้อนกลับ พบว่าโมเดล MMix3PL และ MMix2PLE มีแนวโน้มมีความแม่นยำของการประมาณค่าพารามิเตอร์ของข้อสอบสูงใกล้เคียงกัน ภายใต้สถานการณ์จำลองที่โมเดลวิเคราะห์เป็นโมเดลเดียวกับโมเดลประชากร และการเพิ่มขนาดตัวอย่างยังมีส่วนช่วยทำให้ความแม่นยำสูงขึ้นด้วย ซึ่งผลดังกล่าวสอดคล้องกับสมมติฐานที่ตั้งไว้ว่า

“โมเดล *MMix3PL* และ *MMix2PLE* มีความแม่นยำของการประมาณค่าพารามิเตอร์ของข้อสอบสูงใกล้เคียงกัน และการเพิ่มขนาดตัวอย่างน่าจะช่วยให้ความแม่นยำสูงขึ้น ส่วนความแม่นยำของการประมาณค่าพารามิเตอร์กลุ่มแฝงและความสามารถเฉลี่ยทั้งสองโมเดลน่าจะไม่แตกต่างกัน” อย่างไรก็ตามในภาพรวมอาจสังเกตได้ว่าโมเดล *MMix2PLE* มีแนวโน้มให้ค่าความคลาดเคลื่อน (bias และ RMSE) ค่อนข้างต่ำกว่าโมเดล *MMix3PL* ภายใต้สถานการณ์ดังกล่าว ซึ่งอาจเป็นผลมาจากการนิยามโอกาสในการเดาถูก (g) ในโมเดล *MMix2PLE* ให้มีค่าผันแปรตามคุณลักษณะของผู้สอบและข้อสอบ g จึงไม่ใช่พารามิเตอร์ที่ต้องประมาณค่า แตกต่างจาก c ในโมเดล *MMix3PL* ซึ่งเป็นพารามิเตอร์ของข้อสอบที่ต้องประมาณค่า ดังนั้นในแง่ของการประมาณค่าพารามิเตอร์ โมเดล *MMix3PL* จึงมีความซับซ้อนหรือมีจำนวนพารามิเตอร์ที่ต้องประมาณค่ามากกว่าโมเดล *MMix2PLE* และอาจส่งผลทำให้ความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์ในภาพรวมมีแนวโน้มสูงกว่า (Sen, 2014)

ส่วนการประมาณค่าพารามิเตอร์ c ในโมเดล *MMix3PL* ที่มีค่าสหสัมพันธ์เป็นลบ พบว่าผลการวิเคราะห์นี้มีความสอดคล้องกับงานวิจัยของ Choi (2014) ซึ่งทำการศึกษาโมเดล *Mix3PL* และได้ค่าสหสัมพันธ์ระหว่างค่าจริงกับค่าประมาณของพารามิเตอร์ c เป็นลบเช่นกัน ทั้งนี้ผลดังกล่าวอาจเป็นเพราะพารามิเตอร์ c คือค่าเส้นกำกับล่างต่ำสุด (lower asymptote) ของโค้งลักษณะข้อสอบ ซึ่งแสดงถึงความน่าจะเป็นที่ผู้สอบความสามารถต่ำมาก (θ เข้าใกล้ $-\infty$) จะตอบข้อสอบได้ถูกต้อง แต่เนื่องจากการศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์ในเชิงจำลองทั่วไป นิยมสุ่มพารามิเตอร์ความสามารถของผู้สอบมาจากการแจกแจงปกติมาตรฐานหรือ $N(0, 1)$ ความสามารถของผู้สอบส่วนใหญ่จึงอยู่ในระดับปานกลาง และมีเพียงบางส่วนที่อยู่ในระดับต่ำมากหรือสูงมาก ส่งผลทำให้ผู้สอบที่มีคุณลักษณะตามนิยามของ c มีจำนวนน้อยมาก ค่าประมาณ c ที่ได้จากข้อมูลจำลองจึงน่าจะมีแนวโน้มมีความคลาดเคลื่อนมาตรฐานสูง และไม่สัมพันธ์กับค่าจริง (Lee & Bolt, 2018)

4) การคัดเลือกโมเดลสำหรับข้อมูลเชิงประจักษ์ พบว่าโมเดล *MMix3PL* มีความสอดคล้องกับข้อมูลคะแนนสอบ O-NET มากกว่าโมเดล *MMix2PLE* ทั้งนี้ผลดังกล่าวอาจเป็นเพราะข้อมูลที่นำมาสาธิตเป็นข้อมูลผลการตอบข้อสอบจากวิชาคณิตศาสตร์ ซึ่งมีลักษณะเด่นคือผู้ตอบที่มีความรู้สมบูรณ์มักตอบข้อสอบได้ทั้งหมด แต่หากไม่มีความรู้หรือมีความรู้เพียงบางส่วนจะตอบข้อสอบไม่ได้เลย การเดาถูกในวิชาคณิตศาสตร์จึงมักเป็นการเดาถูกแบบสุ่ม ซึ่งโอกาสในการเดาถูกไม่มีความเกี่ยวข้องกับความสามารถของผู้สอบ (San Martin et al., 2006) จากผลดังกล่าวจึงอาจทำให้โมเดล *MMix3PL* ซึ่งนิยามโอกาสในการเดาถูก (c) ว่าเป็นพารามิเตอร์ของข้อสอบที่มีได้แปรผันตามความสามารถของผู้สอบ มีความสอดคล้องกับข้อมูลมากกว่าโมเดล *MMix2PLE* ซึ่งนิยามโอกาสในการเดาถูก (g) ว่าเป็นค่าที่แปรผันตามความสามารถของผู้สอบ และจากเงื่อนไขดังกล่าวผู้วิจัยจึงคาดการณ์ว่าโมเดล *MMix2PLE* น่าจะมีความสอดคล้องกับข้อมูลผลการตอบข้อสอบในรายวิชาที่ผู้สอบสามารถใช้ความรู้เข้ามาช่วยในการเดาได้มากกว่า เช่น ข้อสอบวิชาทางภาษา เนื่องจากตัวเลือกของข้อสอบในวิชา

ดังกล่าวมักมีลักษณะเป็นข้อความ ซึ่งผู้ตอบที่มีความรู้ไม่เพียงพอในการตอบข้อคำถามสามารถอ่านข้อความในตัวเลือกและใช้ความรู้ที่มีตัดตัวเลือกที่ไม่ถูกต้องออกไปก่อน แล้วจึงเดาตอบแบบสุ่มจากตัวเลือกที่เหลือ โอกาสในการเดาถูกในวิชาทางภาษาจึงมักมีค่าแปรผันตามความสามารถของผู้สอบ

นอกจากนี้ผลการวิเคราะห์กลุ่มแฝง ยังพบว่ากลุ่มที่มีความสามารถสูงมักเป็นนักเรียนที่มีฐานะทางบ้านดี ศึกษาอยู่ในโรงเรียนสังกัดสำนักงานคณะกรรมการการอุดมศึกษา (อว.) และอยู่ในกรุงเทพฯ ขณะที่กลุ่มความสามารถต่ำส่วนใหญ่เป็นนักเรียนที่มีฐานะทางบ้านปานกลางถึงต่ำ ศึกษาอยู่ในโรงเรียนที่มีใช้สังกัด อว. และไม่ได้อยู่ในกรุงเทพฯ ทั้งนี้ผลการวิเคราะห์ดังกล่าวอาจเป็นหลักฐานหนึ่ง que แสดงถึงความเหลื่อมล้ำทางการศึกษาในประเทศไทย สอดคล้องกับงานวิจัยของสังวรณี ังคระโทก (2552) ที่พบว่าผลสัมฤทธิ์ทางการเรียนของนักเรียนมีความแตกต่างกันตามที่ตั้งของสถานศึกษา ขนาดโรงเรียน และฐานะของผู้ปกครอง และงานวิจัยของ ณีฎฐภรณ์ หลาวทอง และคณะ (2563) ที่พบว่าที่ตั้งและสังกัดของโรงเรียน และเศรษฐกิจและสังคมของครอบครัวของนักเรียน เป็นตัวแปรสำคัญที่มีผลต่อคุณภาพการศึกษาของนักเรียน และยังสอดคล้องกับงานวิจัยของ Anderson และ Anderson (2017) ซึ่งพบว่ากระบวนการจัดการสอนของโรงเรียนส่งผลต่อผลสัมฤทธิ์ทางการเรียนของนักเรียนแตกต่างกันตามภูมิหลังทางเศรษฐกิจของผู้ปกครอง

5.3 ข้อเสนอแนะ

1) ข้อเสนอแนะในการนำไปใช้

1.1 นักวัดและประเมินผลการศึกษาสามารถนำรูปแบบและขั้นตอนวิธีการประมาณค่าของโมเดล MMix2PLE ในงานวิจัยนี้ไปเป็นแนวทางในการประยุกต์ใช้กับโมเดลรูปแบบอื่น ๆ เพื่อปรับปรุงโมเดลการวัดให้มีความซับซ้อนสอดคล้องตามสภาพธรรมชาติของข้อมูล หรือเพื่อหลีกเลี่ยงปัญหาที่อาจเกิดขึ้นจากการประมาณค่าพารามิเตอร์ในโมเดล และผ่อนคลายข้อตกลงบางประการของโมเดลที่มักมีการฝ่าฝืนธรรมชาติของข้อมูล

1.2 จากผลการศึกษาข้อมูลเชิงประจักษ์ทำให้ทราบว่าโมเดล MMix2PLE ที่ประยุกต์ขึ้นในงานวิจัยนี้ สามารถใช้เป็นโมเดลเพื่อวิเคราะห์ถึงความเหลื่อมล้ำทางการศึกษาได้เช่นกัน ซึ่งโมเดลดังกล่าวมีจุดเด่นที่เหนือกว่าโมเดลพระระดับทั่วไป เนื่องจากมีการคำนึงถึงภาวะวิวิธพันธุ์ของประชากร ซึ่งเป็นลักษณะธรรมชาติของข้อมูล รวมถึงมีการวัดความสามารถของผู้สอบจากค่าความสามารถแฝง ซึ่งมีแนวโน้มให้ผลการวัดที่มีความถูกต้องและน่าเชื่อถือกว่าการวัดจากคะแนนสอบ ดังนั้นหน่วยงานทางการศึกษาหรือสถาบันการศึกษาจึงสามารถใช้โมเดล MMix2PLE เพื่อวิเคราะห์หากกลุ่มแฝงและคุณลักษณะแฝงของโรงเรียนหรือนักเรียน ซึ่งอาจได้สารสนเทศที่จะนำไปใช้ในการพัฒนาหลักสูตรหรือนำไปวินิจฉัยเพื่อยกระดับการศึกษาให้มีคุณภาพ เพิ่มความเท่าเทียมกันทางการศึกษาให้กับโรงเรียนแต่ละกลุ่ม และวางแผนให้นักเรียนมีผลสัมฤทธิ์ทางการเรียนที่สูงขึ้น

1.3 โมเดลในงานวิจัยนี้ถูกพัฒนาด้วยโปรแกรม MultiBUGS เวอร์ชัน 1.0.0 ซึ่งสามารถกำหนดจำนวนแกนประมวลผลเพื่อการคำนวณแบบคู่ขนาน (parallel computing) ได้หลายแกน (multiple cores) แตกต่างจากโปรแกรมแบบเบสส์ทั่วไป เช่น WinBUGS หรือ OpenBUGS ที่ต้องรันบนแกนประมวลผลไว้เพียงแกนเดียว (single core) ดังนั้นโปรแกรม MultiBUGS จึงมีประสิทธิภาพด้านความเร็วเหนือกว่าโปรแกรมแบบเบสส์ทั่วไปค่อนข้างมาก โดยจากการบันทึกเวลาที่ใช้ในการประมาณค่าพารามิเตอร์ย้อนกลับต่อหนึ่งชุดข้อมูลจำลอง เมื่อกำหนดแกนประมวลผลไว้ที่ 10 แกน พบว่าโมเดล MMix3PL ใช้เวลาประมาณ 3-5 ชั่วโมง และ 7-8 ชั่วโมง สำหรับกรณี $N = 1,250$ และ $2,500$ ตามลำดับ ส่วนโมเดล MMix2PLE ใช้เวลาประมาณ 3-5 ชั่วโมง และ 10-15 ชั่วโมง สำหรับกรณี $N = 1,250$ และ $2,500$ ตามลำดับ อย่างไรก็ตามโปรแกรมในเวอร์ชันดังกล่าวยังมีข้อจำกัดเกี่ยวกับขนาดหน่วยความจำในการจัดเก็บค่าพารามิเตอร์ที่ประมาณได้จากขั้นตอนวิธีทวนซ้ำ ซึ่งยังมีขนาดใหญ่มากนัก ดังนั้นผู้ใช้งานควรตรวจสอบความเหมาะสมระหว่างเวอร์ชันของโปรแกรมกับจำนวนพารามิเตอร์ที่ต้องการประมาณค่าในโมเดล รวมถึงขนาดตัวอย่างของข้อมูลก่อนทำการวิเคราะห์

2) ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 การวิจัยนี้เป็นการประยุกต์ใช้โมเดล IRT ทางเลือกที่คำนึงถึงการเดากับโมเดล MMixIRT ซึ่งกำหนดให้มีการวัดความสามารถแบบเอกมิติเท่านั้น เนื่องจากมีข้อจำกัดเกี่ยวกับโปรแกรมวิเคราะห์และเวลาที่ใช้ในการประมาณค่าพารามิเตอร์ ในการทำวิจัยครั้งต่อไป หากมีโปรแกรมที่มีประสิทธิภาพด้านความเร็วมากกว่านี้ ควรมีการขยายขอบเขตการวัดเป็นแบบพหุมิติ ซึ่งอาจช่วยให้โมเดลมีความสอดคล้องกับข้อมูลการทดสอบในปัจจุบันที่มีการวัดความสามารถแบบหลายองค์ประกอบมากขึ้น

2.2 การวิจัยครั้งต่อไปอาจนำโมเดล MMixIRT ไปบูรณาการกับโมเดล IRT ที่มีการตรวจให้คะแนนหลายค่า เช่น โมเดล graded response โมเดล rating scale โมเดล partial credit ซึ่งการประยุกต์ใช้โมเดลดังกล่าวอาจเป็นประโยชน์ต่อการวิเคราะห์ เพื่อจำแนกกลุ่มบุคคลหรือกลุ่มองค์กรตามคุณลักษณะแฝงที่คล้ายคลึงกันในเชิงจิตวิทยา อาทิ บุคลิกภาพ ความภาคภูมิใจในตนเอง ความเป็นผู้นำ ความพึงพอใจ เจตคติ ซึ่งตัวแปรเหล่านี้นิยมวัดด้วยแบบวัดมาตราการประเมิน (rating scale) หลายระดับ หรือแบบวัดที่มีการตรวจให้คะแนนบางส่วน (partial credit)

2.3 โมเดล 2PLE ที่นำมาบูรณาการในงานวิจัยนี้เป็นโมเดล IRT ทางเลือกที่คำนึงถึงการเดา โมเดลล่าสุดในปี ค.ศ. 2019 โดยปรับปรุงแนวคิดมาจากโมเดล 3PL ด้วยการให้โอกาสในการเดาถูกมีค่าผันแปรตามความสามารถของผู้สอบ ความยาก และอำนาจจำแนกของข้อสอบ เพื่อต้องการให้โมเดลมีความสอดคล้องกับธรรมชาติของข้อมูลการตอบข้อสอบมากขึ้น อย่างไรก็ตามโมเดลดังกล่าวอาจมีความเหมาะสมสำหรับการอธิบายข้อมูลการตอบข้อสอบในบางบริบทเท่านั้น ในการทำวิจัยครั้งต่อไปจึงควรเปรียบเทียบผลการวัดที่ได้จากโมเดล MMixIRT ที่มีการบูรณาการมาจากโมเดล IRT ทางเลือกๆ รูปแบบอื่น เพื่อศึกษาถึงความเหมาะสมระหว่างโมเดลกับข้อมูลในบริบทอื่น ๆ เพิ่มเติม



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บรรณานุกรม

ภาษาไทย

- กระทรวงศึกษาธิการ. (2552). หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551. กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.
- กัลยา วานิชย์บัญชา. (2552). การวิเคราะห์ข้อมูลหลายตัวแปร (พิมพ์ครั้งที่ 4). กรุงเทพฯ: ศูนย์หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ชนะศึก นิชานนท์. (2553). ประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบเบสส์โดยใช้การสุ่มอย่างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (วิทยานิพนธ์ปริญญาคุชฎีบัณฑิต), กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- ณัฐธรรณ์ หลาวทอง, สิวะโชติ ศรีสุทธิยากร, & Gerald, W. F. (2563). การพัฒนาดัชนีวัดความเสมอภาคและความเท่าเทียมทางการศึกษาของประเทศไทย. กรุงเทพฯ: สำนักประสานงาน สกว. ศูนย์ CSPA.
- ดำรง ทิพย์โยธา. (2544). ความน่าจะเป็นและสถิติ สรุปเนื้อหา โจทย์แบบฝึกหัดและเฉลย (พิมพ์ครั้งที่ 1). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- พัชร มีวรรณ. (2540). ผลของจำนวนตัวเลือกที่มีต่อค่าฟังก์ชันสารสนเทศของแบบสอบเลือกตอบแบบถูกผิดในระดับการศึกษาที่ต่างกัน (วิทยานิพนธ์ปริญญาโทบัณฑิต), กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- ไพรัตน์ วงษ์นาม. (2555). การวิเคราะห์การถดถอยพหุระดับด้วยโปรแกรม HLM 6. วารสารศึกษาศาสตร์, 23(3), 27-42.
- ศิริชัย กาญจนวาสี. (2554). การวิเคราะห์พหุระดับ (Multi-level analysis) (พิมพ์ครั้งที่ 5). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2555). ทฤษฎีการทดสอบแนวใหม่ (Modern test theories) (พิมพ์ครั้งที่ 4). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2556). ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) (พิมพ์ครั้งที่ 7). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สถาบันทดสอบทางการศึกษาแห่งชาติ. (2558). ข้อมูลการจัดสอบเบื้องต้น: การทดสอบระดับชาติ O-NET [ออนไลน์]. <http://www.niets.or.th/th/catalog/view/211>
- สถาบันทดสอบทางการศึกษาแห่งชาติ. (2562ก). ข้อมูลสำคัญเกี่ยวกับการสอบ O-NET ปีการศึกษา 2562: รูปแบบข้อสอบ และจำนวนข้อสอบ (Test Blueprint) วิชาคณิตศาสตร์ [ออนไลน์]. <https://www.niets.or.th/th/content/view/13777>
- สถาบันทดสอบทางการศึกษาแห่งชาติ. (2562ข). สรุปผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2561 [ออนไลน์]. http://www.newonetestresult.niets.or.th/AnnouncementWeb/PDF/SummaryONETM6_2562.pdf
- สังวรณัฏฐ์ ังคระโทก. (2552). คุณภาพการสอนวิทยาศาสตร์และความสามารถทางวิทยาศาสตร์ของนักเรียนไทย: ข้อค้นพบ และข้อเสนอแนะเชิงนโยบายจากการประเมินนักเรียนระดับนานาชาติ. กรุงเทพฯ: สภาวิจัยแห่งชาติ.

- สิวะโชติ ศรีสุทธิยากร. (2555). วิธีการประมาณค่าแบบเบย์สำหรับการวิเคราะห์ข้อมูลทุกระดับที่ตัวแปรมีความคลาดเคลื่อนจากการวัด: การศึกษาศาสนาการณณ์จำลองแบบมอนติคาร์โลและข้อมูลจริง (วิทยานิพนธ์ปริญญา ดุษฎีบัณฑิต), กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- อรพรรณ แก้วน้อย. (2557). การเปรียบเทียบคุณสมบัติทางจิตมิติของแบบสอบถูกผิดหลายตัวเลือกกับแบบสอบถูกผิดหลายตัวเลือกปรับปรุงใหม่ที่ใช้วิธีการตรวจให้คะแนนที่ต่างกัน (วิทยานิพนธ์ปริญญาโท), กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

ภาษาอังกฤษ

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Anderson, I. G., & Anderson, S. C. (2017). Student-centered instruction and academic achievement: Linking mechanisms of educational inequality to schools' instructional strategy. *British journal of sociology of education*, 38(4), 533-550.
- Andrich, D., & Marais, I. (2014). Person proficiency estimates in the dichotomous Rasch model when random guessing is removed from difficulty estimates of multiple choice items. *Applied Psychological Measurement*, 38(6), 432-449.
- Asparouhov, T., & Muthén, B. (2008). Multilevel mixture model. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 25-51). Information Age Publishing, Inc.
- Bacci, S., & Ginaldi, M. (2015). A classification of university courses based on students' satisfaction: An application of a two-level mixture item response model. *Quality & Quantity*, 49(3), 927-940.
- Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science*, 60(2), 434-448.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2019). *lme4: Linear Mixed-Effects Models using "Eigen" and S4 (version 1.1-21) [Computer software manual]*. <https://CRAN.R-project.org/package=lme4>
- Ben-Simon, A., Budescu, D., & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. *Applied Psychological Measurement*, 21(1), 65-88.
- Bereby-Meyer, Y., Meyer, J., & Budescu, D. (2003). Decision making under internal uncertainty: the case of multiple-choice tests with different scoring rules. *Acta Psychologica*, 112(2), 207-220.
- Bereby-Meyer, Y., Meyer, J., & Flacher, O. (2002). Prospect theory analysis of guessing in multiple choice tests. *Journal of Behavioral Decision Making*, 15(4), 313-327.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord & Novick (Eds.), *Statistical theories of mental test scores* (pp. 395-479). Addison - Wesley Publishing.
- Bo, Y., Lewis, C., & Budescu, D. (2015). An option-based partial credit item response model. *Quantitative Psychology Research*, 89, 45-72.

- Bolt, D., Cohen, A., & Wollack, J. (2001). A mixture model for multiple choice data. *Journal of Educational and Behavioral Statistics*, 26(4), 381-409.
- Bolt, D., Cohen, A., & Wollack, J. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, 39(4), 331-348.
- Bozdogan, H. (1987). Model Selection and Akaike's Information Criterion (AIC): the General Theory and Its Analytic Extensions. *Psychometrika*, 52(3), 345-370.
- Budescu, D., & Bar-Hillel, M. (1993). To guess or not to guess: a decision theoretic view of formula scoring. *Journal of Educational Measurement*, 30(4), 277-291.
- Budescu, D., & Bo, Y. (2015). Analyzing test-taking behavior: decision theory meets psychometric theory. *Psychometrika*, 80(4), 1105-1122.
- Bush, M. (2015). Reducing the need for guesswork in multiple-choice tests. *Assessment & Evaluation in Higher Education*, 40(2), 218-231.
- Cao, J., & Stokes, S. (2008). Bayesian IRT Guessing Models for Partial Guessing Behaviors. *Psychometrika*, 73(2), 209-230.
- Cho, S. J. (2007). *A multilevel mixture IRT model for DIF analysis*. Unpublished doctoral dissertation, University of Georgia, Athens, GA.
- 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.
- Choi, Y. (2014). *Metric identification in mixture IRT models*. Unpublished doctoral dissertation, University of Georgia, Athens, GA.
- Cohen, A., & Bolt, D. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement*, 42(2), 133-148.
- Cohen, A., Wollack, J., Bolt, D., & Mroch, A. (2002). *A mixture Rasch model analysis of test speededness* the annual meeting of the American Educational Research Association, New Orleans: LA.
- Congdon, P. (2003). *Applied Bayesian modelling*. John Wiley.
- Crosen, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448-474.
- Draper, D. (1995). Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society, Series B*, 57(1), 45-97.
- Dressel, P., & Schmidt, J. (1953). Some modifications of the multiple choice item. *Educational and Psychological Measurement*, 13(4), 574-595.
- Echternacht, G. (1976). Reliability and validity of option weighting schemes. *Educational and Psychological Measurement*, 36(2), 301-309.

- Finch, W., & Finch, M. (2013). Investigation of specific learning disability and testing accommodations based differential item functioning using a multilevel multidimensional mixture item response theory model. *Educational and Psychological Measurement, 73*(6), 973-993.
- Fox, J. P., & Glas, C. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika, 66*(2), 271-288.
- Frary, R. (1989). Partial-credit scoring methods for multiple-choice tests. *Applied Measurement in Education, 2*(1), 79-96.
- Frisbie, D. (1992). The multiple true-false item format: A status review. *Educational Measurement: Issues and Practice, 11*(4), 21-26.
- Frühwirth-Schnatter, S. (2006). *Finite mixture and markov switching models*. Springer.
- Gelman, A., & Rubin, D. (1992). A single series from the Gibbs sampler provides a false sense of security. In J. Bernardo, J. Berger, A. Dawid, & A. Smith (Eds.), *Bayesian Statistics 4* (pp. 625-632). Oxford Univ. Press.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. Bernardo, J. Berger, A. Dawid, & A. Smith (Eds.), *Bayesian Statistics 4* (pp. 169-193). Oxford Uni. Press.
- Gibbons, J., Olkin, I., & Sobel, M. (1977). *Selecting and ordering populations: a new statistical methodology*. Wiley.
- Goudie, R. J., Turner, R. M., De Angelis, D., & Thomas, A. (2017). *Massively parallel MCMC for Bayesian hierarchical models*. <https://arxiv.org/abs/1704.03216>
- Guo, H. (2017). Predicting rights-only score distributions from data collected under formula score instructions. *Psychometrika, 82*(1), 1-16.
- Hambleton, R., Roberts, D., & Traub, R. (1970). A comparison of the reliability and validity of two methods for assessing partial knowledge on a multiple-choice test. *Journal of Educational Measurement, 7*(2), 75-82.
- Hambleton, R., & Swaminathan, H. (1985). *Item response theory: Principles and applications* (2nd Ed.). Kluwer Nijhoff Publishing.
- Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of item response theory*. Sage.
- Han, K. (2012). Fixing the c Parameter in the Three-Parameter Logistic Model. *Practical Assessment, Research & Evaluation, 17*(1), 1-24.
- Heidelberger, P., & Welch, P. (1983). Simulation run length control in the presence of an initial transient. *Operations Research, 31*(6), 1109-1144.
- Henry, K. L., & Muthén, B. (2010). Multilevel latent class analysis: an application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling, 17*(2), 193-215.

- Holzinger, K. (1924). On scoring multiple response tests. *Journal of Educational Psychology*, 15(7), 445-447.
- Hox, J. J., & Maas, C. J. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*, 8(2), 157-174.
- Hsu, T., Moss, P., & Choosak, K. (1984). The merits of multiple-answer items as evaluated by using six scoring formulaes. *Journal of Experimental Education*, 52(3), 152-158.
- Hutchinson, T. P. (1991). *Ability, partial information and guessing: Statistical modelling applied to multiple-choice tests*. Rumsby Scientific Publishing.
- Jacobs, S. (1971). Correlates of unwarranted confidence in responses to objective test items. *Journal of Educational Measurement*, 8(1), 15-19.
- Jang, E. E., & Roussos, L. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement*, 44(1), 1-21.
- 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.
- Lee, W. Y., Cho, S. J., & Sterba, S. K. (2018). Ignoring a multilevel structure in mixture item response models: impact on parameter recovery and model selection. *Applied Psychological Measurement*, 42(2), 136-154.
- 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.
- Lim, R., & Drasgow, F. (1990). Evaluation of two methods for estimating item response theory parameters when assessing different item functioning. *Journal of Applied Psychology*, 75(2), 164-174.
- Lindquist, E., & Hoover, H. (2015). Some notes on corrections for guessing and related problems. *Educational Measurement: Issues and Practice*, 34(2), 15-19.
- Linzer, D., & Lewis, J. (2014). *poLCA: Polytomous variable Latent Class Analysis (version 1.4.1) [Computer software manual]*. <https://CRAN.R-project.org/package=poLCA>
- Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 39(2), 247-264.
- Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Lubke, G., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10(1), 21-39.
- Magidson, J., & Vermunt, J. (2004). Latent class models. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 175-198). Sage.
- Maier, K. S. (2001). A Rasch hierarchical measurement model. *Journal of Educational and Behavioral*

- Statistics*, 26(3), 307-330.
- Mehrens, W., & Lehmann, I. (1984). *Measurement and Evaluation in Education and Psychology*. Holt, Rinehart & Winston.
- Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, 55(2), 195-215.
- Muthén, L., & Muthén, B. O. (1998-2015). *Mplus User's Guide* (7th ed.). Muthén & Muthén.
- Patz, R. J., & Junker, B. W. (1999). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics*, 24(4), 342-366.
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2019). *coda: Output Analysis and Diagnostics for MCMC (version 0.19-3) [Computer software manual]*. <https://CRAN.R-project.org/package=coda>
- Raftery, A. L., & Lewis, S. (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics 4*. Oxford University Press.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. SAGE.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2019). *HLM 8 for Windows [Computer software]*. Scientific Software International, Inc.
- Robitzsch, A. (2020). *sirt: Supplementary Item Response Theory Models (version 3.9-4) [Computer software manual]*. <https://CRAN.R-project.org/package=sirt>
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14(3), 271-282.
- San Martín, E., Del Pino, G., & De Boeck, P. (2006). IRT models for ability-based guessing. *Applied Psychological Measurement*, 30(3), 183-203.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.
- Sen, S. (2014). *Robustness of mixture IRT models to violations of latent normality* University of Georgia]. Athens, GA.
- Sen, S., & Cohen, A. (2020). The impact of test and sample characteristics on model selection and classification accuracy in the multilevel mixture IRT model. *Frontiers in Psychology*, 11(197), 1-8.
- Sen, S., Cohen, A., & Kim, S. (2019). Model selection for multilevel mixture Rasch models. *Applied Psychological Measurement*, 43(4), 272-289.
- Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica*, 7(2), 221-242.
- Smith, R. (1987). Assessing partial knowledge in vocabulary. *Journal of Educational Measurement*, 24(3), 217-231.
- Smith, R. M. (1993). Guessing and the Rasch model. *Rasch Measurement Transactions*, 6(4), 262-263.

- Stephens, M. (2000). Dealing with Label Switching in Mixture Models. *Journal of the Royal Statistical Society, Series B*, 62(4), 795-809.
- Sykes, R., & Hou, L. (2003). Weighting constructed-response items in IRT-based exams. *Applied Measurement in Education*, 16(4), 257-275.
- 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.
- Thissen, D., & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika*, 47(4), 397-412.
- Tsai, T., & Lin, C. (2015, Jan 9-10, 2015). *Modeling Guessing Properties of Multiple-Choice Items in the Measurement of Political Knowledge* the 2015 Asian Political Methodology Conference, Taipei, Taiwan.
- Vermunt, J. (2003). Multilevel latent class models. *Sociological Methodology*, 33(1), 213-239.
- Vermunt, J. (2007). *Multilevel mixture item response theory models: An application in education testing* the 56th session of the International Statistical Institute, Lisbon, Portugal.
- Vermunt, J. (2008). Multilevel latent variable modeling: An application in education testing. *Austrian Journal of Statistics*, 37(3&4), 285-299.
- Vermunt, J., & Magidson, J. (2016). *Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax*. Statistical Innovations Inc.
- von Davier, M. (2001). *WINMIRA 2001 [Computer software manual]*.
<http://208.76.84.140/svflumu/wmira/winmiramanual.pdf>
- Wickham, H., François, R., Henry, L., & Müller, K. (2020). *dplyr: A Grammar of Data Manipulation (version 1.0.2) [Computer software manual]*. <https://cran.r-project.org/package=dplyr>
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*. Elsevier.
- Willse, J. T. (2014). *mixRasch: Mixture Rasch Models with JMLE (version 1.1) [Computer software manual]*. <https://CRAN.R-project.org/package=mixRasch>
- Wise, S., & Kong, X. (2005). Response Time Effort: A New Measure of Examinee Motivation in Computer-Based Tests. *Applied Measurement in Education*, 18(2), 163-183.
- Yang, C. C., & Yang, C. C. (2007). Separating Latent Classes by Information Criteria. *Journal of Classification*, 24(2), 183-203.
- Zhang, J. (2007). Conditional covariance theory and DETECT for polytomous items. *Psychometrika*, 72(1), 69-91.
- 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.



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ก

คำสั่งที่ใช้ในการจำลองข้อมูลแบบมอนติคาร์โลด้วยโปรแกรม R

```

# R Code for generating MMixIRT data
# Number of student-level class (K)=2 & school-level class (G)=2
# Fixing the working directory
setwd("C:/Users/USER/Dropbox/Data_dissertation")
mainFolder <- getwd()
# Loading "dplyr" package
library(dplyr)

# IRT model
model <- "3PL"
#model <- "2PLE"

# Number of items
nItem <- 40

# Number of schools
nSch <- 50

# Number of students (each school)
nStu <- 25
#nStu <- 50

# Number of sample sizes
nSam <- nSch*nStu

# Proportion of school-level class
pi1 <- 0.50; pi2 <- 1-pi1

# Proportion of student-level class (within school-level class)
pi11 <- 0.90; pi21 <- 1-pi11 # school-level class 1
pi12 <- 0.30; pi22 <- 1-pi12 # school-level class 2

# Mean & SD of normal distribution for ability
muTheta11 <- 0; muTheta21 <- 0; muTheta12 <- 0; muTheta22 <- 0
sdTheta11 <- 1; sdTheta21 <- 1; sdTheta12 <- 1; sdTheta22 <- 1

# Number of generating data set
rep <- 10

# Item parameter
# Difficulty parameters
b11 <- c(-2.5, -2.4, -2.3, -2.2, -2.1, -2.0, -1.9, -1.8, -1.7, -1.6,
        -1.5, -1.4, -1.3, -1.2, -1.1, -1.0, -0.9, -0.8, -0.7, -0.6,
         0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5,
         1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5)
b21 <- c(0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5,
        1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5,
        -1.5, -1.4, -1.3, -1.2, -1.1, -1.0, -0.9, -0.8, -0.7, -0.6,
        -2.5, -2.4, -2.3, -2.2, -2.1, -2.0, -1.9, -1.8, -1.7, -1.6)
b12 <- c(-0.5, -0.4, -0.3, -0.2, -0.1, 0.0, 0.1, 0.2, 0.3, 0.4,
        0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4,
        -0.4, -0.3, -0.2, -0.1, 0.0, 0.1, 0.2, 0.3, 0.4, 0.5,
        0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5)

```



```

b22 <- c( 0.6,  0.7,  0.8,  0.9,  1.0,  1.1,  1.2,  1.3,  1.4,  1.5,
         1.6,  1.7,  1.8,  1.9,  2.0,  2.1,  2.2,  2.3,  2.4,  2.5,
        -1.5, -1.4, -1.3, -1.2, -1.1, -1.0, -0.9, -0.8, -0.7, -0.6,
         1.5,  1.6,  1.7,  1.8,  1.9, -2.0, -1.9, -1.8, -1.7, -1.6)
# Discrimination parameters
a11 <- ifelse(b11<=-1,2,1); a21 <- ifelse(b21<=-1,2,1)
a12 <- ifelse(b12<=-1,2,1); a22 <- ifelse(b22<=-1,2,1)
# Guessing parameters
c11 <- ifelse(b11<=-1,0.10,ifelse(b11<1,0.20,0.25))
c21 <- ifelse(b21<=-1,0.10,ifelse(b21<1,0.20,0.25))
c12 <- ifelse(b12<=-1,0.10,ifelse(b12<1,0.20,0.25))
c22 <- ifelse(b22<=-1,0.10,ifelse(b22<1,0.20,0.25))

# Sampling sth school to school-level class 1&2
# Random seeds for generating data set (10 data sets)
# From http://stattrek.com/statistics/random-number-generator.aspx
seed <- c(01522, 23694, 73492, 50656, 69219,
         07267, 96586, 00454, 22885, 98722)
for (iter in 1:rep){
  set.seed(seed[iter])
  index_cSch1 <- sort(sample(x=nSch, size=nSch*pi1, replace=F))
  index_cSch2 <- (1:nSch)[-index_cSch1]
  datSch <- mat.or.vec(nSch,2)
  datSch[,1] <- 1:nSch
  for (i in 1:(nSch*pi1)) {
    datSch[index_cSch1[i],2] <- 1
  }
  for (i in 1:(nSch*pi2)) {
    datSch[index_cSch2[i],2] <- 2
  }

  # Preparing data for each student in each school
  datStu <- mat.or.vec(nSam,4)
  colnames(datStu) <- c("School","cSch","Student","cStu")
  datStu[,1] <- rep(datSch[,1], each=nStu)
  datStu[,2] <- rep(datSch[,2], each=nStu)
  datStu[,3] <- 1:nSam
  datStu <- data.frame(datStu)
  datStu <- arrange(datStu,cSch)

  # Sampling jth student to student-level class 1&2
  # Student-level 1&2 for School-level class 1
  dat.cSch1 <- filter(datStu, cSch==1)
  index_cStu11 <- sort(sample(dim(dat.cSch1)[1], pi1*pi11*nSam,
                             replace=F))
  index_cStu21 <- (1:dim(dat.cSch1)[1])[-index_cStu11]
  for (i in 1:length(index_cStu11)) {
    dat.cSch1[index_cStu11[i],4] <- 1
  }
  for (i in 1:length(index_cStu21)) {
    dat.cSch1[index_cStu21[i],4] <- 2
  }

  # Student-level 1&2 for School-level class 2
  dat.cSch2 <- filter(datStu, cSch==2)
  index_cStu12 <- sort(sample(dim(dat.cSch2)[1], pi2*pi12*nSam,
                             replace=F))
  index_cStu22 <- (1:dim(dat.cSch2)[1])[-index_cStu12]

```

```

for (i in 1:length(index_cStu2)) {
  dat.cSch2[index_cStu2[i],4] <- 1
}
for (i in 1:length(index_cStu22)) {
  dat.cSch2[index_cStu22[i],4] <- 2
}

# Generating response data for cStu=1 (within cSch=1)
dat.cStu11 <- filter(dat.cSch1, cStu==1)
nStu11 <- dim(dat.cStu11)[1]
theta11 <- rnorm(nStu11, mean=muTheta11, sd=sdTheta11)
resp11 <- mat.or.vec(nStu11,nItem)
if(model=="3PL") {
  for (row in 1:nStu11) {
    for (column in 1:nItem) {
      tt <- 1/(1+exp(-a11[column]*(theta11[row]-b11[column])))
      p <- c11[column]+(1-c11[column])*tt
      r <- runif(1,0,1)
      if (r <= p) {
        resp11[row,column] <- 1
      }
      else {
        resp11[row,column] <- 0
      }
    }
  }
}
if(model=="2PLE") {
  for (row in 1:nStu11) {
    for (column in 1:nItem) {
      tt <- 1/(1+exp(-a11[column]*(theta11[row]-b11[column])))
      ss <- 1/(1+exp(a11[column]*(theta11[row]-b11[column])))
      p <- (1/5)^ss+(1-(1/5)^ss)*tt
      r <- runif(1,0,1)
      if (r <= p) {
        resp11[row,column] <- 1
      }
      else {
        resp11[row,column] <- 0
      }
    }
  }
}
resp11 <- data.frame(resp11)
dat11 <- bind_cols(dat.cStu11,resp11)

# Generating response data for cStu=2 (within cSch=1)
dat.cStu21 <- filter(dat.cSch1, cStu==2)
nStu21 <- dim(dat.cStu21)[1]
theta21 <- rnorm(nStu21, mean=muTheta21, sd=sdTheta21)
resp21 <- mat.or.vec(nStu21,nItem)
if(model=="3PL") {
  for (row in 1:nStu21) {
    for (column in 1:nItem) {
      tt <- 1/(1+exp(-a21[column]*(theta21[row]-b21[column])))
      p <- c21[column]+(1-c21[column])*tt
      r <- runif(1,0,1)
      if (r <= p) {

```

```

        resp21[row,column] <- 1
      }
      else {
        resp21[row,column] <- 0
      }
    }
  }
}
if(model=="2PLE") {
  for (row in 1:nStu21) {
    for (column in 1:nItem) {
      tt <- 1/(1+exp(-a21[column]*(theta21[row]-b21[column])))
      ss <- 1/(1+exp(a21[column]*(theta21[row]-b21[column])))
      p <- (1/5)^ss+(1-(1/5)^ss)*tt
      r <- runif(1,0,1)
      if (r <= p) {
        resp21[row,column] <- 1
      }
      else {
        resp21[row,column] <- 0
      }
    }
  }
}
resp21 <- data.frame(resp21)
dat21 <- bind_cols(dat.cStu21,resp21)

# Generating response data for cStu=1 (within cSch=2)
dat.cStu12 <- filter(dat.cSch2, cStu==1)
nStu12 <- dim(dat.cStu12)[1]
theta12 <- rnorm(nStu12, mean=muTheta12, sd=sdTheta12)
resp12 <- mat.or.vec(nStu12,nItem)
if(model=="3PL") {
  for (row in 1:nStu12) {
    for (column in 1:nItem) {
      tt <- 1/(1+exp(-a12[column]*(theta12[row]-b12[column])))
      p <- c12[column]+(1-c12[column])*tt
      r <- runif(1,0,1)
      if (r <= p) {
        resp12[row,column] <- 1
      }
      else {
        resp12[row,column] <- 0
      }
    }
  }
}
if(model=="2PLE") {
  for (row in 1:nStu12) {
    for (column in 1:nItem) {
      tt <- 1/(1+exp(-a12[column]*(theta12[row]-b12[column])))
      ss <- 1/(1+exp(a12[column]*(theta12[row]-b12[column])))
      p <- (1/5)^ss+(1-(1/5)^ss)*tt
      r <- runif(1,0,1)
      if (r <= p) {
        resp12[row,column] <- 1
      }
      else {

```

```

        resp12[row,column] <- 0
      }
    }
  }
}
resp12 <- data.frame(resp12)
dat12 <- bind_cols(dat.cStu12,resp12)

# Generating response data for cStu=2 (within cSch=2)
dat.cStu22 <- filter(dat.cSch2, cStu==2)
nStu22 <- dim(dat.cStu22)[1]
theta22 <- rnorm(nStu22, mean=muTheta22, sd=sdTheta22)
resp22 <- mat.or.vec(nStu22,nItem)
if(model=="3PL") {
  for (row in 1:nStu22) {
    for (column in 1:nItem) {
      tt <- 1/(1+exp(-a22[column]*(theta22[row]-b22[column])))
      p <- c22[column]+(1-c22[column])*tt
      r <- runif(1,0,1)
      if (r <= p) {
        resp22[row,column] <- 1
      }
      else {
        resp22[row,column] <- 0
      }
    }
  }
}
if(model=="2PLE") {
  for (row in 1:nStu22) {
    for (column in 1:nItem) {
      tt <- 1/(1+exp(-a22[column]*(theta22[row]-b22[column])))
      ss <- 1/(1+exp(a22[column]*(theta22[row]-b22[column])))
      p <- (1/5)^ss+(1-(1/5)^ss)*tt
      r <- runif(1,0,1)
      if (r <= p) {
        resp22[row,column] <- 1
      }
      else {
        resp22[row,column] <- 0
      }
    }
  }
}
resp22 <- data.frame(resp22)
dat22 <- bind_cols(dat.cStu22,resp22)

# Combining data set to complete data set
dat <- bind_rows(dat11,dat21,dat12,dat22)
dat <- arrange(dat,Student)
dat <- bind_cols(dat[, (5:length(dat))],
                 select(dat,c(cSch,cStu,School)))

#### The End of generating complete data set ####
# Generating output file
resp <- dat[,1:nItem]
filename1 <- paste("MMix",model,"_N",nSam,"_I",nItem,"_K2G2_",
                  iter,".txt",sep="")

```

```

# Original file containing class information
filename2 <- paste("MMix",model,"_N",nSam,"_I",nItem,"_K2G2_C",
                  iter, ".txt", sep="")

# Data with generated class information
write.table(dat, file=filename2, sep=',', row.names=F, col.names=F)

# Data for MultiBUGS script
write("# Input Data", file=file.path(mainFolder, filename1), append=T)
write(paste("# ", nStu, " students, ", nSch, " schools, ", nItem, "
            items", sep=""), file=file.path(mainFolder, filename1),
      append=T)
write(paste("list(J=", nSam, ", I=", nItem, ", S=", nSch, ", K=2, G=2,
            alpha=c(1,1), alphas=c(1,1),", sep=""),
      file=file.path(mainFolder, filename1), append=T)
write("group=c(", file=file.path(mainFolder, filename1), append=T)

group <- rep(1:nSch, each=nStu)
group <- matrix(group, ncol=nStu, byrow=T)
for (s in 1:nSch) {
  write(group[s,], sep=",", ncolumns=nStu,
        file=file.path(mainFolder, filename1), append=T)
}
write(")", file=file.path(mainFolder, filename1), append=T)

temp <- readLines(file.path(mainFolder, filename1), n=-1)
for (i in 5:(4+nSch-1)) {
  temp[i] <- paste(temp[i], ",", sep="")
}
write(temp, file=file.path(mainFolder, filename1), append=F)
write("resp=structure(.Data=c(",
      file=file.path(mainFolder, filename1), append=T)
for (j in 1:nSam) {
  write(data.matrix(resp[j,]), sep=",", ncolumns=nItem,
        file=file.path(mainFolder, filename1), append=T)
}
write(paste(")", .Dim=c(", nSam, ", nItem, ")), sep=""),
      file=file.path(mainFolder, filename1), append=T)
temp <- readLines(file.path(mainFolder, filename1), n=-1)
for (i in (7+nSch):(6+nSch+nSam-1)) {
  temp[i] <- paste(temp[i], ",", sep="")
}
write(temp, file=file.path(mainFolder, filename1), append=F)
}

```

ภาคผนวก ข

คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์ด้วยโปรแกรม MultiBUGS

ข.1 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล 3PL

```

# I: the number of items
# J: the number of students
# eta: ability
# sigma: the SD of ability
# a: item discrimination; b: item difficulty; c: item guessing

model {
# Response data
  for (j in 1:J) {
    for (i in 1:I) {
      r[j,i] <- resp[j,i]
    }
  }
# Generated responses from 3PL model
  for (j in 1:J) {
    for (i in 1:I) {
      logit(t[j,i]) <- a[i]*(sigma*eta[j] - b[i])
      p[j,i] <- c[i] + (1-c[i])*t[j,i]
      r[j,i] ~ dbern(p[j,i])
    }
  }
# Ability
  for (j in 1:J) {
    eta[j] ~ dnorm(0,1)
  }
# SD of Ability
  sigma ~ dnorm(0,1) T(0,)
# Item Parameters
  for (i in 1:I) {
    a[i] ~ dnorm(0,1) T(0,)
    b[i] ~ dnorm(0,1)
    c[i] ~ dbeta(5,17)
  }
# Log-Likelihood
  for (j in 1:J) {
    for (i in 1:I) {
      l[j,i] <- log(p[j,i])*r[j,i] + log(1-p[j,i])*(1-r[j,i])
    }
  }
  loglik <- sum(l[1:J,1:I])
  par <- 3*I + 1
  AIC <- -2*loglik + 2*par
  BIC <- -2*loglik + par*log(J)
  CAIC <- -2*loglik + par*(log(J)+1)
  DBIC <- -2*loglik + par*(log(J)-log(2*22/7))
}

```

ข.2 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล 2PLE

```

# I: the number of items
# J: the number of students
# eta: ability
# sigma: the SD of ability
# a: item discrimination; b: item difficulty

model {
# Response data
  for (j in 1:J) {
    for (i in 1:I) {
      r[j,i] <- resp[j,i]
    }
  }
# Generated responses from 2PLE model
  for (j in 1:J) {
    for (i in 1:I) {
      logit(t[j,i]) <- a[i]*(sigma*eta[j] - b[i])
      logit(s[j,i]) <- -logit(t[j,i])
      g[j,i] <- pow(1/5, s[j,i])
      p[j,i] <- g[j,i] + (1-g[j,i])*t[j,i]
      r[j,i] ~ dbern(p[j,i])
    }
  }
# Ability
  for (j in 1:J) {
    eta[j] ~ dnorm(0,1)
  }
# SD of Ability
  sigma ~ dnorm(0,1) T(0,)
# Item Parameters
  for (i in 1:I) {
    a[i] ~ dnorm(0,1) T(0,)
    b[i] ~ dnorm(0,1)
  }
# Log-Likelihood
  for (j in 1:J) {
    for (i in 1:I) {
      l[j,i] <- log(p[j,i])*r[j,i] + log(1-p[j,i])*(1-r[j,i])
    }
  }
  loglik <- sum(l[1:J,1:I])
  par <- 2*I + 1
  AIC <- -2*loglik + 2*par
  BIC <- -2*loglik + par*log(J)
  CAIC <- -2*loglik + par*(log(J)+1)
  DBIC <- -2*loglik + par*(log(J)-log(2*22/7))
}

```

ข.3 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล multilevel 3PL

```

# I: the number of items
# J: the number of students
# S: the number of schools
# eta: ability at the student-level
# tau: the SD of ability at the student-level (sqrt of tau in model)
# u: ability at the school-level
# gamma: the mean of ability at the school-level
# zeta: the inverse variance of ability at the school-level
# a: item discrimination; b: item difficulty; c: item guessing

model {
# Response data
  for (j in 1:J) {
    for (i in 1:I) {
      r[j,i] <- resp[j,i]
    }
  }
# Generated responses from multilevel 3PL model
  for (j in 1:J) {
    for (i in 1:I) {
      logit(t[j,i]) <- a[i]*(tau*eta[j]+u[group[j]])-b[i]
      p[j,i] <- c[i] + (1-c[i])*t[j,i]
      r[j,i] ~ dbern(p[j,i])
    }
  }
# Ability at the Student-level
  for (j in 1:J) {
    eta[j] ~ dnorm(0,1)
  }
# SD of Ability at the Student-level
  tau ~ dnorm(0,1) T(0,)
# Ability at the School-level
  for (s in 1:S) {
    u[s] ~ dnorm(gamma,zeta)
  }
# Mean & (1/Variance) of Ability at the School-level
  gamma ~ dnorm(0,1)
  zeta ~ dgamma(0.1,0.001)
# Item Parameters
  for (i in 1:I) {
    a[i] ~ dnorm(0,1) T(0,)
    b[i] ~ dnorm(0,1)
    c[i] ~ dbeta(5,17)
  }
# Log-Likelihood
  for (j in 1:J) {
    for (i in 1:I) {
      l[j,i] <- log(p[j,i])*r[j,i] + log(1-p[j,i])*(1-r[j,i])
    }
  }
  loglik <- sum(l[1:J,1:I])
  par <- 3*I + 3
  AIC <- -2*loglik + 2*par
  BIC <- -2*loglik + par*log(J)
  CAIC <- -2*loglik + par*(log(J)+1)
  DBIC <- -2*loglik + par*(log(J)-log(2*22/7))
}

```


ข.4 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล multilevel 2PLE

```

# I: the number of items
# J: the number of students
# S: the number of schools
# eta: ability at the student-level
# tau: the SD of ability at the student-level (sqrt of tau in model)
# u: ability at the school-level
# gamma: the mean of ability at the school-level
# zeta: the inverse variance of ability at the school-level
# a: item discrimination; b: item difficulty
model {
# Response data
  for (j in 1:J) {
    for (i in 1:I) {
      r[j,i] <- resp[j,i]
    }
  }
# Generated responses from multilevel 2PLE model
  for (j in 1:J) {
    for (i in 1:I) {
      logit(t[j,i]) <- a[i]*(tau*eta[j]+u[group[j]])-b[i]
      logit(ss[j,i]) <- -logit(t[j,i])
      g[j,i] <- pow(1/5, ss[j,i])
      p[j,i] <- g[j,i] + (1-g[j,i])*t[j,i]
      r[j,i] ~ dbern(p[j,i])
    }
  }
# Ability at the Student-level
  for (j in 1:J) {
    eta[j] ~ dnorm(0,1)
  }
# SD of Ability at the Student-level
  tau ~ dnorm(0,1) T(0,)
# Ability at the School-level
  for (s in 1:S) {
    u[s] ~ dnorm(gamma,zeta)
  }
# Mean & (1/Variance) of Ability at the School-level
  gamma ~ dnorm(0,1)
  zeta ~ dgamma(0.1,0.001)
# Item Parameters
  for (i in 1:I) {
    a[i] ~ dnorm(0,1) T(0,)
    b[i] ~ dnorm(0,1)
  }
# Log-Likelihood
  for (j in 1:J) {
    for (i in 1:I) {
      l[j,i] <- log(p[j,i])*r[j,i] + log(1-p[j,i])*(1-r[j,i])
    }
  }
  loglik <- sum(l[1:J,1:I])
  par <- 2*I + 3
  AIC <- -2*loglik + 2*par
  BIC <- -2*loglik + par*log(J)
  CAIC <- -2*loglik + par*(log(J)+1)
  DBIC <- -2*loglik + par*(log(J)-log(2*22/7))
}

```

ข.5 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล Mix3PL

```

# I: the number of items
# J: the number of students
# K: the number of student-level class
# kk: group membership at the student-level
# eta: ability
# mu: the mean of ability
# sigma: the SD of ability
# a: item discrimination; b: item difficulty; c: item guessing

model {
# Response data
  for (j in 1:J) {
    for (i in 1:I) {
      r[j,i] <- resp[j,i]
    }
  }
# Generated responses from Mix3PL model
  for (j in 1:J) {
    for (i in 1:I) {
      logit(t[j,i]) <- a[i,kk[j]]*(sigma[kk[j]]*eta[j]-b[i,kk[j]])
      p[j,i] <- c[i,kk[j]] + (1-c[i,kk[j]])*t[j,i]
      r[j,i] ~ dbern(p[j,i])
    }
  }
# Ability
  for (j in 1:J) {
    eta[j] ~ dnorm(mu[kk[j]],1)
  }
# Mean of Ability
  mu[1] <- 0
  for (k in 2:K) {
    mu[k] ~ dnorm(0,1)
  }
# SD of Ability
  for (k in 1:K) {
    sigma[k] ~ dnorm(0,1) T(0,)
  }
# Student-level membership
  for (j in 1:J) {
    kk[j] ~ dcat(pi[1:K])
  }
  pi[1:K] ~ ddirich(alpha[])
# Item Parameters
  for (i in 1:I) {
    for (k in 1:K) {
      a[i,k] ~ dnorm(0,1) T(0,)
      b[i,k] ~ dnorm(0,1)
      c[i,k] ~ dbeta(5,17)
    }
  }
# Log-Likelihood
  for (j in 1:J) {
    for (i in 1:I) {
      l[j,i] <- log(p[j,i])*r[j,i] + log(1-p[j,i])*(1-r[j,i])
    }
  }
  loglik <- sum(l[1:J,1:I])
}

```

```
par <- (3*I*K) + (2*K-1) + (K-1)
AIC <- -2*loglik + 2*par
BIC <- -2*loglik + par*log(J)
CAIC <- -2*loglik + par*(log(J)+1)
DBIC <- -2*loglik + par*(log(J)-log(2*22/7))
}
```



ข.6 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล Mix2PLE

```

# I: the number of items
# J: the number of students
# K: the number of student-level class
# kk: group membership at the student-level
# eta: ability
# mu: the mean of ability
# sigma: the SD of ability
# a: item discrimination; b: item difficulty

model{
# Response data
  for (j in 1:J) {
    for (i in 1:I) {
      r[j,i] <- resp[j,i]
    }
  }
# Generated responses from Mix2PLE model
  for (j in 1:J) {
    for (i in 1:I) {
      logit(t[j,i]) <- a[i,kk[j]]*(sigma[kk[j]]*eta[j]-b[i,kk[j]])
      logit(s[j,i]) <- -logit(t[j,i])
      g[j,i] <- pow(1/5, s[j,i])
      p[j,i] <- g[j,i] + (1-g[j,i])*t[j,i]
      r[j,i] ~ dbern(p[j,i])
    }
  }
# Ability
  for (j in 1:J) {
    eta[j] ~ dnorm(mu[kk[j]],1)
  }
# Mean of Ability
  mu[1] <- 0
  for (k in 2:K) {
    mu[k] ~ dnorm(0,1)
  }
# SD of Ability
  for (k in 1:K) {
    sigma[k] ~ dnorm(0,1) T(0,)
  }
# Student-level membership
  for (j in 1:J) {
    kk[j] ~ dcat(pi[1:K])
  }
  pi[1:K] ~ ddirich(alpha[])
# Item Parameters
  for (i in 1:I) {
    for (k in 1:K) {
      a[i,k] ~ dnorm(0,1) T(0,)
      b[i,k] ~ dnorm(0,1)
    }
  }
# Log-Likelihood
  for (j in 1:J) {
    for (i in 1:I) {
      l[j,i] <- log(p[j,i])*r[j,i] + log(1-p[j,i])*(1-r[j,i])
    }
  }
}

```

```
loglik <- sum(l[1:J,1:I])  
par <- (2*I*K) + (2*K-1) + (K-1)  
AIC <- -2*loglik + 2*par  
BIC <- -2*loglik + par*log(J)  
CAIC <- -2*loglik + par*(log(J)+1)  
DBIC <- -2*loglik + par*(log(J)-log(2*22/7))  
}
```



ข.7 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล MMix3PL

```

# I: the number of items
# J: the number of students
# S: the number of schools
# K: the number of student-level class
# G: the number of school-level class
# kk: group (student) membership at the student-level
# gg: group (school) membership at the school-level
# eta: ability
# mu: the mean of ability
# sigma: the SD of ability
# a: item discrimination; b: item difficulty; c: item guessing

model{
# Response data
  for (j in 1:J) {
    for (i in 1:I) {
      r[j,i] <- resp[j,i]
    }
  }
# Generated responses from MMix3PL model
  for (j in 1:J) {
    for (i in 1:I) {
      logit(t[j,i]) <- a[i,kk[j],gg[group[j]]]*
        (sigma[kk[j],gg[group[j]]]*eta[j] - b[i,kk[j],gg[group[j]]])
      p[j,i] <- c[i,kk[j],gg[group[j]]] + (1 -
        c[i,kk[j],gg[group[j]]])*t[j,i]
      r[j,i] ~ dbern(p[j,i])
    }
  }
# Ability
  for (j in 1:J) {
    eta[j] ~ dnorm(mu[kk[j],gg[group[j]]], 1)
  }
# Mean of Ability
  mu[1,1] <- 0
  for (k in 2:K) {
    mu[k,1] ~ dnorm(0,1)
  }
  for (g in 2:G) {
    mu[1,g] ~ dnorm(0,1)
  }
  for (k in 2:K) {
    for (g in 2:G) {
      mu[k,g] ~ dnorm(0,1)
    }
  }
# SD of Ability
  for (k in 1:K) {
    for (g in 1:G) {
      sigma[k,g] ~ dnorm(0,1) T(0,)
    }
  }
# Student Level
  for (j in 1:J) {
    kk[j] ~ dcat(pi[gg[group[j]],1:K])
  }
  for (g in 1:G) {

```

```

    pi[g,1:K] ~ ddirich(alpha[])
  }
# School Level
for (s in 1:S) {
  gg[s] ~ dcat(pi1[1:G])
}
pi1[1:G] ~ ddirich(alpha1[])
# Item Parameters
for (i in 1:I) {
  for (k in 1:K) {
    for (g in 1:G){
      a[i,k,g] ~ dnorm(0,1) T(0,)
      b[i,k,g] ~ dnorm(0,1)
      c[i,k,g] ~ dbeta(5,17)
    }
  }
}
# Log-Likelihood
for (j in 1:J) {
  for (i in 1:I) {
    l[j,i] <- log(p[j,i])*r[j,i]+log(1-p[j,i])*(1-r[j,i])
  }
}
loglik <- sum(l[1:J,1:I])
par <- (3*I*K*G) + (2*K*G-1) + (G-1) + G*(K-1)
AIC <- -2*loglik + 2*par
BIC <- -2*loglik + par*log(J)
CAIC <- -2*loglik + par*(log(J)+1)
DBIC <- -2*loglik + par*(log(J)-log(2*22/7))
}

# Initial value of School-Level group membership
# Required for MMixIRT (K=2,G=2) model
list(
  gg=c(1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,
    1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,
    ...
    1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2)
)

```

ข.8 คำสั่งที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับโมเดล MMix2PLE

```

# I: the number of items
# J: the number of students
# S: the number of schools
# K: the number of student-level class
# G: the number of school-level class
# kk: group (student) membership at the student-level
# gg: group (school) membership at the school-level
# eta: ability
# mu: the mean of ability
# sigma: the SD of ability
# a: item discrimination; b: item difficulty

model{
# Response data
  for (j in 1:J) {
    for (i in 1:I) {
      r[j,i] <- resp[j,i]
    }
  }
# Generated responses from MMix2PLE model
  for (j in 1:J) {
    for (i in 1:I) {
      logit(t[j,i]) <- a[i,kk[j],gg[group[j]]]*
        (sigma[kk[j],gg[group[j]]]*eta[j] -
        b[i,kk[j],gg[group[j]]])
      logit(ss[j,i]) <- -logit(t[j,i])
      guess[j,i] <- pow(1/5, ss[j,i])
      p[j,i] <- guess[j,i] + (1-guess[j,i])*t[j,i]
      r[j,i] ~ dbern(p[j,i])
    }
  }
# Ability
  for (j in 1:J) {
    eta[j] ~ dnorm(mu[kk[j],gg[group[j]]], 1)
  }
# Mean of Ability
  mu[1,1] <- 0
  for (k in 2:K) {
    mu[k,1] ~ dnorm(0,1)
  }
  for (g in 2:G) {
    mu[1,g] ~ dnorm(0,1)
  }
  for (k in 2:K) {
    for (g in 2:G) {
      mu[k,g] ~ dnorm(0,1)
    }
  }
# SD of Ability
  for (k in 1:K) {
    for (g in 1:G) {
      sigma[k,g] ~ dnorm(0,1) T(0,)
    }
  }
# Student Level
  for (j in 1:J) {
    kk[j] ~ dcat(pi[gg[group[j]],1:K])
  }

```



```

}
for (g in 1:G) {
  pi[g,1:K] ~ ddirich(alpha[])
}
# School Level
for (s in 1:S) {
  gg[s] ~ dcat(pi1[1:G])
}
pi1[1:G] ~ ddirich(alpha1[])
# Item Parameters
for (i in 1:I) {
  for (k in 1:K) {
    for (g in 1:G){
      a[i,k,g] ~ dnorm(0,1) T(0,)
      b[i,k,g] ~ dnorm(0,1)
    }
  }
}
# Log-Likelihood
for (j in 1:J) {
  for (i in 1:I) {
    l[j,i] <- log(p[j,i])*r[j,i]+log(1-p[j,i])*(1-r[j,i])
  }
}
loglik <- sum(l[1:J,1:I])
par <- (2*I*K*G) + (2*K*G-1) + (G-1) + G*(K-1)
AIC <- -2*loglik + 2*par
BIC <- -2*loglik + par*log(J)
CAIC <- -2*loglik + par*(log(J)+1)
DBIC <- -2*loglik + par*(log(J)-log(2*22/7))
}

# Initial value of School-Level group membership
# Required for MMixIRT (K=2,G=2) model
list(
  gg=c(1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,
       1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,
       ...
       1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2,1,2)
)

```



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ประวัติผู้เขียน

ชื่อ-สกุล	ธีรุตม์ สุขสกุลวัฒน์
วัน เดือน ปี เกิด	12 สิงหาคม พ.ศ. 2529
วุฒิการศึกษา	วท.บ. คณิตศาสตร์ (เกียรตินิยม) คณะวิทยาศาสตร์ จุฬาฯ, พ.ศ. 2551 สต.ม. สถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาฯ, พ.ศ. 2554
ผลงานตีพิมพ์	1) ธีรุตม์ สุขสกุลวัฒน์, ญัฐภรณ์ หลาวทอง และสิวะโชติ ศรีสุทธิยากร. (2564). การเดากับการพัฒนาโมเดลทฤษฎีการตอบสนองข้อสอบทางเลือกที่ คำนึงถึงการเดา. วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม ปี ที่ 27 ฉบับที่ 1 (มกราคม - มิถุนายน 2564) (บทความวิชาการ). 2) ธีรุตม์ สุขสกุลวัฒน์, ญัฐภรณ์ หลาวทอง และสิวะโชติ ศรีสุทธิยากร. (2564). โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ. วารสาร ศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 24 ฉบับที่ 1 (มกราคม - มีนาคม 2564) (บทความวิชาการ). 3) ธีรุตม์ สุขสกุลวัฒน์, ญัฐภรณ์ หลาวทอง และสิวะโชติ ศรีสุทธิยากร. (2565). การพัฒนาโมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับ โดยใช้โมเดลทฤษฎีการตอบสนองข้อสอบทางเลือกที่คำนึงถึงการเดา. วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 24 ฉบับที่ 4 (ตุลาคม - ธันวาคม 2565) (บทความวิจัย).

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY